<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Just tweet it - The collection, processing, classification and analysis of 2 million fitness tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Vickey, Theodore</td>
</tr>
<tr>
<td><strong>Publication Date</strong></td>
<td>2018-06-13</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>NUI Galway</td>
</tr>
<tr>
<td><strong>Item record</strong></td>
<td><a href="http://hdl.handle.net/10379/7649">http://hdl.handle.net/10379/7649</a></td>
</tr>
</tbody>
</table>

Some rights reserved. For more information, please see the item record link above.
Just Tweet It
The Collection, Processing, Classification and Analysis of 2 Million Fitness Tweets.

Theodore A. Vickey

Submitted in fulfillment of the requirements for the degree of
Doctor of Philosophy

Supervisor:
Dr. John Breslin

External Examiner: Prof. Dr. Alan Smeaton
Internal Examiner: Prof. Dr. Gearóid ÓLaighin

National University of Ireland, Galway (NUIG)
June 2018
Table of Contents

1 Introduction .................................................................................................................. 1
  1.1 Background and Motivation .................................................................................. 1
  1.2 Research Questions ............................................................................................... 2
  1.3 Hypotheses ............................................................................................................. 3
  1.4 Research Method .................................................................................................... 4
  1.5 Conceptual Model .................................................................................................. 6
  1.6 Expectations .......................................................................................................... 6
  1.7 Ethical Considerations of Twitter Research.......................................................... 7
    1.7.1 Informed Consent ............................................................................................. 7
    1.7.2 Privacy ............................................................................................................. 8
  1.8 Approval from NUIG Research Ethics Committee ................................................... 9
  1.9 Structure ............................................................................................................... 9
  1.10 Summary .............................................................................................................. 11

2 Literature Review and Background .......................................................................... 12
  2.1 Online Social Networks ....................................................................................... 12
  2.2 Twitter .................................................................................................................. 13
    2.2.1 The Structure of a Tweet ................................................................................ 14
    2.2.2 Overall Twitter Usage .................................................................................... 15
    2.2.3 User Intention – What People Tweet ............................................................... 15
  2.3 Physical Activity .................................................................................................... 17
    2.3.1 The Rise of Physical Inactivity ....................................................................... 18
    2.3.2 Reasons for Physical Inactivity ....................................................................... 19
    2.3.3 Costs Associated with Physical Inactivity ....................................................... 20
  2.4 Related Work ......................................................................................................... 21
  2.5 How Social Networking Is Used to Share Personal Health Data.............................. 21
    2.5.1 Social Networking and Physical Activity ....................................................... 22
    2.5.2 The Influence of Those Who Tweet ................................................................. 23
    2.5.3 Measurement of Physical Activity .................................................................. 24
  2.6 Twitter Data Mining and Text Classification .......................................................... 25
  2.7 Klout ..................................................................................................................... 26
  2.8 Summary ............................................................................................................... 27

3 Data Collection and Processing ............................................................................... 28
  3.1 Content Analysis .................................................................................................... 29
    3.1.1 Determination of Major Classifications ......................................................... 32
    3.1.2 Testing of Method to Independent Review Coders .......................................... 32
    3.1.3 Percentage Agreement ................................................................................... 33
  3.2 Selection Criteria for Mobile Fitness Apps .............................................................. 34
    3.2.1 Endomondo ..................................................................................................... 38
    3.2.2 RunKeeper ....................................................................................................... 39
    3.2.3 Nike+ ............................................................................................................... 41
    3.2.4 MyFitnessPal ................................................................................................... 42
    3.2.5 DailyMile ........................................................................................................ 44
  3.3 Data Collection from Twitter .................................................................................. 47
    3.3.1 TwapperKeeper ............................................................................................... 48
    3.3.2 The Fitness Tweet Crawler ............................................................................. 51
  3.4 The Fitness Tweet Classification Model ................................................................... 55
    3.4.1 Activity ............................................................................................................. 56
    3.4.2 Blarney ............................................................................................................. 58
    3.4.3 Conversation ................................................................................................... 59
    3.4.4 Fitness Tweet Classification Model Script ....................................................... 60
  3.5 Conclusion .............................................................................................................. 67

4 Results Analysis and Interpretation of the Fitness Tweets ..................................... 68
  4.1 Descriptive Analysis of App-shared Tweets ............................................................. 68
  4.2 Analysis of Fitness Tweet Classifications ............................................................... 71
Table List

Table 3-1 - Previous Classification Research ................................................................. 31
Table 3-2 - Percent Agreement with Original Scale among Four Independent Review Coders .................................................. 33
Table 3-3 – Popularity Ranking of Mobile Fitness Apps in US iTunes Store, December 2010 and July 2018 .................. 35
Table 3-4 - Popularity Ranking of Mobile Fitness Apps in UK iTunes Store, December 2010 and 2018 ..................... 36
Table 3-5 - Popularity Ranking of Android Fitness Apps, December 2010 and July 2018 .................................................. 36
Table 3-6 - Combined Popularity Ranking of Mobile Fitness Apps, December 2010 .................................................. 37
Table 3-7 - Beta Test of Data-collection Results .................................................................. 38
Table 3-8 - Mobile Fitness App Comparisons (as of December 2012) .................................................. 46
Table 3-9 - Collected Twitter Data Point Descriptions from TwapperKeeper .......................... 50
Table 3-10 - Twitter Data Point Descriptions from Twitter for the Fitness Tweet Crawler ......................... 51
Table 3-11 - Twitter Data Point Descriptions from Klout for the Fitness Tweet Crawler ......................... 52
Table 3-12 - Sample of Klout Data Demographics ................................................................ 54
Table 3-13 - Sample Activity Tweets .................................................................................. 58
Table 3-14 - Sample Blarney Tweets .................................................................................. 59
Table 3-15 - Sample Conversation Tweets ........................................................................... 60
Table 4-1 - Statistical Analysis of DailyMile Fitness Tweets ................................................... 78
Table 4-2 - DailyMile Top 10 Users .................................................................................. 79
Table 4-3 - Statistical Analysis of MyFitnessPal Fitness Tweets ............................................... 80
Table 4-4 - MyFitnessPal Top 10 Users .............................................................................. 81
Table 4-5 - Statistical Analysis of Endomondo Fitness Tweets ............................................... 82
Table 4-6 - Endomondo Top 10 Users ................................................................................... 83
Table 4-7 - Statistical Analysis of Nike+ Fitness Tweets ......................................................... 84
Table 4-8 - Top 10 Nike+ Users ......................................................................................... 85
Table 4-9 - Statistical Analysis of RunKeeper Fitness Tweets .................................................. 86
Table 4-10 - Top 10 RunKeeper Users ............................................................................... 87
Table 4-11 - Demographics of Workout+ Tweeters .............................................................. 89
Table 4-12 - Top 10 Twitter Accounts Followed by Workout+ Tweeters ............................... 93
Table 4-13 - Top 10 Topics Discussed by Workout+ Tweeters ............................................... 94
Table 4-14 - Top 10 Brands Followed by Workout+ Tweeters ............................................... 94
Table 4-15 - Top 10 Likes and Interests of Workout+ Tweeters ............................................ 95
Table 4-16 - Top 10 Professions of Workout+ Tweeters ....................................................... 95
Table 4-17 - Top 10 Countries of Origin of Workout+ Tweeters .......................................... 96
Table 4-18 - Top 10 Cities of Origin of Workout+ Tweeters .................................................. 96
Table 5-1 - Klout Score quartiles ....................................................................................... 106
Table 5-2 - Klout Score by activity tweet and gender (N=583,252) ..................................... 106
Table 5-3 - Workout and workout+ tweets by Klout quartile ............................................. 107
Table 5-4 - Minutes exercised by gender and Klout Score among workout group ................ 108
Table 5-5 - Total number of tweets by sentiment and app .................................................. 110
Table 5-6 - Healthy People 2020 Baseline and Targets ...................................................... 120
Table 5-7 - User Tables with First Use Date, Total Minutes and Weekly Status .................... 125
Table 6-1 - Summary of Hypotheses ............................................................................... 132
**Figure List**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1-1</td>
<td>Hypotheses</td>
<td>4</td>
</tr>
<tr>
<td>Figure 1-2</td>
<td>Use of social networks to share from mobile fitness app</td>
<td>5</td>
</tr>
<tr>
<td>Figure 1-3</td>
<td>Data Flow for Research</td>
<td>6</td>
</tr>
<tr>
<td>Figure 1-4</td>
<td>Sequence of Chapters</td>
<td>10</td>
</tr>
<tr>
<td>Figure 3-1</td>
<td>Content Analysis Steps and Actions</td>
<td>30</td>
</tr>
<tr>
<td>Figure 3-2</td>
<td>Top 200 Health and Fitness Apps</td>
<td>35</td>
</tr>
<tr>
<td>Figure 3-3</td>
<td>Top 200 Health and Fitness Apps from</td>
<td>36</td>
</tr>
<tr>
<td>Figure 3-4</td>
<td>Endomondo Screenshots</td>
<td>39</td>
</tr>
<tr>
<td>Figure 3-5</td>
<td>RunKeeper Screenshots</td>
<td>40</td>
</tr>
<tr>
<td>Figure 3-6</td>
<td>RunKeeper Member Page (Web)</td>
<td>40</td>
</tr>
<tr>
<td>Figure 3-7</td>
<td>Nike+ Screenshots</td>
<td>41</td>
</tr>
<tr>
<td>Figure 3-8</td>
<td>Nike+ Member Page (Web)</td>
<td>42</td>
</tr>
<tr>
<td>Figure 3-9</td>
<td>MyFitnessPal Member Page (Web)</td>
<td>43</td>
</tr>
<tr>
<td>Figure 3-10</td>
<td>MyFitnessPal Screenshots (Web)</td>
<td>44</td>
</tr>
<tr>
<td>Figure 3-11</td>
<td>DailyMile Member Page (Web)</td>
<td>45</td>
</tr>
<tr>
<td>Figure 3-12</td>
<td>Screenshot of TwapperKeeper</td>
<td>49</td>
</tr>
<tr>
<td>Figure 3-13</td>
<td>Step 1 of the Fitness Tweet Crawler</td>
<td>52</td>
</tr>
<tr>
<td>Figure 3-14</td>
<td>Step 2 of the Fitness Tweet Crawler</td>
<td>52</td>
</tr>
<tr>
<td>Figure 3-15</td>
<td>Fitness Tweet Classification Model</td>
<td>55</td>
</tr>
<tr>
<td>Figure 3-16</td>
<td>Flowchart of Classification Script</td>
<td>56</td>
</tr>
<tr>
<td>Figure 4-1</td>
<td>Collected Fitness Tweets by Language (Excluding English)</td>
<td>69</td>
</tr>
<tr>
<td>Figure 4-2</td>
<td>English vs. Non-English Tweets</td>
<td>70</td>
</tr>
<tr>
<td>Figure 4-3</td>
<td>Total Users by Classification</td>
<td>71</td>
</tr>
<tr>
<td>Figure 4-4</td>
<td>Daily Tweets by Apps</td>
<td>72</td>
</tr>
<tr>
<td>Figure 4-5</td>
<td>Fitness Tweet Classification by Major Category</td>
<td>73</td>
</tr>
<tr>
<td>Figure 4-6</td>
<td>Classification of Fitness Tweets by Mobile Fitness App by Number of Tweets</td>
<td>74</td>
</tr>
<tr>
<td>Figure 4-7</td>
<td>Cross App Comparison of Sub-category</td>
<td>75</td>
</tr>
<tr>
<td>Figure 4-8</td>
<td>Percentage of Users Who only Shared Activity Once via Twitter</td>
<td>77</td>
</tr>
<tr>
<td>Figure 4-9</td>
<td>DailyMile Breakdown of Fitness Tweets</td>
<td>78</td>
</tr>
<tr>
<td>Figure 4-10</td>
<td>DailyMile Time Period Analysis by Date</td>
<td>79</td>
</tr>
<tr>
<td>Figure 4-11</td>
<td>MyFitnessPal Breakdown of Fitness Tweets</td>
<td>80</td>
</tr>
<tr>
<td>Figure 4-12</td>
<td>MyFitnessPal Time Period Analysis</td>
<td>81</td>
</tr>
<tr>
<td>Figure 4-13</td>
<td>Endomondo Breakdown of Fitness Tweet</td>
<td>82</td>
</tr>
<tr>
<td>Figure 4-14</td>
<td>Endomondo Time Period Analysis by Date</td>
<td>83</td>
</tr>
<tr>
<td>Figure 4-15</td>
<td>Nike+ Breakdown of Fitness Tweets</td>
<td>84</td>
</tr>
<tr>
<td>Figure 4-16</td>
<td>Nike+ Time Period Analysis by Date</td>
<td>85</td>
</tr>
<tr>
<td>Figure 4-17</td>
<td>RunKeeper Breakdown of Fitness Tweets</td>
<td>86</td>
</tr>
<tr>
<td>Figure 4-18</td>
<td>RunKeeper Time Period Analysis by Date</td>
<td>87</td>
</tr>
<tr>
<td>Figure 4-19</td>
<td>Gender Demographics</td>
<td>90</td>
</tr>
<tr>
<td>Figure 4-20</td>
<td>Relationship Demographics</td>
<td>91</td>
</tr>
<tr>
<td>Figure 4-21</td>
<td>Age Demographics</td>
<td>92</td>
</tr>
<tr>
<td>Figure 4-22</td>
<td>Income Demographics</td>
<td>92</td>
</tr>
<tr>
<td>Figure 4-23</td>
<td>Ethnic Demographics</td>
<td>92</td>
</tr>
<tr>
<td>Figure 5-1</td>
<td>The Fitness Tweet Classification Model</td>
<td>104</td>
</tr>
<tr>
<td>Figure 5-2</td>
<td>Overall percentage of sentiment by all mobile fitness apps</td>
<td>109</td>
</tr>
<tr>
<td>Figure 5-3</td>
<td>Percentage of sentiment by mobile fitness app</td>
<td>110</td>
</tr>
<tr>
<td>Figure 5-4</td>
<td>Word cloud of Workout+ Tweets</td>
<td>111</td>
</tr>
<tr>
<td>Figure 5-5</td>
<td>User Page from Shortened URL Link in Fitness Tweet</td>
<td>121</td>
</tr>
<tr>
<td>Figure 5-6</td>
<td>Fitness Tweet Classification Model</td>
<td>122</td>
</tr>
<tr>
<td>Figure 5-7</td>
<td>Mobile Fitness App vs. Healthy People 2020</td>
<td>123</td>
</tr>
<tr>
<td>Figure 5-8</td>
<td>Average Weekly Minutes of Physical Activity by App</td>
<td>124</td>
</tr>
<tr>
<td>Figure 6-1</td>
<td>Number of Activity Tweets per Klout Quartile</td>
<td>133</td>
</tr>
<tr>
<td>Figure 6-2</td>
<td>Number of Activity Tweets per Klout Quartile by Gender</td>
<td>134</td>
</tr>
<tr>
<td>Figure 6-3</td>
<td>Number of Workout Tweets per Klout Quartile by Gender</td>
<td>135</td>
</tr>
<tr>
<td>Figure 6-4</td>
<td>Number of Workout+ Tweets per Klout Quartile by Gender</td>
<td>136</td>
</tr>
<tr>
<td>Figure 6-5</td>
<td>Positive vs. Negative Workout+ Tweets</td>
<td>139</td>
</tr>
<tr>
<td>Figure 6-6</td>
<td>Mobile Fitness Apps vs. Healthy People 2020 (copy of Figure 6.7)</td>
<td>140</td>
</tr>
</tbody>
</table>
Declaration

I declare that this thesis is composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified. The research presented in this dissertation was supported by the Irish Research Council Enterprise Partnership Scheme, the American Council on Exercise, and Science Foundation Ireland under grant number SFI/08/CE/I1380 (Lion 2) and a University Write-up Bursary in 2015. I also declare that I was at the time of this research an unpaid member of the American Council on Exercise (ACE) Board of Directors and no member other person at NUIG was affiliated with ACE during the time of this research.

Theodore A. Vickey
July 4, 2018
Published Papers in Peer-reviewed Journals:


Conference Papers:


Book Chapters:


Additional Journal Paper Contributions:

Presentations:

1. Mobile Fitness Apps and Data Collection (2015) in Redmond, WA – Microsoft – Oral Presentation


Acknowledgements
As with many parts of life, things cannot be accomplished alone; and my research is no different. There have been many people who, through their unselfish acts of kindness, allowed me to achieve this accomplishment. It was a truly international effort.

To my USS team at DERI, thank you for the support, laughs and insight. A piece of each of you is reflected in this dissertation, grazie, dziękuję. To Mengjiao Wang, 感谢您为您的计算机知识和友谊。你永远是我的兄弟。To Dr. Maciej Dabrowski, Dziękuję za nauczenie mej starej dupy nowych sposobów na wiele rzeczy. To my advisor, mentor and friend Dr. John Breslin, Thóg tú an deis ar an madra d'aois le smaoineamh nua, go raibh maith agat do gach go atá déanta agat.

To Scott Goudeseune, Dr. Cedric Bryant and the rest of the team from the American Council on Exercise, thank you for believing in my dream. To my Canadian team of Dr. Kathleen Martin Ginis and Dr. Shane Sweet, thank you for seeing the potential in my idea and your help in validating it.

To my Vickey and O’Neill family, you continue to inspire me to greatness, especially to Dad for your unending love and support, sister Katie for your wizardry with the pencil—you are my Picasso—and Uncle Bob and Aunt Carrie for a quiet spot in Maine for inspiration and support and the many smiles and laughs from nephews Sean and Tate, nieces Maeve, Sophia, Norah and Lila.

To my Point Loma University family lead by Dr. Jeff “Hezekiah” Sullivan and Dr. Nicole Cosby, thanks for showing me the light at the end of the tunnel.
ABSTRACT

In 2013, the World Health Organization coined the term “Globesity” to highlight the importance of the epidemic and impact as a major health problem in many parts of the world (World Health Organization, 2013). Many believe that technology has been a key to this decrease in physical and increase in sedentary behaviors. However, the effective use of technology may be one of the keys towards better health, fitness and wellness for the global population using a device they already use, a mobile phone.

This research investigates a novel and scalable method of physical activity data collection through the use of Twitter. The publicly available fitness tweets can provide a wealth of demographic and activity information from Twitter users from around the world. Specifically, this dissertation reviews the use of five mobile fitness apps that allow participants to share their workouts from the app through Twitter. At the time of the data collection, this was the first known academic research in the world that combined Twitter, mobile fitness apps and physical activity. The underlying hypotheses and research questions explore if researchers can learn enough information to help decrease the incidence of physical inactivity from a population level.

There were three research questions to be answered. The first was to determine if an automated data collector could be created and to then accurately identify fitness tweets shared from one’s mobile fitness app from the Twitter stream. If so, how can these fitness tweets be collected and processed and what are the limitations in the data processing of these fitness tweets? The second was to determine if an automated fitness tweet classification model could accurately quantify characteristics of physical activity such as but not limited to duration, type and intensity. If so, what is the process and what are the limitations in the collection of physical-activity minutes using Twitter? The third and final was to determine if additional demographic information from those who share their workouts online using Twitter be generated to give additional insights into the characteristics of the type of person whom fitness tweets?

Building on existing research tools and literature on data collection, analysis and classification from computer science research, a Fitness Tweet Collection Tool and a Fitness Tweet Classification Model were created to process and catalogue over 2 million fitness tweets. The tools were designed in such a way that would allow future health and wellness researchers, without an in-depth knowledge of programming or coding, to modify the tools to conduct their own Twitter and health research. One such research project that used this data collection model was by researchers at Harvard studying skin cancer.

Four hypotheses are presented from the resulting fitness tweets: that as one’s online influence increases, the busier they will be, thus less time to exercise; that women are more expressive in their tweets and would tend to fitness tweet more; that people feel better when they exercise, and this will be reflected
in their tweets and that those who use mobile fitness apps are more interested in their physical activity, thus will report greater minutes of activity than those that were self-reported through surveys.

The findings of this research and the three data experiments conducted from the collected data provide implications for not only future researchers, but also for mobile fitness developers. This research demonstrates that it is possible to collect self-reported health information, in this case physical activity, from a diverse population at different levels of physical activity from around the world.
1 INTRODUCTION

As much as technology has enriched society and expanded global communication, it can be argued that it has also negatively affected overall global health by reducing opportunities for physical activity and contributing to an overall decline in physical activity participation rates (Foster, Linehan & Kirman, 2010). The reasons for, and potential solutions to, this lack of physical activity are complex, multifaceted and interdisciplinary.

Physical inactivity contributed to approximately 3.2 million deaths in 2014 and was the fourth leading risk factor for premature death around the world (World Health Organization, 2014). In 2018, much of the world is becoming less active as countries develop economically, levels of inactivity increase with levels as high as 70% with changes in transportation patterns, increased use of technology, urbanization and cultural values (World Health Organization, 2018). Despite the fact that one-third of adults and four-fifths of adolescents do not reach established daily physical-activity requirements, smartphone apps that collect and in some cases promote physical activity are popular (Middelweerd, Mollee, van der Wal, Brug & Te Velde, 2014). It is, therefore, worthwhile to study how these mobile fitness apps can be better utilized to effectively promote physical activity, as more than half of U.S. adults own a smartphone (Fox & Duggan, 2013).

1.1 Background and Motivation

With the increased use of both social networks and smartphones, can there be a way to measure physical activity by using these technology tools? The motivation for this research is to create a new model to allow for the collection of self-reported physical activity data that would allow for researchers to determine if such collected data can be a valid and reliable alternative research tool.

Recent data-collection enhancements using technology have provided researchers with a new method to collect real-time, accurate and valid data. These technologies offer different opportunities to expand the way researchers think of survey data collection, increasing the ways researchers can interact with survey respondents and expanding the range of stimulus materials that can be used (Couper, 2005). Large-scale physical activity assessment using on body technology now provides objective measurement to reliably detect association with multiple health outcomes (Doherty et al., 2017).

One such enhancement is the use of sensors to collect physical activity data. This type of data collection differs from the traditional self-reported interviews on which previous health and fitness research has been based, with data collected through face-to-face interviews or telephone surveys. For example, in the Healthy People 2010 data collection, researchers telephoned randomly selected adults and asked two series of questions that assessed (1) how often during leisure time they participated in vigorous-intensity activities that caused heavy sweating or large increases in breathing or heart rate for greater
than 10 minutes at a time and (2) how often they participated in light- or moderate-intensity activities that caused only light sweating or a slight to moderate increase in breathing or heart rate. Participants provided both frequency of participation (per day, week, month or year) and duration of each session (Carlson & Fulton, 2010).

To be able to keep pace with the development of technology, innovative tools are needed that can collect data in real time, from a diverse population around the world, using any number of mobile fitness apps through which users share their physical activity on their social network. A standard model of classification will allow researchers in epidemiology to work in tandem with researchers in digital science.

1.2 Research Questions

The purpose of this research was to develop an understanding of the types of information specific to physical activity that is being shared from mobile fitness apps via Twitter. The specific research questions were threefold:

RQ1 – Can an automated data collector accurately identify fitness tweets shared from one’s mobile fitness app from the Twitter stream?
- If so, how can these fitness tweets be collected and processed?
- What are the limitations in the data processing of these fitness tweets?

RQ2 – Can an automated fitness tweet classification model quantify characteristics of physical activity such as but not limited to duration, type, and intensity?
- If so, what is the process?
- What are the limitations in the collection of physical-activity minutes using Twitter?

RQ3 – Can additional demographic information from those who share their workouts online using Twitter be generated to give additional insights into the characteristics of the type of person whom fitness tweets?

This research highlights that researchers can use Twitter for physical-activity measurement on a wide scale using hundreds, if not tens of thousands, of users. While tweets are limited to only 140 characters, a wealth of additional information can be collected through data crawling and data scraping using hyperlinks from users’ physical-activity tweets from their mobile fitness app (including, but not limited to, heart rate, route, music selection, elevation, and duration). Academic research using Twitter is in its infancy (Zimmer & Proferes, 2014), which does cause challenges and concerns. However, these challenges and concerns also provide opportunities such as academic disciplines successfully using Twitter as a robust data-collection tool (Lee et al., 2011; Gaffney & Puschmann, 2014; Bruns & Stieglitz, 2013). At the time of this writing, this research is the first known use of Twitter data collection using mobile fitness apps within physical-activity research.
This research also highlights the method of data collection using Twitter and the data processing to include additional user demographic information (including, but not limited to, self-reported location, GPS location, time zone, followers, followings, Klout influence and overall Twitter activity). This designed data-collection and data-processing model has already been used in other health-related research, including a Harvard-based skin cancer research project\(^1\). In addition, this research introduces a classification model specific to physical-activity research using Twitter that can be replicated for future research.

### 1.3 Hypotheses

After the creation of the research questions, four hypotheses were established for this research, as shown in Figure 1-3 (the reasons and outcomes will be described in Chapter 7).

- **H1** – The more influential a person is, the busier they will be, and the less time they will have to exercise;
- **H2** – Women are more expressive in their tweets and would tend to fitness tweet more;
- **H3** – People feel better when they exercise, and this will be reflected in their tweets;
- **H4** – Those that use mobile fitness apps are more interested in their physical activity, thus will report greater minutes of activity than those that were self-reported through surveys.

The creation of the tools and the subsequent data mining used in the research questions allows for a number of different hypotheses to be addressed. With the advancement of the use of social media to influence our daily lives, such as the recent election of Donald Trump, Brexit and the UK and the constitutional amendment in Ireland regarding the right to life, one may surmise if similar influence could be seen with health and physical activity (H1) and general online influence. In addition, having a greater understanding of who uses Twitter to share health information, in this case, physical activity, the developers of disruptive health technology apps (DHT) can better design the DHT to allow for the greatest possible outcome of health (H2). As the mobile fitness apps allow for a real-time reporting of not only the quantitative data from the physical activity but also the emotional demeanour of the user, the sentiment of the physical activity can be measured and by doing so provide additional insights into the dynamics that make up a positive physical activity experience (H3). Finally, could the use of mobile technology help decrease inherent survey bias such as recall and ego when collecting national data on the health of a nation (H4).

---

1.4 Research Method

The workflow model for this research was to identify a process to collect fitness-related tweets from the general public using five predetermined mobile fitness apps. To accomplish this, a fitness tweet crawler for data collection was created using previous models from other academic fields of study. As it collected tweets, the fitness tweet crawler also connected with Twitter to collect demographics for each user. Once the tweet data was collected, these fitness tweets were categorized using a newly...
defined classification model. This new database of fitness tweets was then used in three separate experiments to explore different and unique uses of the collected fitness data.

Some of the more popular mobile fitness apps at the time this research was conducted allowed users to share their workouts with their social network. This sharing ability was one of the criteria used to determine which apps were used in this research. At that time, there were three places users could share their workouts online from within these apps: (1) the mobile fitness app website itself, along with user-connected (2) Twitter and (3) Facebook accounts (Figure 1-1). As of 2018, these three possible social networks used to share physical activity data have stayed the same.

Twitter was chosen as the social network for this research because the mobile fitness app user tweets were publicly available. Attempts were made to request anonymized data from the mobile fitness app developers to analyse the sharing of physical-activity data among the three social network platforms, but none of the contacted developers would release this internal data. Thus, the publicly available Twitter data was chosen as the main data source. The primary source for the data collected for this research was Twitter, through the use of five mobile fitness apps (Nike+, RunKeeper, MyFitnessPal, Endomondo, and DailyMile). At the time of this research, there were no other peer-reviewed studies that used Twitter to collect physical activity data and limited knowledge about how people use mobile fitness apps to share their physical activity with their social networks.

The data for this research was collected from publicly available tweets between April 21, 2011, and September 21, 2011, using five mobile fitness apps (Nike+, Endomondo, RunKeeper, MyFitnessPal and DailyMile) and Twitter profile pages. There was no recruitment of users, only collection of publicly available data from Twitter. There were 165,768 unique users within the 184-day data set, accounting for 1,982,653 individual tweets (English only). A more detailed explanation of the data is provided starting in Chapter 4.

While this dissertation does not answer all of these needs, it does provide the framework that will allow other researchers to expand upon the findings with the hopes of continued collaboration of interdisciplinary fields of research.
1.5 Conceptual Model

To help guide this research, the following data flow model was created (Figure 1-2) as well as the overall conceptual model (Figure 1-3).

The data flow model, which is the foundation for this research, was created using peer-reviewed concepts from other academic disciplines. The conceptual model highlights the four main hypotheses for this research that is a result of the Fitness Tweet Database that was created. These models are described in more detail in subsequent chapters.

1.6 Expectations

This research aims to introduce a new and novel way for health/fitness/wellness researchers to use publicly available self-reported health-related data (in this case physical-activity data). In particular, this dissertation investigates the possibility of using Twitter data as a way to collect valuable and insightful physical-activity data from one of five mobile fitness apps. As suggested in the research
questions and hypotheses, it is possible to collect a wide and diverse set of data points specific to physical activity that can provide additional insights into metrics such as the type of exercise, duration, social influence and online conversations around such activity.

In addition, from the research questions and hypotheses, a framework and model for collecting and classifying fitness tweets is possible. This integrated database of physical-activity data from Twitter and subsequent links to user profiles from Twitter could allow for a robust reporting model that can be used for comparison between demographic data, user characteristics and other sources of user data.

It is also hoped that all of the tools used in this dissertation have been developed in such a way that future health, wellness, and fitness researchers can easily modify them for their own future research, thus enhancing the possible tools for academic research in a number of related fields of study.

1.7 Ethical Considerations of Twitter Research

Informed consent and privacy are basic ethical tenets of scientific research involving people (Eysenbach & Till, 2001), yet there is a belief among some computer science researchers that there is a need for more simplified consideration of research ethics when compared to non-computer science research (Bruns, Burgess, Crawford & Shaw, 2011).

Findings from Zimmer & Proferes (2014) suggest benchmarks on the privacy, data sharing and ethical concerns for Twitter-based research, but also that the rapid growth in the diversity of disciplines in which Twitter research takes place challenges the ability of scholars to agree on the particular ethical concerns.

In a study of 310 Twitter-based studies, only 16 (4%) made any mention of ethical issues or considerations in relation to the research design and data collection. Of these 16, only six made reference to obtaining ethical review board approval, with five acknowledging the presence of ethical issues that shaped how the Twitter data was collected and managed (Zimmer & Proferes, 2014). With such an overwhelming volume of available data, there is a lack of ethical standards around the use of human subjects by computer science researchers (Carpenter & Dittrich, 2013).

1.7.1 Informed Consent

The first tenant of the ethics in research is the informed consent of the participant. However, in computer science research, the participant could be anyone with an Internet connection. Therefore, the attainment of informed consent may not be possible (Carpenter & Dittrich, 2013).

Some computer science researchers view informed consent differently from other disciplines since tweets are publicly available and there is no human interaction with the person sending the tweet. In addition, Twitter’s Terms of Service explicitly state that tweets are published to the Internet and thus
“public,” and, therefore, the collection of the tweets does not require ethical review approval, as such disciplines are historically outside the purview of ethical review boards (Bruns et al., 2011; Carpenter & Dittrich, 2011). Whilst there remains conflicting issues that contribute to the complexity of good ethical practice in social media research, this should not deter researchers from conducting this type of research but rather to understand that this type of research should be based on current evidence and standardization to avoid discrepancies between different institutions and jurisdictions (Golder, Ahmed, Norman, & Booth, 2017).

1.7.2 Privacy

The second tenant of ethics in research is privacy. Privacy and security within data and information management are very related to each other, while at the same time complex and multifaceted representing a wide range of issues and challenges, to which a variety of solutions have been applied (Kirrane, Villata, & D’Aquin, 2018). The Internet holds possible privacy dangers for computer science researchers, who can easily and unintentionally violate the privacy of the individuals being observed (Eysenbach & Till, 2001). Specific to Twitter users, in 2009 fewer than 10% of users took steps to ensure privacy by restricting access to their accounts (Moore, 2009). In response to an inquiry from United States Senator Chuck Grassley in April of 2018, Twitter responded to the inquiry regarding privacy as follows:

“Twitter is public by its nature. That is the key feature of our platform and what sets us apart from many other internet companies. Through this public platform, we are committed to providing a service that fosters and facilitates free and open democratic debate, and we do so by making it possible for people to react to, comment on, engage with, and criticize content that they or other accounts choose to share. Public Tweets are viewable and searchable by any person who accesses our platform. While we do offer people the ability to share non-public, protected Tweets with their authorized followers, the vast majority of people on Twitter choose to engage with public content and to post their content for public engagement. Users thus make decisions about what they are sharing with the world, as opposed to just their friends or followers.” (Twitter, 2018)

The Twitter Terms of Service acknowledge that users retain their rights to their content and that by submitting, posting or displaying content on or through Twitter, users grant Twitter a worldwide, non-exclusive, royalty-free licence to use, copy, reproduce, process, adapt, modify, publish, transmit, display and distribute such content in any and all media or distribution methods now known or later developed (Twitter, 2011a).

Academic social-science researchers have been quick to recognize the value in studying Twitter data on many topics (Weller, Bruns, Burgess, Mahrt & Puschmann, 2014). Recent research in Ireland around the #RepealThe8th vote in May of 2018 will provide researchers a dearth of information that years prior
would have been difficult to collect and analyze. Never before have researchers had access to such a rich collection of everyday, real-time communication. Yet, in spite of the steps already used by researchers, much more remains to be done to fully develop an interdisciplinary suite of methodologies, tools and conceptual frameworks for the study of Twitter and its impact upon society (Weller et al., 2014).

1.8 Approval from NUIG Research Ethics Committee

Data collection for academic research using Twitter is often challenging due to the ethical concerns already highlighted. To alleviate these privacy concerns for this research, approval was sought and given by the NUI Galway Research Ethics Committee. Individual users were not contacted for written consent for participation in the study, as it was determined that by agreeing to the terms and conditions set forth by Twitter, consent was granted for research.

It was agreed that even though Twitter activity is technically in the public domain, some users may have a concern with the research use of the data linked to their names as an infringement on their privacy. To alleviate this concern, privacy measures were put in place to anonymize tweets from Twitter usernames. Each user was assigned a unique identifier (USERxxxxx). This unique identifier allowed for the linking of databases without the need to use the user’s Twitter name, allowing measurements and analyses between the databases.

All reporting and publications from this research used these unique identifiers. All records linking study participant identification with identifying features were stored confidentially and complied with the university’s policy of Data Retention.

As this research was based on publicly available and anonymized data, no interventions were needed, and there were no identified potential risks or cause of discomfort or distress, either mental or physical. Data was stored on password-protected cloud servers and/or a password-protected laptop.

1.9 Structure

The remainder of this dissertation has been organized in the following manner (Figure 1-4). Chapter 2 introduces two concepts: social networking and physical activity within the realm of mobile health. In addition, Chapter 2 presents a multidisciplinary approach that combines research on computer science and physical activity.

Chapter 2 provides a literature review and background on social networks (more specifically Twitter), physical activity and the concept of mobile health and the importance of how health information is shared using online social networks, the concept of online influence and existing methods to measure physical activity. The chapter concludes with a discussion of existing models and procedures to collect information using Twitter through data mining and a description of text classification.
Chapter 3 discusses the design of the tools used within this research—a data-collection model for publicly available online social network posts from Twitter and a data-classification model that can classify data from the collection model specific to the needs of the research. Both of these tools were created to allow for future replication and enhancement by future researchers.

Chapter 4 provides a discussion of the results of analysis and interpretation of the fitness tweets, including a descriptive analysis using the five different mobile fitness apps, an analysis of the fitness tweet classifications, and the addition of sub-classifications for a more in-depth analysis opportunity and concludes with an overview of the demographic characteristics of those who shared Workout+ tweets including, but not limited to, their inferred gender, age range, income level, ethnicity, religion and location.

Chapter 5 combines the concepts from Chapter 3 (the tools) and Chapter 4 (the data) and provides an analysis of the collected data as three distinct experiments using the collected and processed data from the Fitness Tweet Data Set to address both the research questions and hypotheses. The first of these
experiments is an analysis of fitness tweets to determine the significance of influence. The second is an analysis of fitness tweets to correlate physical activity with the sentiments provided by the participants. Finally, the third is an analysis of a subsection of the fitness tweets to examine the differences between the data used for Healthy People 2020 and the self-reported physical-activity measurements from mobile fitness apps and Twitter.

Chapter 6 is a critical assessment of this work, including a review of the research questions; the precision, thoroughness, and contribution of this dissertation to the field of study, and a comparison with other types of data collection of health, fitness and/or wellness information using online social networks.

1.10 Summary
Technology has disrupted our world, and when used effectively has the ability to improve our health, rather than hinder it. The adoption of apps on smartphones that many of us use not on a daily basis, but dozens of time during our waking day (in in some cases our sleepless nights) in academic and research terms a new and emerging field. With technological capacity more than a million times more powerful than the computer used to put to put a man on the moon in 1969, we keep smartphones in our pockets, ready and able to provide insights, motivation and education on better health and wellness. I believe the effective use of the types disruptive health technologies presented within the dissertation is my generation’s moon-shot, and I am proud that this research has allowed me to be considered a leader in this global movement.
2 LITERATURE REVIEW AND BACKGROUND

This research focuses on the enhanced potential of researchers to use mobile phones and health and fitness apps to collect wellness information from both the general public in a covert observational situation as well as in a more structured laboratory participant observational setting. This chapter discusses the two foundations for this research. The first foundation is Twitter, specifically the structure, usage and user intention of the online social network. The second foundation is physical activity.

2.1 Online Social Networks

With interfaces that allow people to follow the lives of friends, acquaintances and family members, the number of people on social networks has grown exponentially since the turn of this century (Huberman, Romero & Wu, 2009). A social networking site (SNS) is an online platform that focuses on building and enhancing online and/or offline relationships among people who have common ideas, activities, interests or passions. For those who use online SNSs such as Facebook, LinkedIn, Twitter and countless others, participation has become an indispensable part of their daily lives (Fraser & Dutta, 2010).

Scholars, advertisers, and political activists see the massive growth of online social networks as a representation of social interactions that can be used to study the propagation of ideas, social bond dynamics and viral marketing, among others (Huberman et al., 2009).

Gottlieb & Bergen (2010) define a social network as “…a unit of a social structure composed of the individual’s social ties and the ties among them.” In their research, Bramoullé, Djebbari, and Fortin (2009) defined a social network as:

“…a social structure made of nodes (which are generally individuals or organizations) that are tied by one or more specific types of interdependency, such as friendship, values, beliefs, conflict or trade. The resulting graph-based structures are often very complex.”

(p.42)

The linkages between nodes in a social network enable communication and influence (O’Malley & Marsden, 2008). In the past few years, there has been an exponential growth of many SNSs cutting across national boundaries and representing every conceivable community (Fraser & Dutta, 2010). Although their dramatic increase and popularity may have plateaued, social networks will remain an important part of the Internet (Breslin & Decker, 2007).

Research by Kumar, Novak & Tomkins (2010) suggests that many online social networks have similar characteristics and growth patterns. The density of an online social network measures the number of interconnections per person. SNSs can provide powerful tools and opportunities that society has never experienced before. Breslin and Decker (2007) suggest that:
“…social networking sites usually offer the same basic functionalities: a network of friend’s listings (showing a person’s “inner circle”), private and group messaging, discussion forums or communities, events management, blogging and media uploading. With such features, SNSs demonstrate how the Internet continues to better connect people for various social and professional purposes.”

Individuals connected to each other via an SNS may display similar characteristics, such as similar political views, academic performance, body size and health behaviors (Christakis & Fowler, 2009). Recent events such as the Brexit vote and the US presidential election of 2016 have shown how social media is becoming entrenched in our daily lives. Researchers from the Insight Centre in Ireland published a Twitter sentiment gold standard for the Brexit referendum by providing a resource for observing the social and discourse dynamics behind the referendum categorized in five main classes (Hurlimann et al., 2016). In the 2016 US presidential election, social networks including Twitter were heavily used by both candidates as a direct source of news and access to the electorate, bypassing the traditional media (Enli, 2017)

2.2 Twitter

Online social networking services have eliminated the four walls of brick and mortar found in traditional networking and social interaction (Vickey, 2010). Twitter is one of the main social networks that users of mobile fitness apps can use to share their workouts. To understand the relationship between mobile fitness apps and Twitter, we must first understand Twitter.

Twitter is a microblogging service. Microblogging is a form of communication in which users can describe their current status in short posts distributed by instant messages, mobile phones, email or the web (Java, Song, Finin & Tseng, 2007). Twitter serves as a platform for developing new connections and building conversations and has become an important tool in personal, professional and academic settings (Vega, Parthasarathy & Torres, 2010). Founded in 2006 by Jack Dorsey and Biz Stone, Twitter became mainstream after winning the South By Southwest (SXSW) web award in 2007 (Twitter, 2011b). What set Twitter apart from other microblogging services is how its users have adopted Twitter in ways that were never intended by its developers, including as a political tool to physically and geographically organise communities (Vega et al., 2010). Twitter has fast become commonplace in our daily lives and is also becoming a subject for academic research. From its beginnings as a novelty “status updating” service, it has evolved into its current form, which is more focused on mirroring the events of the real world (Cheong & Ray, 2011).

With the increasing research work on Twitter, the fields of study and application of the microblogging service have also broadened. Cheong and Ray (2011) originally reviewed literature from disciplines such as social media and the web, anthropology, computer-human interaction, data mining, knowledge discovery, and visualization. More recently, their Twitter and microblogging research has been
expanding to include areas such as terrorism informatics, user modeling and personalization, online security, spam detection and information streaming. The research for this dissertation is one of the first examinations of the relationship between Twitter and physical activity.

2.2.1 The Structure of a Tweet

The structure of Twitter is simple: users send messages (tweets) to a network of people (followers) from a variety of devices (desktop computers, mobile phones, etc.). Tweets are text-based messages of up to 140 characters in length\(^2\) (Twitter, 2017a). The default setting for tweets is public, which permits Twitter users to follow and read each other’s tweets. Each user has a Twitter page where all of his or her tweets are aggregated into a single list (Jansen, Zhang, Sobel & Chowdury, 2009). Twitter has also embraced third-party developers from the onset by providing a versatile application programming interface (API), and it also enjoys unprecedented popularity with celebrities (Curran, O’Hara & O’Brien, 2011). This minimal design was in stark contrast with competing services at the time, where the trend was to allow users full customization of their personal page, often resulting in a cluttered and confusing design (Curran et al., 2011).

Unlike on other online SNSs, the relationship of following and being followed on Twitter requires no reciprocation. A user can follow any other user, and the user being followed need not follow back. Being a follower on Twitter means that the user receives all the messages from those he or she follows (Kwak, Lee, Park & Moon, 2010). A study of social interactions within Twitter reveals that the driver of usage is a sparse and hidden network of connections underlying the “declared” set of friends and followers (Huberman et al., 2009).

This unusual simplicity of Twitter—the users and their content—continues to warp perception of how the relationship between user and platform operates. Many of the popularized studies examining the influence on Twitter fail to identify the nuances of social interaction in the system (Leavitt, Burchard, Fisher, & Gilbert, 2009). The variety of uses of Twitter as more than just a status-update service is what makes Twitter both appealing and challenging to study. Users use Twitter for online activities such as sharing recommendations, tracking events, making new friends and the sharing of pictures (Vega et al., 2010).

The Twitter platform offers access to the vast collection of data via their public APIs. The Twitter API is an official online interface for a software developer to collect Twitter data, allowing high-throughput of near real-time access to various subsets of public Twitter data. Only non-protected public accounts can create public statuses. The collected data can be filtered in various ways, including, but not limited to, UserID, keyword, random sampling and geographic location (Twitter, 2011b).

---

\(^2\) In November 2017, Twitter expanded limits from 140 to 280 characters.
2.2.2 Overall Twitter Usage

The number of Twitter users has dramatically increased over the past years, with 30 million active users in Q1 2010 to now over 336 million active users in Q1 2018 (Statista, 2018a). These users provide over 500 million tweets per day (or 6,000 tweets per second), making it one of the largest microblogging service (InternetLiveStats, 2018). With its ongoing popularity and global mainstream media coverage, Twitter is growing in both usage and revenue when compared to Facebook (Investopedia, 2018).

While an active Twitter user may post several “tweets” in a single day, the median number of lifetime tweets is one and the average Twitter user posts just once every 74 days (Silverthorne, 2009). Following the established protocols of previous research, the following definition of user activity and retention was established (Java et al., 2007):

- A user is considered active during a week if he or she has posted at least one post during that week.
- An active user is considered retained for the given week if he or she reposts at least once in the following week.

In a 2009 conference speech, Peterson (2009) posited three types of “active” Twitter users:

- The Vast Majority of “Active” Users – with an average of 403 followers, 398 following and 44 updates per week (6.3 tweets per day)
- The Average “Active” Twitter User – with an average of 4,664 followers, 1,165 following and 108 updates per week (15.4 tweets per day)
- The Truly Exceptional “Active” Twitter User – with an average of 46,000 followers, 8,600 following and 567 updates per week (80+ tweets per day)

Each individual Twitter user sees updates from the people he or she follows on a timeline. The number of tweet updates depends on the number of followers and the tendency of those followers to tweet. If the number of followers is the higher side, the number of tweets that appear on the timeline could be overwhelming, thus causing an information overload (Vega et al., 2010).

In 2018, Twitter has become ubiquitous in daily communication as one of the most popular social media networks with a global reach of over 1 billion monthly visits to the site with roughly 82% of these visits from a mobile device, thus providing an opportunity to track geolocation and the time of each individual tweet (Cao et al., 2018). Demographics from the Pew Research Institute suggest that the average Twitter user is a college educated young white male in an urban setting (Pew Research Center, 2018).

2.2.3 User Intention – What People Tweet

It is important to understand why and how people use online social networks. By understanding these reasons, improvements to the overall structure of the network will occur (Java et al., 2007). Research
has focused on how Twitter has been used as a communication platform. From this work, researchers were able to derive standard metrics for measuring a user’s Twitter behavior, including, but not limited to, the measurement of the number of tweets, retweets, and followers (Vega et al., 2010).

Research suggests that usage patterns for Twitter are very different from other typical online social networks, as there is a small group of users who are very active. The top 10% of prolific Twitter users account for over 90% of total tweets. By comparison, in other typical online social networks, the top 10% of users account for 30% of all production. Twitter is more aligned with Wikipedia, where the top 15% of the most prolific editors account for 90% of Wikipedia’s edits. This research suggests that Twitter is used more like a one-way, one-to-many publishing service than a two-way, peer-to-peer communication network (Heil & Piskorski, 2009).

People use Twitter for different reasons. Java et al. (2007) identified four main user intentions on Twitter:

1. **Daily Chatter** – Most posts on Twitter talk about the user’s daily routine or what people are currently doing. This is the most common use of Twitter.

2. **Conversations** – Since there is no direct way for people to comment or reply to their friends’ posts, early adopters started using the @ symbol followed by username for replies. About one-eighth of all posts in the research contains a conversation, and this form of communication was used by almost 21% of users in the research.

3. **Sharing Information** – About 13% of all the posts in the research contained some URL in them. Due to the small character limit, a URL-shortening service is frequently used to make this feature feasible.

4. **Reporting News** – Many Twitter users report the latest news or comment about current events on Twitter. Some automated users or agents post updates like weather reports and new stories from RSS feeds.

A study by Kelly (2009) found similar results. The study categorised tweets into six user intentions:

1. **News** – Any sort of mainstream news that you might find on your national news stations such as CNN, Fox or others

2. **Spam** – Nonsense tweets such as “See how I got 3,000 followers in one day”

3. **Self-promotion** – Corporate tweets about products, services or “Twitter only” promos

4. **Pointless babble** – These are the “I am eating a sandwich now” tweets

5. **Conversational** – Tweets that go back and forth between folks, almost in an instant message fashion, as well as tweets that try to engage followers in conversation

6. **Pass-along value** – Any tweets with an “RT” in them (suggesting a Re-Tweet).
With regards to the number of tweets per category, pointless babble accounted for 40.55% of the total tweets captured, conversational was a close second at 37.55%, and the pass-along value was 8.7% of the tweets captured.

In the past decade, researchers have determined that those that tweet prefer a mobile platform versus a desktop website, with the most popular message type as an undirected message with no specific recipient and a difference between how a business tweets rather than an individual (Han, Hong, Lee, & Kim, 2017). In addition, users learn about news and events from what other users post, as they are exposed to news incidentally by their contacts, with some of these contacts being of influence (Halpern, Valenzuela, & Katz, 2017).

2.3 Physical Activity

Physical activity is a complex concept that can be classified qualitatively into major categories of locomotion, work, sedentary behaviours, leisure activities and exercise (Butte, Ekelund & Westerterp, 2012). Similar to the Butte definition, The World Health Organization (2010) defined physical activity as bodily movement produced by skeletal muscles that require energy expenditure. This includes not only traditional notions of physical activity during exercise, sport, and recreation, but also non-exercise activities including housework and social activities such as play and travel. In 2012, The Lancet presented its own definition of physical activity, suggesting physical activity is more than just running outside or exercising in a gym. Physical activity is about how we as humans relate to our own environment on a daily basis, using our bodies in the ways it was designed to move, no matter our movement nor our specific task, be it during work, play, transport or leisure (Das & Horton, 2012).

Physical activity has been assessed based on different age segments from child to adult to elderly, with special consideration for disabled persons. Males have been shown to be more physically active than females, with a steady decline in physical activity across age groups (Troiano et al., 2008). For the purposes of this dissertation, references to physical activity will be focused on the adult population.

In October 2008, the U.S. Department of Health and Human Services released the first comprehensive federal Physical Activity Guidelines for Americans, providing the general public with science-based guidance regarding the types and amounts of physical activity needed to maintain health and prevent disease (Pate, 2009). One of the main highlights from the report suggested that adults perform 150 minutes of physical activity per day. These Physical Activity Guidelines became the physical-activity objectives for Healthy People 2020, a multidisciplinary approach to promoting physical activity (United States Department of Health and Human Services, 2013).

In Ireland, The National Guidelines on Physical Activity was released by the Department of Health and Children and the Health Service Executive (HSE) in 2009. The Guidelines suggested that promoting physical activity complements existing national strategies to improve nutrition and to reduce tobacco,
Literature Review and Background

drug and alcohol use. It also positively impacts efforts to enhance social environments through reduced violence and improved social interaction and integration at a local and national level. The report stated that it was critical that despite the overwhelming evidence, much of the Irish population is not sufficiently active to reap the health gains associated with physical activity (Department of Health and Children and The Health Service Executive, 2009).

At all ages, the benefits of being physically active outweigh the potential harm (World Health Organization, 2010). It is well documented that regular physical activity is associated with reduced morbidity and mortality attributable to several chronic diseases such as diabetes, cardiovascular disease and some cancers (Blair, 1989).

The benefits of physical activity are far-reaching and extend beyond health alone. Being physically active is a major contributor to one’s overall physical and mental well-being, including improved sleep, a sense of purpose and value, quality of life, reduced stress, stronger relationships and social connectedness (Das & Horton, 2012). An active lifestyle does not require a regimented, vigorous exercise program. Instead, small changes that increase daily physical activity will enable individuals to reduce their risk of chronic disease and may contribute to enhanced quality of life (Pate et al., 1995). Besides the positive physical attributes of an active lifestyle, promoting active modes of travel, such as cycling and walking, is good for the environment, which in turn can have a positive impact on health (Das & Horton, 2012). Although no amount of physical activity can stop the biological aging process, there is evidence that regular exercise can minimize the physiological effects of an otherwise sedentary lifestyle and increase active life expectancy by limiting the development and progression of chronic disease and disabling conditions (Chodzko-Zajko et al., 2009).

Given the many benefits of a physically active lifestyle, several professional organizations and the vast majority of health and fitness practitioners strongly recommend that all healthy adults regularly engage in a sound exercise program. The key is to identify an exercise prescription that is results-oriented, time-efficient and safe (Bryant & Peterson, 1999).

2.3.1 The Rise of Physical Inactivity

The World Health Organization suggests that approximately 2.3 million people die each year because they are physically inactive. Those defined as inactive have a 20 to 30% increased risk of death compared to those who engage in at least 30 minutes of moderate physical activity on most days of the week (World Health Organization, 2014).

The risks of physical inactivity have been studied for years. But unlike other disease risk factors such as diet, tobacco and alcohol, the importance of physical activity has been slow to be recognized, and the emphasis to tackle it at a population level has not been forthcoming (Das & Horton, 2012). To that end, physical inactivity levels are rising in many countries, with major implications for the prevalence
of non-communicable diseases and the general health of the population worldwide. Physical inactivity is now identified as the fourth leading risk factor for global mortality at 6% of worldwide deaths, after high blood pressure (13%), tobacco use (9%) and high blood glucose (6%) (World Health Organization, 2014).

There has been a recent increase in research associated with the health impact of sedentary behaviour. Even when adults meet the physical-activity guidelines, sitting for prolonged periods can compromise their metabolic health and increase premature mortality risk (Owen & Healy, 2010). It is not uncommon for people to spend half of their waking day in a seated position with relatively idle muscles (Hamilton, Hamilton & Zderic, 2007). A proposal to journal editors was published in 2012 to adopt a consistent definition of sedentary as “…any waking behaviour characterized by an energy expenditure of less than or equal to 1.5 METs while in a sitting or reclining posture” (Tremblay, 2012). The same proposal suggested authors use the term “inactive” to describe those who are performing insufficient amounts of moderate- to vigorous-intensity physical activity.

2.3.2 Reasons for Physical Inactivity

One challenge is that physical activity is often perceived only in the context of controlling obesity, and therefore physical inactivity is regarded as a minor or secondary risk factor for non-communicable diseases (Das & Horton, 2012). Rates of physical inactivity vary tremendously across the globe, from Australia on the low end (15% of adults considered to be inactive), and Brazil on the high end (87% inactive), with the United States roughly in the middle (40% inactive). Unfortunately, even when adults do engage in physical activity, they often do so at levels far below the levels needed to elicit health improvement (Lox, Martin-Ginis, & Petruzzello, 2006).

As the availability of new technology has increased, physical labour and human energy expenditure have decreased (Haskell et al., 2007). The use of many of these technologies has been driven by the goal of increased individual worker productivity and reduced physical hardships and disabilities caused by jobs entailing continuous heavy labor but has in effect caused a lack of physical activity (Hallal, Andersen, Bull & Guthold, 2012). Over the course of one day, physical inactivity may induce negative effects on cellular processes in skeletal muscles or other tissues regulating risk factors, including plasma triglycerides and high-density lipoprotein (HDL) cholesterol (Hamilton et al., 2007). The human body has evolved in that most of its biological systems (muscle, skeletal, metabolic and cardiovascular) do not develop and function in an optimal way unless stimulated by frequent physical activity (Booth, Laye, Lees, Rector & Thyfault, 2008). Although the technological revolution has been of great benefit to many throughout the world, it has come at a major cost in terms of the contribution to global physical inactivity (World Health Organization, 2010).
2.3.3 Costs Associated with Physical Inactivity

Strong evidence shows that physical inactivity increases the risk of many adverse health conditions, including major non-communicable diseases such as coronary heart disease, type 2 diabetes, and breast and colon cancers, and shortens life expectancy. Because much of the world’s population is inactive, this link presents a major public health issue (Lee et al., 2012).

Worldwide, it is estimated that physical inactivity causes 6% of the burden of disease from coronary heart disease, 7% of type 2 diabetes, 10% of breast cancer and 10% of colon cancer. If inactivity were decreased by 10%, 533,000 deaths could be averted each year. This number could rise to 1.3 million if physical inactivity were decreased by 25%. If inactivity were eliminated, the life expectancy of the world’s population would increase by 0.68 years (Lee et al., 2012). Other epidemiologic studies have shown that low levels of physical activity and physical fitness are associated with markedly increased all-cause mortality rates. Even during midlife, an increase in physical activity is associated with a decreased risk of mortality (Pate et al., 1995).
2.4 Related Work
A literature review was conducted to evaluate previous research covering various topics, including mobile health applications, mobile health, use of Twitter in academic research, sentiment analysis and physical activity. The review crossed a number of disciplines, including exercise science, information technology, communication and health promotion.

A number of databases were used in the search for relevant academic published articles. These database searches were initiated from the James Hardiman Library at the National University of Ireland at Galway and included searches from ARAN, Scopus, and PubMed. Additional searches for relevant academic research were conducted through Google Scholar to identify research, then through the James Hardiman Library to acquire papers. Initial terms for search included but were not limited to, physical activity, mobile health, mobile fitness applications and Twitter. Additional search terms were included after review of the initial literature search for relevant topics.

Articles and papers for review were drawn from a number of sources, including peer-reviewed journals, conference presentations, conference papers, research by recognized independent institutions, textbooks, and company reports. The search terms to determine the literature review included but not limited to: “mobile fitness app”, “smartphone and fitness”, “mHealth”, “Twitter”, “Endomondo”, “MyFitnessPal”, “Nike+”, “RunKeeper”, “DailyMile”, “physical activity”, “physical activity tracking”. From these keywords, peer-reviewed articles were collected and evaluated. From the references to these papers, additional journal articles were found and reviewed. These articles included English-only publications.

2.5 How Social Networking Is Used to Share Personal Health Data
Survey evidence shows that people are using technology to support their desire for overall health. The 2013 Health Online Report by the Pew Research Group suggests that 81% of Americans use the internet, and of those, 72% are “online health seekers,” having looked online for health information of one kind or another within the past year. This percentage is up from 57% in 2009 (Fox & Duggan, 2013).

The same report suggests that one in three Americans are “online diagnosers,” people using the internet to self-diagnose a medical condition. When asked how they started their search for health information online, 77% of online health seekers said they began at a search engine, while 13% began at a site that specialized in health information (e.g., WebMD) (Fox & Duggan, 2013).

Recent data show that people with health conditions are turning to online social networks to build social ties with others who have similar health conditions (Pagoto, Schneider, Evans, et al., 2014). An online
phenomenon called “peer to peer healthcare” was discovered in 2011, where people seek counsel from fellow patients and/or caregivers, a trend that shows a new model of personal health research. The pursuit of health information is now taking place within a widening network of both online and offline sources (Fox, 2011).

Specific to Americans who used the internet in 2012, 24% reported that they turned to others who had the same health condition, 26% have read or watched online content about someone else’s experience with a medical or health issue and 16% have gone online to find others who might share the same health or medical condition (Fox & Duggan, 2013).

2.5.1 Social Networking and Physical Activity
People interact with their social network with regards to their health. Christakis & Fowler (2009) concluded that “… a person with more friends and social contacts generally has better health than a person with fewer friends, and a person at the centre of a network is more susceptible to both the benefits and risks of social connection than those at the periphery of a network” (p.5). Whereas in the past people may have called a health practitioner, they now are also searching the Internet for health information, posting questions to their Facebook account and listening to shared health and wellness podcasts (Fox & Jones, 2009).

The social life of health information is robust. Fox and Jones (2009) reported that 57% of survey respondents reported having looked for health information online, with nearly 66% sharing that information with someone else. The same report suggests that 60% of the online information affected a health decision, while 49% said it changed the way they think about diet and exercise. Only 3% of all adults surveyed said they knew someone that was harmed by following online health information (a finding that has remained stable since 2006) (Fox & Jones, 2009). Research suggests that by identifying and understanding how a person’s social environment impacts physical activity and exercise, policymakers could fund and implement interventions that promote and sustain physical activity and developers could make products and services that could lead to the prevention of a wide array of diseases and better overall health (Almeida, 2008).

Wireless connections are associated with deeper engagements in social networks and the accelerated pace of information exchange (Fox & Jones, 2009). In addition, younger generations (ages between 18 and 49) are more likely than their older counterparts to participate in social technologies related to health and fitness. As these younger generations face more health concerns (for themselves and those within their social network, including family and friends), they may turn to the tools they have mastered in other areas of their lives to gather and share health and fitness advice (Fox & Jones, 2009).
2.5.2 The Influence of Those Who Tweet

The Merriam-Webster dictionary defines influence as “the power or capacity of causing an effect via indirect or intangible ways” (Merriam-Webster, 2012a). Each Twitter user profile serves as the foundation of communities in which users regularly meet, talk, provide support and help each other. In other words, they exhibit communal characteristics of a community in that users display a sense of belonging and have the ability to influence each other through their replies and retweeting (Quercia, Ellis, Capra & Crowcroft, 2011). Despite research and a large number of theories of influence in sociology, there is not yet a tangible way to measure such a force, nor is there a concrete definition of what influence means (Cha, Haddai, Benevenuto & Gummadi, 2010). Scoring of online influence is subjective and is for now considered imperfect. Most analytics companies rely heavily on a user’s Twitter and Facebook profiles, leaving out other online activities, like blogging or posting YouTube videos, and do not take into consideration a user’s influence in the offline world (Rosenbloom, 2011).

Traditional communication theory states that a small minority of users, called influentials, are exceptional in the persuasion of others (Watts & Dodds, 2007). This theory predicts that by targeting the influentials in the network, one may achieve a large-scale chain reaction of influence driven by word-of-mouth, with a very small marketing cost (Katz & Lazarsfeld, 1956). With the advent of Twitter, influence can be measured more easily since tweets and messages can be traced back to their original source (Cha et al., 2010). This more modern view de-emphasizes the role of influentials and posits that the key factors determining influence are (1) the interpersonal relationship among ordinary users and (2) the readiness of society to adopt an innovation (Domingos & Richardson, 2001; Watts & Dodds, 2007).

Directed links in social media could represent anything from friendships to common interests, and as such these directed links can determine the flow of information and hence indicate a user’s influence on others—a concept that is crucial in sociology and viral marketing (Cha et al., 2010). However, when the focus of influence is solely on the connections between users, there is a misconception regarding the number of followers and influence:

“A popular metric of perceived influence on Twitter measures the number of a user’s followers. In general, the more followers a user possesses, the more impact he appears to make in the Twitter environment, because he seems more popular (namely, that users follow him). This statement makes sense assuming that Twitter acts as a successful broadcast medium, where a user publishes a tweet, and it is read by every follower. However, this view of Twitter as a broadcast medium ignores the potential for users to interact with the content on the platform.”

Leavitt et al. (2009)

Cha et al. (2010) reported that studying influence patterns can help one better understand why certain trends or innovations are adopted faster than others. Studying influence patterns, however, has been difficult because such a study does not lend itself to readily available quantification, and essential
components like human choices and the ways societies function cannot be reproduced within the confines of the lab and/or study. More recent research found that influentials are highly active users and consequently defined a new influence measure based on user activity. This measure accurately predicts URL clicks (hence influence) on Twitter, suggesting that influence on Twitter is not gained accidentally but strongly depends on audience engagement. No study has yet established what type of engagement translates into influence (Quercia et al., 2011).

Quercia et al. (2011) suggest that the specific use of vocabulary and prescribed ways of communicating is what links social influence with others on Twitter. Expressing a sense of community correlates with influence while expressing negative emotions reflects one’s mood, which in turn impacts one’s influence. As a result, a theoretical implication is that Twitter is a distal communication modality (users are separated in space and time) that was originally designed not as a social networking tool, but rather as a broadcasting platform of publicly available news and opinions. However, insights from research suggest that the medium partly resembles proximal communication between individuals embedded in offline social networks: Influence is not gained spontaneously, but partly depends on linguistic qualities that reflect one’s personality and mood.

As of 2013, there were a very small number of Twitter users who follow more than 10,000 others. There are only 101 users with more than 2 million followers, most of whom are either celebrities or mass media outlets, with most of the top Twitter users not following back those users who follow them (TwitterCounter, 2013). In 2018, the top three Twitter users in terms of followers over 100,000,000 included: @katyperry (109.53M), @justinbieber (106.49M), @barakobama (102.97M) (Statista, 2018b).

2.5.3 Measurement of Physical Activity

Reliable and accurate measurement of physical activity remains challenging for researchers, exercise scientists, and epidemiologists, with current measurement techniques falling into three general categories: subjective reporting, direct observations and use of portable monitors (Troiano et al., 2008). Each of these methods has its own known limitations (Lee, Macfarlane, Lam & Stewart, 2011).

In many countries, including the United States, physical-activity data are collected for national health surveys through self-reporting (Troiano et al., 2008). However, this type of self-reported physical-activity measurement has the inherent limitation of recall bias (Hallal et al., 2012). Other research suggests the correlation between objective measures of physical activity and surveys such as the International Physical Activity Questionnaire Short Form (IPAQ-SF) in the large majority of studies was lower than the acceptable standard and that the use of IPAQ-SF typically overestimated physical activity (Lee et al., 2011).
In the first part of this century, the use of objective motion sensors was not considered practical for large-scale studies because of high cost, uncertain reliability, and difficulty in data interpretation (Wood, 2000). However, with enhancements in their objective measurement, data-storage capacity and small and unobtrusive size, motion sensors (such as accelerometers) have grown in popularity (Troiano et al., 2008). The use of online social networking as a possible method to collect publicly available physical-activity data is promising for both small research studies following specific individuals as well as large-scale observational research across activity levels, location, and experience. Two prominent challenges exist. First, researchers must properly set up the data collection from devices or apps and, second, researchers must determine what percentage of the population use Twitter (or whatever online social network is being used) and the device or app being used in an observational research study. Future work in using online social networking for physical activity data collection can also analyse common terms used when discussing physical activity and thus create a specific database for these terms to collect.

There is ample evidence to show that physical inactivity is a major contributor to death and disability from non-communicable diseases worldwide (Das & Horton, 2012). At the same time, mobile communication technologies are widely available, with a majority of American adults owning a mobile phone (Fox & Duggan, 2013). To reach the targets for increasing physical activity levels and improving health in the United States as set forth by Healthy People 2020, a multidisciplinary approach is critical. One possible approach involves using mobile technology and social networking.

2.6 Twitter Data Mining and Text Classification

Data mining is a relatively young and interdisciplinary field of computer science. The process results in the discovery of new patterns in large data sets by using methods at the intersection of artificial intelligence, machine learning, statistics and database systems. The overall goal is to extract knowledge from an existing data set and transform it into a human-understandable structure for further use (Chakrabarti, Ester, Fayyad, & Gehrke, 2006).

Text classification is the labelling of natural language texts into one or more categories drawn from a predefined set. This may be done manually or algorithmically. For the purposes of this research, the text in question is a tweet. Text classification is one of the most important research fields in information retrieval and data mining. Its solutions are at the core of several technology applications, ranging from the automatic cataloguing of newspaper pages and web pages to the management of incoming emails, from the annotation of DNA genome sequences to sentiment analysis of tweets (Vitale, Ferragina & Scaiella, 2012). By tapping into the world’s collective brain, researchers have found that if they make an effort to dig through the mundane comments, these live conversations offer an early glimpse into public sentiment and activity, and perhaps can even help shape it (Miller, 2009).
Literature Review and Background

Constructing a classification model is a perplexing task, as tweets by their very nature are short, with a maximum of just 140 characters. Additionally, to allow for these short messages, users often use a specific language, with word fragmentation, incorrect grammar and specific abbreviations, which can be difficult for a computer to interpret (Yerva, Miklos & Aberer, 2012). Current Twitter classifications have been designed on a macro-level timeline at the expense of the richness of depth from individual histories and shared experiences (Dann, 2010). During the course of this research, it became apparent that while tweet-classification models existed in other areas of research, no such model existed in the health and fitness areas of academic study—thus the Fitness Tweet Classification Model was created.

From the data-collection and data-processing tools as defined in this research, we have been able to create a growing dataset of publicly shared information via Twitter that provides a wealth of data about a person’s physical activity that includes, but is not limited to, exercise type, length, day of the week, geographical location and time. This combined information will allow research on how technology can be used to monitor and possibly motivate individuals, the exercise habits of those who share their workout information via mobile fitness apps via Twitter and how a person’s social network influences his or her fitness activities.

The classification model created for this Twitter and fitness research was based on the Cheong and Ray (2011) model. It has since been backed by additional research by Cormode, Krishnamurthy, and Willinger (2010), who studied the modelling and measurement of online social media and identified two central objects in Twitter—the user and the tweet itself. In this research, findings from the state-of-the-art Twitter and fitness research will relate to either one or both of these domains.

2.7 Klout

As suggested in this literature review, the measurement of one’s influence on Twitter can be difficult to accurately measure. Online influence measurement services feed public data from sites such as Twitter, LinkedIn, and Facebook into secret formulas that generate scores that gauge users’ influence comparable to “the credit score of friendship” or “the S&P of social relationships” (Miller, 2011). These services are in the process of scoring millions, eventually billions, of people on their level of influence. To proponents, the measurement of online influence is an inspiring tool that is encouraging the democratization of influence where one no longer must be a celebrity, a politician or a media personality to be considered influential. However, to critics, social scoring is a brave new techno-measure where one’s rating could help determine how well they are treated by everyone with whom they interact (Rosenbloom, 2011).

For the purposes of this research, online influence was defined as the ability to drive a person to action, where that action could be defined as a reply, a retweet, a comment, or in a broader sense, the impetus of the recipient of the influence to be physically active. To be able to quantify the influence, it was
determined that a numerical influence score was needed. After a review of a number of different online measures, the third-party tool Klout was used to measure online influence.

The Klout Score is a factor of over 35 variables broken into three categories; True Reach, Amplification Score and Network Score (Klout, 2011). Klout scores measure the “size and strength of a person’s sphere of influence” online. Rather than simply count up how many tweets each member has posted, Klout scores reflect each member’s “True Reach” (actual audience size), “Amplification Ability” (number of posts that get replied to or discussed by other Twitter users), and “Network Score” (where a member’s Klout score is boosted if that member’s readers also have high Klout scores). Klout scores range from 0 to a possible 100, although scores above 50 are rare (Lassen & Brown, 2011). A Klout score fully ignores a user’s number of followers and number of tweets but rather the extent to which the user’s content is retweeted (Quercia et al., 2011).

Klout has classified influencers into 16 different categories based on their communication style, audience, and engagement. A person’s Klout Class is like a personality test for their online influence. Users of the same class might not share the same topics, Klout score, or audience, but they share a similar communication style (Klout, 2011).

2.8 Summary
This chapter introduced the combination of online social networking and physical activity and concluded with the concept of data mining and text classification using Twitter. Besides the already-popular research on the exploration of Twitter by groups led by Huberman (2009) and Java (2007), several new research endeavours have been established that enhance the knowledge of this subject, from the combined perspective of both Twitter users and messages. It is from this established base of Twitter research that the Fitness Tweet Classification Model was built.
Data Collection and Processing

3 DATA COLLECTION AND PROCESSING

This chapter introduces the concept of content analysis and explains how the major classifications were created. In addition, the criteria used for the selection of the five mobile fitness apps used in this research is detailed. Furthermore, the chapter introduces the data collection and processing tools created for this research, the Fitness Tweet Crawler and Fitness Tweet Classification Model, respectively. It is through the Fitness Tweet Classification Model that this work contributes to future research, as the continued improvement in the methods used to monitor physical activity can help guide the development of new and improved programs and policies to increase physical activity and thereby reduce the incidence of non-communicable diseases in the 21st century (Hallal et al., 2012).

One of the objectives of this research was to develop a method for collecting physical-activity data using Twitter. While there is consistently a tremendous number of tweets that contain words or phrases associated with exercise or physical activity, a more structured approach to collecting them was desired. Yin (2002) defines data analysis as the process of examining, categorizing, charting or otherwise amalgamating the evidence in order to address the initial intentions of the study. Because qualitative data collection typically takes the form of large amounts of unstructured textual materials, analysis of the data is not always straightforward. Therefore, clear-cut rules about how this data should be analysed needed to be developed (Bryman & Bell, 2007).

In May 2013, iTunes offered 23,490 smartphone applications that were categorised as Health and Fitness (Middelweerd et al., 2014), an increase of 95% from August 2012 (Dolan, 2011). In 2018, that number doubled to over 47,911 smartphone applications (Statistica, 2018).

The advent of smartphones has greatly enlarged both the number and reach of mobile apps for health purposes by providing a platform for developers to design third-party applications (apps), which expand the functionality and utility of these mobile devices (West et al., 2012). In addition to allowing users to track their fitness activities via a global positioning satellite (GPS) from their smartphone, these mobile applications allow the immediate sharing of a workout with friends and family members who make up one’s online community through a website hosted by the app company or by third-party social networks such as Facebook or Twitter.

With hope comes caution, as the speed to market of mobile fitness apps may outpace the desire to create a well-rounded, scientifically based technology. Given the current status of the development of mobile fitness apps and the rates of innovation for research, technology, and theory within academic research, it may be challenging for researchers to evaluate the safe and effective methods of increasing physical activity by using mobile fitness apps (Conroy, Yang, & Maher, 2014). The vast majority of commercial apps have not been evaluated using scientific methods (Pagoto & Bennett, 2013). In addition, many of
the existing commercial mobile fitness apps have been developed without a keen understanding of theories of health behaviour change, as the developers of these apps come from a variety of backgrounds and many are not trained in the application of health-behaviour theory (Cowan et al., 2013). One of the results of the literature review suggested that at the time of this research, while other researchers outside of health promotion and exercise science were creating and using models to collect and analyse a wide range of topics shared over online social networks, the adoption of such data-collection models within health promotion and exercise science was limited.

3.1 Content Analysis

Content analysis is a highly flexible research method that facilitates the examination of written and oral communication with varying research goals and objectives. It can be applied in quantitative, quantitative and sometimes mixed modes of research frameworks to employ a wide range of analytical techniques to generate findings and then put them into context (White & Marsh, 2006; Insch, Moore & Murphy, 1997).

For the purposes of this research, the content-analysis procedure was based on a synthesis of recommendations regarding the use of content analysis as described by Insch et al. (1997). This procedure had eight steps, as highlighted in Figure 3-1. The steps listed in blue form the established procedure, while the process in green maps each step conducted for this research, matching them to the established procedure.
The data for the content analysis in this research were the 2,856,534 tweets from the five mobile fitness apps (Step 1). The tweets from these mobile fitness apps were chosen because of the popularity of each app and the understanding that these fitness tweets implied the users desired to share their workouts on Twitter. These tweets were identified by searching Twitter for tweets containing specific hashtags, which are an online social networking tool originally innovated for the purpose of information organization and management, and which have gained many expressive functionalities over time (Shapp, 2014). For the purposes of this research, the organization of information was the use of hashtags from each of the five mobile fitness apps.

The specified unit of measure for the analysis was English-only tweets, allowing for 1,982,653 tweets for analysis (Step 2). This was accomplished using the demographic information of the Twitter data. Twitter identifies the language and assigns a two-letter code for each tweet, and only EN (i.e., English) tags were used for analysis. Any tweets that were miscoded by Twitter were manually assigned as non-English.
Data Collection and Processing

A review of the literature on Twitter classification suggested that there were no established criteria for fitness tweet categorization (Vickey, Breslin, Ginis & Dabrowski, 2013). Therefore, criteria for the content analysis were created using prior literature (Dann, 2010) and sourced from the research cited below to establish broad categories of content.

<table>
<thead>
<tr>
<th>Conversational</th>
<th>Jansen &amp; Zhang, 2009</th>
<th>Kelly, 2009</th>
<th>Naaman et al., 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>“conversations”</td>
<td>Information Seeking</td>
<td>Conversational</td>
<td>Questions to Followers</td>
</tr>
<tr>
<td>Pass Along</td>
<td>“information sharing”</td>
<td>Information Providing</td>
<td>Pass-along or Self-Promotion</td>
</tr>
<tr>
<td>News</td>
<td>“reporting”</td>
<td>Information Providing</td>
<td>News</td>
</tr>
<tr>
<td>Status</td>
<td>“daily chatter”</td>
<td>Comment or Sentiment</td>
<td>Pointless Babble</td>
</tr>
<tr>
<td>Spam</td>
<td></td>
<td></td>
<td>Spam</td>
</tr>
</tbody>
</table>

Table 3-1 - Previous Classification Research

This criterion provided a guide for a manual review of the initial English-only tweets was conducted by the researcher. During this manual review, various categories emerged from the data. These categories were not predetermined, but rather inferred from the data as each tweet was assigned to the most appropriate category. If a tweet did not fit into one of the established inferred categories, it was coded as an unknown category for later review. This method is unlike Insch and Moore (1997), where unclassified words or phrases were dropped from the analysis. A total of 500 English-only tweets were manually processed, 100 from each mobile fitness app (Step 3). These tweets were collected over a two-week period. The counts of occurrences at this initial level of analysis provided raw frequencies. However, the true value added by the content analysis is the classification of units into categories, as the categories defined for the study will arise from previous classification models and data analysis (Insch & Moore, 1997). Each tweet was only counted once in a single category. The dataset (N=500) was sorted and evaluated manually by the researcher to determine any apparent similarities. This evaluation allowed the algorithms to be established. This dataset became the study standard and referred to in the remaining analysis as study standard.

After the initial analysis, two different categories emerged from the data (Step 4). The first was specific to fitness-related tweets. There were a large number of tweets that described a person’s exercise. Based on the similar word structure of the tweets, it was apparent that they were auto-generated by the mobile fitness app. These tweets were categorized as “Activity.” The second type was similar to previous Twitter research by Java et al. (2007) and Naaman et al. (2010), in that the user was using the hashtag (either from within the app or on Twitter itself) to have a conversation. These tweets were categorized
as “Conversation.” From this step, a sample coding scheme was created, based upon the evaluation of the initial data set of 500 random tweets (Step 5). This scheme became the Text Classification algorithm. Samples of the types of tweets collected can be found in Chapter 4.4.

### 3.1.1 Determination of Major Classifications

Additional tweets were then submitted to a computerized Text Classification algorithm to identify Activity and Conversation tweets. This procedure revealed subcategories within the Activity and Conversation groupings, as well as a third category, subsequently labelled “Blarney.” Specifically, further analysis of the Activity tweets showed that some users added additional messages along with the information about their actual workout (e.g., I just ran 4 mi using #RunKeeper in the sunshine of San Diego, felt great); thus, the “Workout+” subcategory was added. A Workout+ tweet has the same foundation of a Workout tweet but has the additional variable of information.

Further analysis of the Conversation category indicated tweets pertaining to four areas: requests for technical support (requests to the app company or the broader community); marketing (e.g., press releases or updates, which came from the app company itself or the community); statements of support (where people within the app community congratulated others on reaching milestones, personal bests, etc.); or information sharing (e.g., those within the app community who wanted to run together in an upcoming 10K race would post messages using the hashtag per the app). Thus, the following subcategories were added to the Conversation category: Technical Support, Marketing, Statements of Support and Information Sharing.

In addition, a third category was added (Blarney) that tagged spam tweets (tweets with only a URL) or tweets that had little relevance to exercise (e.g., Test FB http://t.co/IKIQjTi #myfitnesspal). Blarney is defined as skillful flattery, nonsense or blandishment (Merriam-Webster, 2012b). Tweets that were classified as Blarney were further classified into Pointless Babble or Spam. Any tweet that contained just a URL or appeared to be an unsolicited commercial tweet was classified as Spam, while all other tweets classified as Blarney were sub-classified as Pointless Babble (Step 6). This phase is crucial to developing a robust coding scheme, as it can highlight interpretation issues, the realism of decision rules, and the appropriateness of text selection (Insch et al., 1997). These resulting categories are similar to categories derived by Honey and Herring (2009) and Naaman et al. (2010).

### 3.1.2 Testing of Method to Independent Review Coders

After the creation of the Text Classification algorithm, to examine the strength of the algorithm, four review coders were given the same 500 tweets previously used and were asked to manually classify them according to the coding scheme. This step ensured that the classification rules were applied correctly and assessed the semantic validity. Weber (1990) defines semantic validity as the extent to which others familiar with the language (in this case English) examine lists of words (in this case a
Data Collection and Processing

tweet) placed in the same category and agree they have similar meanings or relate to the category in a similar fashion.

3.1.3 Percentage Agreement

Whilst there are several calculations to measure inter-rater agreement for qualitative items to measure validity, including but not limited to Cohen’s kappa coefficient, the percentage agreement method was used. Percentage agreement, also called simple agreement, percentage of agreement, crude agreement or raw percent agreement, is the percentage of all coding decisions made by coders on which they agree. The use of percentage agreement has numerous advantages, such as being simple, easy to calculate and intuitive (Lombard, Snyder-Duch & Campanella, 2002). To calculate the percentage agreement, the total number of times in which the raters agree is divided by the total number of classifications. The percentage agreement did not account for chance. The percentage agreement between independent raters and the study standard (TV) on both major and minor categorizations was calculated to determine which category system was more reliable. Each reviewer was asked to provide a classification using a descriptive number (not a math number) between 1 and 8 (1 = Workout, 2 = Workout+, 3 = Pointless Babble, 4 = Spam, 5 = Technical Support, 6 = Corporate Marketing, 7 = Statements of Support, 8 = Information Sharing). These categorical scores used nominal numbers for informational purposes only, as the numerical value was irrelevant and did not indicate quantity or rank. After submission of the scores, numeric categories 1 and 2 were recoded as Category A, while numbers 3 and 4 were placed into Category B, and numbers 5, 6, 7 and 8 were recoded as Category C. This allowed for a percentage agreement scores for both major and minor categorizations.

The analysis found that overall, raters agreed with the study standard on 76.55% for the minor category and 88.00% for the major category of the total tweets in the data set (Table 3-2). The field of study determines the percentage acceptability, however for most, a score over 75% is considered acceptable (Wolbert, 2018).

<table>
<thead>
<tr>
<th>Overall</th>
<th>Rater 1</th>
<th>Rater 2</th>
<th>Rater 3</th>
<th>Rater 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># Agree</td>
<td>Percentage</td>
<td># Agree</td>
<td>Percentage</td>
</tr>
<tr>
<td>Minor (1-8)</td>
<td>1531</td>
<td>76.55%</td>
<td>361</td>
<td>72.20%</td>
</tr>
<tr>
<td>Major (A, B, C)</td>
<td>1760</td>
<td>88.00%</td>
<td>457</td>
<td>91.40%</td>
</tr>
</tbody>
</table>

Table 3-2 - Percent Agreement with Original Scale among Four Independent Review Coders

This evidence suggests that the minor categorization within the Fitness Tweet Classification Model is usable but can suffer from misinterpretation and wider variation due to eight choices instead of three. The major categorization into Activity, Blarney or Conversation reduces the model into theoretical concepts that can be more easily separated and explained.
After establishing that the percent agreement was sufficient, the analysis code used in the percent agreement analysis was finalized into a Java code for computerized classification of the tweets, which became the Fitness Tweet Classification Model. Once finalized, the entire tweet dataset was analysed through the model (Step 8). Three manual verification checks were conducted randomly, which allowed misclassified tweets to be corrected. Enhancements to the algorithms and reclassification of the entire database were conducted as needed.

### 3.2 Selection Criteria for Mobile Fitness Apps

As the goal for this research was to collect large amounts of data without regard for the scientific functionality of the app from which the data was sent, a computer science-based selection process was used to identify the mobile fitness apps that would be used in this research. Conclusions from Breton, Fuemmeler, and Abroms (2011) suggested additional research is needed to develop, improve and evaluate mobile apps specific to health promotion. One way to evaluate and improve these types of apps is through the data analysis that this research has achieved.

The mobile fitness apps used for this research were available for download for free from the iTunes store in December of 2010. Only apps that were designed for an iPhone were selected because it was determined that most, if not all, activities from which physical-activity data would be reported through each app would most likely be from an iPhone due to the physical size restrictions.

A ranked list of the top 200 free apps available from the Health and Fitness category from the United States iTunes store was downloaded. From that list, only those apps that were specific to fitness were selected (N=56). From those fitness apps, only those fitness apps that tracked physical activity were selected (N=23). From those fitness apps, only those that had the ability to share workouts via Twitter were selected (N=11) (Figure 3-2). The final list of 11 apps that met all of the aforementioned criteria was then ranked according to popularity (Table 3-3). The 2018 rankings are provided to show change over time (SimilarWeb, 2018).

<table>
<thead>
<tr>
<th>App</th>
<th>Rank 2010</th>
<th>Rank 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nike</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>MyFitnessPal</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>RunKeeper</td>
<td>3</td>
<td>44</td>
</tr>
<tr>
<td>DigiFit</td>
<td>4</td>
<td>50+</td>
</tr>
<tr>
<td>MapMyRun</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>Endomondo</td>
<td>6</td>
<td>50+</td>
</tr>
<tr>
<td>miCoach</td>
<td>7</td>
<td>50+</td>
</tr>
</tbody>
</table>
The same process was used for the UK Health and Fitness apps available as of December 2010. From the initial list of 200 Health and Fitness free apps, only those apps that were specific to fitness were selected (N=46). From those fitness apps, only those fitness apps that tracked physical activity were selected (N=17). From those apps, only those that had the ability to share workouts via Twitter were selected (N=13) (Figure 3-3). The final list of 13 apps that met all of the aforementioned criteria was then ranked according to popularity (Table 3-4). The 2018 rankings are provided to show change over time (SimilarWeb, 2018).

<table>
<thead>
<tr>
<th>App</th>
<th>Rank 2010</th>
<th>Rank 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>MyFitnessPal</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>RunKeeper</td>
<td>2</td>
<td>50+</td>
</tr>
<tr>
<td>MapMyRun</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>MapMyRIDE</td>
<td>4</td>
<td>50+</td>
</tr>
<tr>
<td>miCoach</td>
<td>5</td>
<td>50+</td>
</tr>
<tr>
<td>Nike</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>CycleMeter</td>
<td>7</td>
<td>50+</td>
</tr>
</tbody>
</table>
At the time of this analysis, there was no way to determine the actual number of downloads from the iTunes store. To understand possible variances in the popularity of each downloaded app from iTunes, the total number of downloads of each app was determined from the Android store in January 2011. It was found that the most popular Android mobile fitness app was Endomondo with 53,491 downloads, followed by MyFitnessPal with 18,693 downloads and RunKeeper with 12,427 downloads. The 2018 rankings are provided to show change over time (SimilarWeb, 2018). The total downloads for the remaining seven mobile fitness apps were not available, but like in the iTunes store, the ranking was available (Table 3-5).

<table>
<thead>
<tr>
<th>App</th>
<th>Rank 2010</th>
<th>Rank 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endomondo</td>
<td>1</td>
<td>50+</td>
</tr>
<tr>
<td>MyFitnessPal</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>RunKeeper</td>
<td>3</td>
<td>50+</td>
</tr>
<tr>
<td>miCoach</td>
<td>4</td>
<td>50+</td>
</tr>
<tr>
<td>MapMyRun</td>
<td>5</td>
<td>50+</td>
</tr>
<tr>
<td>WalkMeter</td>
<td>6</td>
<td>50+</td>
</tr>
<tr>
<td>Runmeter</td>
<td>7</td>
<td>50+</td>
</tr>
<tr>
<td>Log Your Run</td>
<td>8</td>
<td>50+</td>
</tr>
<tr>
<td>MapMyFitness</td>
<td>9</td>
<td>50+</td>
</tr>
<tr>
<td>Cyclemeter</td>
<td>10</td>
<td>50+</td>
</tr>
</tbody>
</table>

Table 3-5 - Popularity Ranking of Android Fitness Apps, December 2010 and July 2018

A combined ranking system taking the average of the popularity of each app by downloads in 2010 allowed for the following overall ranking (Table 3-6).
Data Collection and Processing

<table>
<thead>
<tr>
<th>App</th>
<th>USA IOS</th>
<th>UK IOS</th>
<th>Android</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nike</td>
<td>1</td>
<td>1</td>
<td>N/A</td>
<td>1.00</td>
</tr>
<tr>
<td>MyFitnessPal</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2.00</td>
</tr>
<tr>
<td>Endomondo</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>3.33</td>
</tr>
<tr>
<td>RunKeeper</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>3.67</td>
</tr>
<tr>
<td>DigiFit</td>
<td>4</td>
<td>N/A</td>
<td>N/A</td>
<td>4.00</td>
</tr>
<tr>
<td>miCoach</td>
<td>7</td>
<td>8</td>
<td>4</td>
<td>6.33</td>
</tr>
<tr>
<td>MapMyRun</td>
<td>5</td>
<td>10</td>
<td>5</td>
<td>6.67</td>
</tr>
<tr>
<td>CycleMeter</td>
<td>8</td>
<td>4</td>
<td>10</td>
<td>7.33</td>
</tr>
<tr>
<td>RunMeter</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>7.67</td>
</tr>
<tr>
<td>WalkMeter</td>
<td>10</td>
<td>9</td>
<td>6</td>
<td>8.33</td>
</tr>
<tr>
<td>MapMyFitness</td>
<td>11</td>
<td>6</td>
<td>9</td>
<td>8.67</td>
</tr>
<tr>
<td>LogYourRun Free</td>
<td>N/A</td>
<td>13</td>
<td>N/A</td>
<td>13.00</td>
</tr>
</tbody>
</table>

Table 3-6 - Combined Popularity Ranking of Mobile Fitness Apps, December 2010

While the overall popularity ranking for each mobile fitness app was established, it was unclear if the users of each mobile fitness app actually used the sharing feature of the app on Twitter. Therefore, each of these apps was included in a beta test for collecting hashtags. This beta test of data collection started on March 11, 2011 and continued through April 1, 2011. At the end of the beta test, the data in Table 3-7 had been collected.

<table>
<thead>
<tr>
<th>App</th>
<th>Hashtag Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>RunKeeper</td>
<td>125,315</td>
</tr>
<tr>
<td>Nike</td>
<td>90,575</td>
</tr>
<tr>
<td>DailyMile</td>
<td>54,220</td>
</tr>
<tr>
<td>Endomondo</td>
<td>42,447</td>
</tr>
<tr>
<td>MyFitnessPal</td>
<td>36,724</td>
</tr>
<tr>
<td>MapMyRun</td>
<td>16,547</td>
</tr>
<tr>
<td>MapMyFitness</td>
<td>13,560</td>
</tr>
<tr>
<td>miCoach</td>
<td>1,348</td>
</tr>
<tr>
<td>DigiFit</td>
<td>298</td>
</tr>
<tr>
<td>CycleMeter</td>
<td>0</td>
</tr>
<tr>
<td>RunMeter</td>
<td>0</td>
</tr>
<tr>
<td>WalkMeter</td>
<td>0</td>
</tr>
<tr>
<td>LogYourRun</td>
<td>0</td>
</tr>
</tbody>
</table>
Since the foundation of this research involved massive amounts of shared workout data, and as a result of the beta test of data collection, the top four mobile fitness apps—RunKeeper, Nike, Endomondo, and MyFitnessPal—were chosen for final data collection.

During the course of the beta data collection, it was determined that CycleMeter, RunMeter, WalkMeter and 20 other mobile fitness apps were able to sync with DailyMile, at which time the workout tweet was sent from DailyMile using the #dailymile hashtag. For that reason, DailyMile was added as the fifth mobile fitness app used in the research.

### 3.2.1 Endomondo

Described as a personal athletics tracker, Endomondo is a free mobile/GPS-powered Sports Tracker app that runs on multiple platforms, including iPhone, Android and Garmin watches (Endomondo, 2012). Endomondo can be used for distance-based activities and utilises both GPS and Microsoft Bing interactive maps to track routes, distance, duration, split times and calorie consumption, while also providing audio feedback on performance (Figure 3-4). What makes Endomondo stand out from other fitness-tracking apps is its integrated social feature. By incorporating elements found in leading social networks, Endomondo helps users find the motivation to both become active and stay active. Users can send real-time pep talks to friends while they are exercising, compete against friends, challenge co-workers and share it all on Facebook or Twitter. Endomondo also has an online community to share tips and provide motivation for users.
As of October 2010, the mobile fitness app had more than 1 million downloads, with 500,000 registered users. The application had growth from 40,000 registered users in January 2010 to 100,000 in April 2010, a doubling of its user base every 10 weeks (O’Hear, 2010).

In addition to the basic tracking of a workout route, split times, calorie consumption and challenges, Endomondo provides the user with an audio coach. For each mile or kilometre traveled, a voice will inform the user about distance and speed. In addition, the app enables friends to follow the user’s run in real-time from their computers, from which they can send messages of encouragement that are converted to audio and played during the workout (Endomondo, 2012).

3.2.2 RunKeeper

RunKeeper is a mobile fitness app that “makes tracking your workouts fun, social, and easy to understand so that you can improve the quality of your fitness” (RunKeeper, 2012). Like Endomondo, RunKeeper uses the smartphone GPS to track time, route, distance, pace and elevation of exercise sessions and has an online community (Figures 3-5 and 3-6). The app allows for data sync between the user’s mobile device and the RunKeeper website, as well as for sharing on social networks such as Facebook and Twitter. As of June 2011, RunKeeper has an online community of 6 million fitness enthusiasts, and in addition to the iPhone app, RunKeeper is available on Android and Windows Phone 7 platforms (Jacobs, 2011).
Data Collection and Processing

Figure 3-5 - RunKeeper Screenshots

Figure 3-6 - RunKeeper Member Page (Web)
In December 2010, RunKeeper changed their business model to allow for the app to be available for free. This type of free promotion strategy is common, with a goal of increasing the total number of downloads. This increases the likelihood that the app appears in the top-selling lists of the Apple App Store, where the increased visibility will allow for additional downloads over the long term (Ha, 2010). As of December 1, 2010, the RunKeeper app was downloaded from the Apple App store over 171,000 times, a tenfold increase from its distribution up to that point (Ha, 2010).

3.2.3 Nike+

Nike+ is a mobile fitness app developed by global shoemaker Nike. Like Endomondo and RunKeeper, Nike+ “…allows a user to measure distance, pace, map runs, track progress and get the motivation needed to go even further” and provides an online support community (Figures 3-7 and 3-8) (Nike, 2011).
The original design of the Nike+ mobile fitness app was introduced in May 2006 as part of the technology that was attached or embedded into a Nike shoe (Nike, 2011). In September 2010, Nike released the Nike+ GPS App on the Apple platform, with the ability to allow users’ Facebook friends to provide moral support as they run (Van Grove, 2010).

By creating a simple way to collect data, along with tools to use and share it, Nike has created a community of more than 1.2 million runners. Data analysis of their collected runs would suggest that the group has tracked more than 130 million miles and burned more than 13 billion calories (McClusky, 2009). Nike provides users with observations about interesting personal habits from their activity database, such as workout patterns during the winter months (people in the U.S. run more often than those in Europe and Africa but for shorter distances), the average duration of a run worldwide (35 minutes), and the most popular Nike+ Powersong, which runners can set to give them extra motivation—“Pump It” by the Black Eyed Peas (McClusky, 2009).

### 3.2.4 MyFitnessPal

MyFitnessPal is an online health and fitness community that offers useful tools, advice, and support to help people meet their weight-loss and fitness goals (Figure 3-9). Like Endomondo, RunKeeper, and Nike+, the MyFitnessPal app allows for the measurement of daily physical activity (time, distance and type) and has an online community that offers tips and support to help with motivation along the way.
While similar to the other mobile fitness apps discussed in this paper, MyFitnessPal also includes a robust daily food-tracking option with a database of over 1.2 million searchable items maintained by the USDA. One feature of the set-up process is personalised goal setting with respect to body weight. A user can decide to gain, lose or maintain weight.

Based on the user’s fitness profile, MyFitnessPal recommends a daily net calorie target. The tracking of exercise (calories out) and food consumption (calories in) throughout the day adjusts the daily net calorie target. Reoccurring exercises and/or food can be saved as “favourites,” thus allowing for quick logging.

In 2011, a Walden University study proposed that positive social change by tracking calories via smartphones using MyFitnessPal could encourage users to make healthy choices and thus reduce the overall prevalence and incidence of obesity and related health conditions (e.g., hypertension, type 2 diabetes and cardiovascular diseases) within their communities (Hijazi, 2011). MyFitnessPal is a free online service, with supplemental apps on iPhone and Android platforms for additional methods of data collection. MyFitnessPal integrates with Facebook and Twitter, allowing for customized sharing of activities (Figure 3-10).
3.2.5 DailyMile

DailyMile is a San Francisco–based company described by founders Kelly Korevec and Ben Weiner as “a social experience for active people, a community of people just off the couch to ultra-marathoners alike, who encourage and inspire one another as we achieve our goals” (DailyMile, 2012).

Founded in 2008, DailyMile was originally designed to cater to active types such as runners and cyclists who often trained alone by incorporating the sharing of workouts via social media, which allows people to train together virtually (Figure 3-11). The service is a combination fitness log, motivational tool and social-networking hub aimed at using social media to help people achieve their health and wellness goals, such as training for a big race or losing weight, all the while connecting with others who are trying to bring fitness and health into their offline lifestyles (Henning, 2010).
As of November 2011, DailyMile reported that more than 10.1 million workouts by members were completed, with over 8.9 million-member interactions via comments posted, and total member activity accounting for over 72 million doughnuts being burned. The site has reached over 200,000 members and adds over 3,000 new members weekly (DailyMile, 2012). DailyMile currently interfaces with devices such as Nike+, Garmin, Apple mobile platforms and Android clients. Members can download and embed personalised widgets of code that can be added to their own blog or website to track exercise mileage.

The DailyMile website has three areas of focus: profile, training, and community. Within the community section, a member can interact with other members, participate in challenges and forums, view shared exercise routes and enter local fitness events. Unlike the other mobile fitness applications discussed in this research, DailyMile does not have its own app, but rather uses an API to enable third-party developers to build applications on the DailyMile Social Workout Platform. These third-party apps include, but are not limited to, Electric Miles, Runmeter, LogYourRun, Kinetic and Jog Log. These apps allow functions such as data entry and deletion, comments, likes, friends, routes and GPS location (DailyMile, 2012). DailyMile also uses members to provide crowdsourcing ideas to developers. Members have suggested concepts such as mobile apps, blog integration, Google Health data transfer, nutritional information exchange and workout logging via SMS.
A breakdown of the functions of the five fitness apps used in this research is presented in Table 3-8.

<table>
<thead>
<tr>
<th></th>
<th>Nike+</th>
<th>RunKeeper</th>
<th>MyFitnessPal</th>
<th>DailyMile</th>
<th>Endomondo</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Size (iOS version)</strong></td>
<td>18.6 MB</td>
<td>11.1 MB</td>
<td>15.9 MB</td>
<td>n/a</td>
<td>5.3 MB</td>
</tr>
<tr>
<td><strong>Native App</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No⁴</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Platform</strong></td>
<td>iOS</td>
<td>iOS, Android</td>
<td>iOS</td>
<td>n/a</td>
<td>iOS, Android, BlackBerry, Windows</td>
</tr>
<tr>
<td><strong>Cost (iOS version)</strong></td>
<td>$1.99</td>
<td>Free</td>
<td>Free</td>
<td>n/a</td>
<td>Free (Pro $3.99)</td>
</tr>
<tr>
<td><strong>Live Cheering</strong></td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>n/a</td>
<td>Yes</td>
</tr>
<tr>
<td><strong># Activities</strong></td>
<td>2¹</td>
<td>14²</td>
<td>350³</td>
<td>n/a</td>
<td>51¹</td>
</tr>
<tr>
<td><strong>Sharing with Twitter and Facebook</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>On Web</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Real-time Coaching</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>n/a</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Real-time Tracking</strong></td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>n/a</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Challenge Friends</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>On Web</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>HR Integration</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>n/a</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>GPS</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>n/a</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Auto Pause</strong></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>n/a</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Languages</strong></td>
<td>English, Chinese, French, German, Italian, Japanese, Portuguese, Spanish</td>
<td>English</td>
<td>English</td>
<td>n/a</td>
<td>English, Danish, French, German, Spanish</td>
</tr>
</tbody>
</table>

Table 3-8 - Mobile Fitness App Comparisons (as of December 2012)

1 - Walk, Run
2 - Running, Cycling, Mountain Biking, Walking, Hiking, Downhill Ski, Cross Country Ski, Snowboarding, Skating, Swimming, Wheelchair, Rowing, Elliptical, Other
3 - Activities are manually entered into the app or website
4 - DailyMile is a website portal that uses a number of different mobile apps to enter data
5 - Walking, Cricket, Running, Cycling Transport, Cycling Sport, Mountain Biking, Skating, Roller Skating, Skating Downhill, Skating Cross Country, Snowboarding, Kayaking, Ice Skate Climbing, Cross Training, Dancing, Fencing, Football, Rugby, Soccer, Handball, Hockey, Pilates, Polo, Scooter, Squash, Tennis, Table Tennis, Beach Volleyball, Volleyball, Weight Training, Yoga, Martial Arts, Gymnastics
3.3 Data Collection from Twitter

There are a number of tools that have been used to collect tweets for academic research, a number of methods to determine the online influence of each user and a number of systems that can be used to determine the sentiment of each tweet. However, the unique contribution of this research to exercise science is the full ecosystem of the processes used to create a unique data set that can be shared with other researchers. These processes included the following: (1) the data collection of physical-activity tweets (such as date, time, location, exercise type, exercise duration and location), (2) the addition of user demographics from one’s Twitter profile (such as location, Twitter use, followers and followings), (3) the integration of one’s social influence (as measured by Klout score) and (4) the ability to assign gender based on government data (Social Security names database). At the time of this research, no other computer science research using collected physical-activity tweet data had been published as the foundation of a method to assess the physical-activity levels within a community.

Researchers typically collect their own data on social networks using one of three general approaches, each of which has its own set of limitations (Cormode et al., 2010).

1. **API Driven** – To assist in data collection, many online social networks provide an Application Programming Interface (API), which is used to query the entities, properties, and relationships of the network. The major limitation is the assumption that the answers to queries via the API are both up-to-date and accurate. Online social networks often impose a limit on the number of API calls per user per day, so large data sets can take days or weeks to process.

2. **Scraping Based** – This data-collection method involves researchers directly accessing the online social networks via a Web client, which then imitates the actions of a user to capture HTML data that is parsed with a hand-crafted, site-specific parser. This method of data collection is considered to be more arduous than API-based methods and may still be limited due to bandwidth restrictions imposed by the site. This method also has to contend with site redesigns by the online social network, which can create errors within the parser. These redesign changes tend to occur with greater frequency and with less notice than changes to the API.

3. **Passive Network Measurement** – This data-collection method allows the researcher to monitor network traffic and selectively collect and parse requests to and from the online social networks of interest. In some ways, this provides the most honest view of the network in use, in that it can capture properties of the network as its users experience it. However, there are significant privacy issues around monitoring and parsing individual’s activities within an online social network. The plethora of access options for modern online social networks (direct Web-based, mobile Web, mobile app, external apps using API) means that it is hard to capture all accesses from any meaningful subpopulation of users.

For the purposes of this research, both API Driven and Scraping Based approaches were used, with the API Driven approach used to collect all Twitter information from all mobile fitness apps. A subset of desired information of data was then collected from one mobile fitness app using the Scraping Based approach.
Data Collection and Processing

Given the hundreds of millions of entities on most popular online social networks, it is not feasible to gather complete information on all properties and activity within the network. Necessarily, any measurement study yields only a sample of the full data, and it is important to make explicit the description of what has been sampled in order to materialize the biases and systematic errors (Cormode et al., 2010).

For academic research purposes, measurement of online social network data has tended to focus on quantity rather than the quality of data. Whilst a larger sample gives one greater confidence in the statistical measures derived, there can be a danger in assuming that bigger is better. The earliest data sets from sociology research, although minuscule by today’s standards, were hand-compiled and carefully curated. Collecting huge volumes of data means that there is no detailed examination of any portion, and the quality of the obtained data is often unknown (Cormode et al., 2010).

There were two main limiting considerations in this research, the first being the participant having access to, and use of, one of the mobile fitness applications identified for research for this study (i.e., RunKeeper, Endomondo, Nike+, MyFitnessPal or DailyMile). The second limiting consideration was that the participant, in addition to the use of a mobile fitness application, must also have a Twitter account and agree to tweet information regarding his or her workouts through the mobile fitness application. The frequency of Twitter use was not a determinant for participation in the research. Additional mobile fitness applications (and thus additional participants) were available for this research; however, it was determined by the researcher that additional information would not elicit further insight into that which was being observed.

3.3.1 TwapperKeeper

After an online review for tweet collections, an open-source program called TwapperKeeper was chosen for Tweet data collection specific to mobile fitness app hashtags (Figure 3-12). TwapperKeeper is a web application designed to archive social-media data via Twitter to allow for long-term archival and analysis. The application uses a Twitter-enabled API that acts as an interface between the Twitter search function and a database for tweet storage. The application allows users to monitor and archive specific hashtags and to provide additional metadata to describe an archive that can later be viewed. TwapperKeeper was installed on a cloud server and began collecting tweets in March 2011. The collected tweets were ones that met the criteria of mobile fitness apps used for this research. The application allows users to monitor and archive specific hashtags and to provide additional metadata to describe an archive that can later be viewed in multiple. The Fitness Tweet Crawler API was used to gather information that linked mobile fitness app tweeters and their publicly available demographic data.
The research team then collected tweets from the five mobile fitness apps by gathering tweets that used the following hashtags: #endomondo, #myfitnesspal, #Nike+, #runkeeper, and #dailymile. These are the hashtags that the apps automatically attach to a tweet to indicate it has come from that particular application. It is through these hashtags that common themes or information can be grouped within Twitter. Tweet collection was done by TwapperKeeper, which began the archiving process by searching publicly available tweets, identifying those that contained the desired hashtags and inserting the identified tweets into a database for later processing. The type of information collected from each tweet is shown in Table 3-9. These datapoints became part of the database schema.

<table>
<thead>
<tr>
<th>Data Point</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archive Source</td>
<td>Twitter Search or Twitter Stream</td>
</tr>
<tr>
<td>Text</td>
<td>The actual tweet</td>
</tr>
<tr>
<td>To_User</td>
<td>Name of recipient user if the tweet was sent to a specific Twitter user</td>
</tr>
<tr>
<td>From_User</td>
<td>Name of the Twitter user that sent the Tweet</td>
</tr>
<tr>
<td>ID</td>
<td>Specific Twitter identification number for the associated Tweet</td>
</tr>
<tr>
<td>From_User_ID</td>
<td>Specific Twitter identification number for the associated Twitter username</td>
</tr>
<tr>
<td>Iso_Language_Code</td>
<td>Identified language of the tweet</td>
</tr>
<tr>
<td>Source</td>
<td>Twitter platform used to send tweet</td>
</tr>
<tr>
<td>Profile Img_URL</td>
<td>URL to the picture of the Tweeter</td>
</tr>
</tbody>
</table>
Data Collection and Processing

<table>
<thead>
<tr>
<th>Geo_Type</th>
<th>Either “point” if geolocation was used with Tweet or blank if not</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geo_Coordinates_0</td>
<td>Latitude of the location where the tweet was sent</td>
</tr>
<tr>
<td>Geo_Coordinates_1</td>
<td>Longitude of the location where the tweet was sent</td>
</tr>
<tr>
<td>Created_At</td>
<td>Day, date and time the tweet was sent</td>
</tr>
<tr>
<td>Time</td>
<td>UNIX time the tweet was sent</td>
</tr>
</tbody>
</table>

*Table 3-9 - Collected Twitter Data Point Descriptions from TwapperKeeper*

Once the hashtags were defined, the application began two archiving processes (TwapperKeeper, 2011):

- **The Stream** – A persistent connection was also created with the Twitter Streaming API for the desired hashtags. The archiving process inserted all inbound tweets into a database table for later processing. A second process was run to analyse each tweet in the table and moved the tweets into the proper archive table.

- **The Crawl** – For the keyword defined (by a hashtag), the crawling processes began to poll the Twitter Search API to find all tweets in the search cache that match the desired hashtag. This allowed for TwapperKeeper to fill in older tweets (limited by the Twitter API, the total number of tweets and date), as well as to continually monitor tweets that might be missed by the Stream archive process. A disruption of service is possible during disconnects/reconnects with the Twitter Streaming API, due to rate limits imposed by Twitter and possible service interruptions on the Twitter service itself.

For the purposes of this research, both the Streaming API and the Search API were used. The Streaming API is the real-time sample of the Twitter Firehose. This API is for those developers with data-intensive needs, such as data mining analytics research. Streaming API allows for large quantities of keywords to be specified and tracked, the retrieval of geo-tagged tweets from a certain region or the return of the public statuses of a user set returned. The Search API is designed for a specific query for Twitter content. This may include finding a set of tweets with specific keywords, tweets referencing a specific user or tweets from a particular user (Twitter, 2011c).

In order to integrate Twitter API with a program, authorizations are required by both of the data sources using OAuth, which is a simple way to interact with and publish protected data (Yahoo! Developer Network, 2010). The method used to generate an OAuth token was adapted from the process as described on the EMC² website (EMC, 2009). The Twitter API rate limits of the Developer and Common User levels were, at the time of this research, 20,000 per hour and 300 per hour, respectively. This research had access to the Developer level API rate limits.
3.3.2 The Fitness Tweet Crawler

In order to collect fitness-related tweets, a tool was developed at the Digital Enterprise Research Institute at the National University of Ireland at Galway. The Fitness Tweet Crawler is a PHP scripting language and JavaScript system that takes Twitter information collected from the TwapperKeeper database and requests additional publicly available information about a user from his or her Twitter profile. At the time of this research, the Fitness Tweet Crawler was the first and only classification model to incorporate individual tweets, Twitter-user demographics, and a user’s Klout score, an online influence score for individuals who are active on social networking services such as Twitter.

After limiting the process to unique users, the Fitness Tweet Crawler has the ability to request additional demographic information for each Twitter user from two different sources: Twitter for demographic and user information (Table 3-10) and Klout for the unique user’s Klout score and style (Table 3-11). Applying effective ranking techniques such as Klout for online users similar to the ranking of webpages to determine influential users on the internet has the potential to lead to many new and useful applications (Rao, Spasojevic, Li, & Dsouza, 2015). These datapoints became part of the database schema. This data was requested once at the beginning of the data collection period.

<table>
<thead>
<tr>
<th>Data Point</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>User_Name</td>
<td>Twitter user name of the person that sent the tweet</td>
</tr>
<tr>
<td>Location</td>
<td>Location of tweeter as recorded in his or her Twitter profile</td>
</tr>
<tr>
<td>Tweets</td>
<td>Total tweets sent by user at the time of the query</td>
</tr>
<tr>
<td>Following</td>
<td>Number of people the user is following</td>
</tr>
<tr>
<td>Followers</td>
<td>Number of people that follow the user</td>
</tr>
<tr>
<td>Access Date</td>
<td>Date of the query</td>
</tr>
<tr>
<td>Access Time</td>
<td>Time of the query</td>
</tr>
<tr>
<td>Twitter_Startdate</td>
<td>Date the user started using Twitter</td>
</tr>
<tr>
<td>Times_Per_Day</td>
<td>Number of times per day the user sends any tweet</td>
</tr>
</tbody>
</table>

Table 3-10 - Twitter Data Point Descriptions from Twitter for the Fitness Tweet Crawler

3 The concept of the Fitness Tweet Crawler was that of Theodore A. Vickey. Programming assistance was provided by DERI teammates Mengjiao Wang and Pavan Kapanipathi.
In the first step, tweets were collected that included any of the five fitness-related hashtags: #endomondo, #myfitnesspal, #Nike+, #runkeeper and #dailymile (Figure 3-13).

In the second step, the Fitness Tweet Crawler was used to collect demographic/user profile information from Twitter (Figure 3-14). A second data request process was created to pull information from the Klout website. This data request was established for future research regarding social influence and fitness tweets.

By collecting the Twitter username from TwapperKeeper, the Fitness Tweet Crawler was able to collect additional data points using the Twitter and Klout APIs (Table 3-12).
All information collected was publicly available, with each user of Twitter agreeing to this public sharing of information per the terms and conditions of his or her Twitter account.

Using these tools, a growing dataset was created of public information that provides a wealth of data about the person’s shared exercise that include, but are not limited, to exercise type, length, the day of the week, mood, geographical location and time. In addition, Twitter user information was collected from each user including, but not limited, to Twitter ID, Twitter activity, number of followers, number of followees, number of tweets, Klout influence score and Twitter start date. This combined information will allow research on how technology can be used to monitor and motivate individuals, the exercise habits of those who share their workout information via mobile fitness apps and Twitter and how a person’s social network influences his or her fitness activities.

One aspect of these mobile devices is the ability to share health and fitness information with others in a social network. Collecting one’s personal physical-activity data via a mobile fitness app, and then sharing it via one’s social network, could facilitate a change in physical activity due to an increase in social support. The importance of receiving adequate amounts of social support for physical activity is apparent.
<table>
<thead>
<tr>
<th>Username</th>
<th>Location</th>
<th>Tweets</th>
<th>Following</th>
<th>Followers</th>
<th>Klout</th>
<th>True Reach</th>
<th>Amp</th>
<th>Network</th>
<th>Style</th>
<th>Twitter Start</th>
<th>Times per Day</th>
<th>Once Every</th>
</tr>
</thead>
<tbody>
<tr>
<td>User1</td>
<td>East Lansing, MI</td>
<td>1012</td>
<td>469</td>
<td>110</td>
<td>38.61</td>
<td>37</td>
<td>22.87</td>
<td>41.51</td>
<td>Explorer</td>
<td>16-Jan-11</td>
<td>7.61</td>
<td>0.13</td>
</tr>
<tr>
<td>User2</td>
<td>Dubai</td>
<td>10</td>
<td>125</td>
<td>35</td>
<td>10.66</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>Explorer</td>
<td>22-Feb-11</td>
<td>0.10</td>
<td>9.60</td>
</tr>
<tr>
<td>User3</td>
<td>Sutton</td>
<td>390</td>
<td>153</td>
<td>22</td>
<td>10.39</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>Explorer</td>
<td>8-Sep-10</td>
<td>1.48</td>
<td>0.67</td>
</tr>
<tr>
<td>User4</td>
<td>Southern Germany</td>
<td>6688</td>
<td>84</td>
<td>30</td>
<td>43.55</td>
<td>10</td>
<td>20.08</td>
<td>47.79</td>
<td>Feeder</td>
<td>6-May-09</td>
<td>8.90</td>
<td>0.11</td>
</tr>
<tr>
<td>User5</td>
<td>LA</td>
<td>927</td>
<td>89</td>
<td>29</td>
<td>16.88</td>
<td>12</td>
<td>11.22</td>
<td>10</td>
<td>Explorer</td>
<td>2-Jun-09</td>
<td>1.28</td>
<td>0.78</td>
</tr>
<tr>
<td>User6</td>
<td>New Orleans, LA</td>
<td>5262</td>
<td>870</td>
<td>1526</td>
<td>44.22</td>
<td>535</td>
<td>24.07</td>
<td>56.04</td>
<td>Thought Leader</td>
<td>2-Jul-09</td>
<td>7.57</td>
<td>0.13</td>
</tr>
<tr>
<td>User7</td>
<td>Arica, Chile</td>
<td>1438</td>
<td>157</td>
<td>209</td>
<td>12.2</td>
<td>72</td>
<td>10</td>
<td>10</td>
<td>Explorer</td>
<td>11-May-09</td>
<td>1.92</td>
<td>0.52</td>
</tr>
<tr>
<td>User8</td>
<td>Nijmegen-area</td>
<td>3255</td>
<td>739</td>
<td>620</td>
<td>48.83</td>
<td>238</td>
<td>31</td>
<td>56.96</td>
<td>Explorer</td>
<td>18-May-10</td>
<td>8.76</td>
<td>0.11</td>
</tr>
<tr>
<td>User9</td>
<td>Houston, Texas</td>
<td>89</td>
<td>61</td>
<td>12</td>
<td>10.02</td>
<td>4</td>
<td>10</td>
<td>10</td>
<td>Explorer</td>
<td>30-Apr-10</td>
<td>0.23</td>
<td>4.43</td>
</tr>
<tr>
<td>User10</td>
<td>Philippines</td>
<td>210</td>
<td>94</td>
<td>43</td>
<td>36.55</td>
<td>16</td>
<td>18.9</td>
<td>39.36</td>
<td>Explorer</td>
<td>9-May-11</td>
<td>10.50</td>
<td>0.10</td>
</tr>
<tr>
<td>User11</td>
<td>Waterloo, Ontario</td>
<td>36</td>
<td>40</td>
<td>20</td>
<td>10.01</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>Explorer</td>
<td>2-Mar-11</td>
<td>0.41</td>
<td>2.44</td>
</tr>
<tr>
<td>User12</td>
<td>Scottsdale, AZ</td>
<td>95</td>
<td>1967</td>
<td>1095</td>
<td>17.57</td>
<td>451</td>
<td>10.3</td>
<td>10</td>
<td>Explorer</td>
<td>8-Jun-10</td>
<td>0.27</td>
<td>3.74</td>
</tr>
<tr>
<td>User13</td>
<td>Brazil</td>
<td>569</td>
<td>1775</td>
<td>863</td>
<td>15.92</td>
<td>339</td>
<td>10.19</td>
<td>10</td>
<td>Explorer</td>
<td>18-Sep-09</td>
<td>0.92</td>
<td>1.09</td>
</tr>
</tbody>
</table>
3.4 The Fitness Tweet Classification Model

This section introduces the Fitness Tweet Classification Model and highlights the three main classification types—Activity, Blarney, and Conversation. An analysis of the resulting database of the fitness tweets from the five mobile fitness apps is also presented⁴. In addition to overall fitness tweeting as a group, individual analysis of each of the five mobile fitness apps is presented⁵.

The Fitness Tweet Classification Model was based on available macro-topic classification models where Tweets were broadly categorised. During the course of this research, it became apparent that, while tweet classification models existed in other areas of research, no such model existed in the health and fitness areas of academic study (Vickey et al., 2013). Thus, the Fitness Tweet Classification Model was created (Figure 3-15).

A processing script was created at the National University of Ireland Galway that followed the resulting logic to assign classification (Figure 3-16).

---


⁵ Parts of this chapter were published in Vickey, T. A., Martin Ginis, K., & Dabrowski, M. (2013). Twitter Classification Model: The ABC of Two Million Fitness Tweets. *Translational Behavioural Medicine (3)*, 3, 304–311.
The Fitness Tweet Classification Model allows researchers of mobile fitness applications to classify the fitness tweets into three main categories:

1. Activity
2. Blarney
3. Conversation

### 3.4.1 Activity

Tweets that share a person’s workout, specific to the Tweet structure as defined by the five different mobile fitness apps, were classified as Activity. Each mobile fitness app used a different data structure that was able to be defined. The text of each Activity tweet contained the following data points:

- $d$ – distance
- $u$ – unit
- $a$ – activity
- $t$ – time
- $p$ – pace
Data Collection and Processing

Examples:

Context: Just posted a *d u a* with @runkeeper. Check it out! http://runkpr.com/xxxxxxx #RunKeeper

Tweet: Just posted a 5.02 mi run with @runkeeper. Check it out! http://runkpr.com/aldj5k #RunKeeper

Context: I just finished a *d u a* with a time of *t* with Nike+ GPS. #Nike+

Tweet: I just finished a 6.99 km run with a time of 36:00 with Nike+ GPS. #Nike+

Context: I just finished a *d u a* with a pace of *p* with Nike+

Tweet: I just finished a 5.58 mi run with a pace of 15'10"/mi with Nike+ Running. #nike+

Some mobile fitness apps allow the user to add additional information to the tweet. This would indicate a different level of sharing information about a workout. Thus, an additional sub-category was created—Workout+. A Workout+ tweet has the same foundation of a Workout tweet but adds the additional variable of information.

---

*i* – information (an indication of a Workout+)

Examples:

Context: Just completed a *d u a i*. Check it out! http://runkpr.com/xxxxxxx #RunKeeper

Tweet: Just completed a 3.27 mi hike - Trees down from storm. Check it out! http://t.co/XIYTjf2 #RunKeeper

Context: Was out *a d u* with #Endomondo. *i* See it here: http://t.co/xxxxxxx

Tweet: Was out running 2.4 miles with #Endomondo. First run since surgery. See it here: http://bit.ly/tAV69

Table 3-13 provides examples of Workout and Workout+ within the Activity category.
3.4.2 Blarney

Over the course of data processing, it became apparent that some tweets did not contain either Activity or Conversation, thus the Blarney category was created. Blarney is defined as skillful flattery, nonsense or blandishment (Merriam-Webster, 2012b). Tweets that were classified as Blarney were further classified into Pointless Babble or Spam.

As the popularity of Twitter has increased, so has Twitter spam. Spam becomes a problem as soon as an online communication medium becomes popular. Twitter’s behavioural and structural properties make it a fertile breeding ground for spammers to proliferate (Yardi, Romero, Schoenebeck & Boyd, 2010). Spammers use Twitter as a tool to post malicious links, send unsolicited messages to legitimate users and hijack trending topics.

Spam is a moving target and thus can often be difficult to measure. A study shows that more than 3% of messages on Twitter are spam (A. Wang, 2010). By comparison to email, in 2009, Microsoft reported that 97% of all email messages sent over the Web were unwanted, Google reported that spam hovered between 90 and 95%, and Symantec reported that spam accounted for 90.4% of all email (Yardi et al., 2010). Most e-mail spam is filtered by email servers and goes unnoticed, yet spammers persist in creative and sophisticated ways. The economics of email spam relates directly to how much money spammers can make off of Internet users who click on their spam links (Yardi et al., 2010). This same economic metric can be used for Twitter spam.

Twitter has strict rules regarding spam accounts. As outlined in the Twitter Spam policy, what constitutes “spamming” will evolve as new tricks and tactics are created by spammers (Twitter, 2016).

<table>
<thead>
<tr>
<th>Activity</th>
<th>Workout</th>
<th>Workout+</th>
</tr>
</thead>
<tbody>
<tr>
<td>RunKeeper</td>
<td>Just completed a 5.69 km walk with @runkeeper. Check it out! [<a href="http://runkeeper.com/ahr0yw">http://runkeeper.com/ahr0yw</a> #RunKeeper](<a href="http://runkeeper.com/ahr0yw">http://runkeeper.com/ahr0yw</a> #RunKeeper)</td>
<td>Just posted a 2.25 mi run - just some miles to stretch the legs before the race tomorrow. [<a href="http://runkeeper.com/ahrne649">http://runkeeper.com/ahrne649</a> #RunKeeper](<a href="http://runkeeper.com/ahrne649">http://runkeeper.com/ahrne649</a> #RunKeeper)</td>
</tr>
<tr>
<td>Nike+</td>
<td>I just finished a 2.00 mi run with a time of 20:05 with Nike+ GPS. #nikeplus</td>
<td>9.69 miles this morning. Longest run since the move to Colorado. #nikeplus #fb</td>
</tr>
</tbody>
</table>
| MyFitnessPal | burned 157 calories doing 60 minutes of "Yoga" #myfitnesspal

Table 3-13 - Sample Activity Tweets
A standard feature of Twitter is the filtering of URLs linked to known malicious sites. However, great vulnerability is the presence of shortened URLs. Since Twitter only allows users to post a short message of 140 or fewer characters, URL-shortening services have become popular to meet the requirements. Many of the mobile fitness apps reviewed for this research use URL-shortening services to link tweets to websites containing additional workout information, including running routes. Shortened URLs can hide the source URLs and obscure the malicious sites behind them. As a result, these shortened links provide an opportunity for attackers to prank, phish and spam. While Twitter does not check these shortened URLs for malware, updates are considered spam if they consist mainly of links and not personal updates, according to Twitter’s policy (Wang, 2010).

Any tweet that contained just a URL link or appeared to be an unsolicited commercial tweet, as determined by the lead researcher after a manual review of selected tweets, was classified as Spam; all other tweets classified as Blarney were sub-classified as Pointless Babble (Table 3-14).

<table>
<thead>
<tr>
<th>BLARNEY</th>
<th>Pointless Babble</th>
<th>Spam</th>
</tr>
</thead>
<tbody>
<tr>
<td>RunKeeper</td>
<td>Other Activity 0.00 km</td>
<td>RunKeeper <a href="http://t.co/24Oy6MvJ">http://t.co/24Oy6MvJ</a> via @addthis</td>
</tr>
<tr>
<td>Nike+</td>
<td>#nikeplus software is crap. Runs far too slow?</td>
<td>I just finished not caring w/a time of 2.5 seconds using #nikeplus</td>
</tr>
</tbody>
</table>

Table 3-14 - Sample Blarney Tweets

### 3.4.3 Conversation

Since its inception, Twitter has been used for branding campaigns by corporations, as an information-sharing medium during elections by political parties and as an information-dissemination tool by news media (Lee et al., 2011). These types of conversations are also occurring with mobile fitness apps (Table 3-15). After a review of the initial data set of tweets, it was evident that within the Conversation category, people talked about four main areas—asking for Technical Support (could come from the app itself or the community), Corporate Marketing such as press releases and information from the apps regarding updates (could come from the app company itself or the community, mostly by re-tweeting),
Statements of Support (where people within the app community congratulated others on reaching milestones, achieving personal bests, etc.) or Information Sharing (example – those within the app community that wanted to run together in an upcoming 10K race would post messages using the hashtag per the app).

<table>
<thead>
<tr>
<th>CONVERSATION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technical Support</strong></td>
</tr>
<tr>
<td>DailyMile</td>
</tr>
<tr>
<td>Endomondo</td>
</tr>
<tr>
<td>RunKeeper</td>
</tr>
<tr>
<td>Nike+</td>
</tr>
<tr>
<td>MyFitnessPal</td>
</tr>
</tbody>
</table>

Table 3-15 - Sample Conversation Tweets

### 3.4.4 Fitness Tweet Classification Model Script

After conducting a review of previous classification models in other research domains and months of manual analysis of collected tweet and reliability testing with independent review coders, the Fitness Tweet Classification Model code was created. As identified in the manual data analysis, there were three main classifications:

- **Activity (1)**
- **Blarney (2)**
- **Conversation (3)**

For coding purposes, the names were replaced by the corresponding number.
Each main classification has a number of different and unique sub-classifications:

**ACTIVITY**
1.1 Workout
1.2 Workout+

**BLARNEY**
2.1 Pointless Babble
2.2 Spam

**CONVERSATION**
3.1 Technical Support
3.2 Corporate Marketing
3.3 Statements of Support
3.4 Information Sharing

**CLASSIFICATION PROCESS:**

**Procedure 1 (for all hashtags)**

After review of the sample dataset, the following classifications were created for the script.

Additional criteria for classifications were added after processing. As more than 70% of the tweets were easily identified as Activity, the remaining Blarney and Conversation classifications were less tedious to create by hand after sorting all tweets in alphabetical order and filtering the previously classified Activity tweets.

For ALL hashtags (dailymile, endomondo, myfitnesspal, nike+ and runkeeper) – This is the first classification. If it has any of these, then classify accordingly

Any RT - CLASSIFY AS INFORMATION SHARING (3.4)  
Tweet starts with # - CLASSIFY AS INFORMATION SHARING (3.4)  
Tweet starts with @ - CLASSIFY AS INFORMATION SHARING (3.4)  
Tweet starts with “ - CLASSIFY AS UNCLASSIFIED CONVERSATION (3.0)  
Tweet starts with ( - CLASSIFY AS UNCLASSIFIED CONVERSATION (3.0)  
Tweet starts with , - CLASSIFY AS UNCLASSIFIED CONVERSATION (3.0)  
Tweet starts with * - CLASSIFY AS UNCLASSIFIED CONVERSATION (3.0)  
Tweet starts with ) - CLASSIFY AS UNCLASSIFIED CONVERSATION (3.0)  
Tweet starts with [ - CLASSIFY AS UNCLASSIFIED CONVERSATION (3.0)  
Tweet starts with \ - CLASSIFY AS UNCLASSIFIED CONVERSATION (3.0)  
Tweet starts with + - CLASSIFY AS UNCLASSIFIED CONVERSATION (3.0)  
Tweet starts with < - CLASSIFY AS UNCLASSIFIED CONVERSATION (3.0)  
Tweet starts with = - CLASSIFY AS UNCLASSIFIED CONVERSATION (3.0)  
Tweet that contains “Awesome” - CLASSIFY AS STATEMENT OF SUPPORT (3.3)  
Tweet that contains “WOW” - CLASSIFY AS STATEMENT OF SUPPORT (3.3)  
Tweet that contains “Congrats” - CLASSIFY AS STATEMENT OF SUPPORT (3.3)  
Tweet that contains “great job” - CLASSIFY AS STATEMENT OF SUPPORT (3.3)  
Tweet that contains #FF - CLASSIFY AS STATEMENT OF SUPPORT (3.3)  
Tweet that contains “good job” - CLASSIFY AS STATEMENT OF SUPPORT (3.3)


Procedure 2 (for specific hashtags)

For any activity, if the distance is less than 0.1 mile or 0.16 km, then classify at BLARNEY: SPAM (2.2).

FOR HASHTAG #DAILYMILE
These classify as ACTIVITY: WORKOUT (1.1)

Did a X (type) workout in Y (distance) in Z (time) and felt ZZ.

EXAMPLE: Did a cc skiing workout 1.1 miles in 15 mins and felt good.
http://dailymile.com/e/R3zA

X could equal: [cc skiing, core fitness, cross training, CrossFit, commute, elliptical, fitness, hiking, inline skating, rock climbing, rowing, spinning, weights, yoga]

Y is user entry with a label of distance as [mile, miles, kilometre, kilometers, km, kms, meter, meters]

Z is a user entry with a label of time: [sec, secs, min, mins, hour, hours, X hour and XX min]

ZZ could equal mood: [great, good, alright, blah, tired, injured]

Explanation – when sharing information via Twitter, the app does one of two things. The first is that the app shares just the workout information (distance and time). The second is that the person, instead of just sharing the default information, actually takes the time to include additional information about the workout (such as their mood, a short statement, etc.). We want to be able to see that difference. Thus, if it is just distance and time, it is 1.1, but if it contains more, then it is classified as 1.2.

(These will contain ran, rode, swam or walked just in front of Y.)

I ran Y (distance) in Z (time) CLASSIFY AS WORKOUT (1.1)

Just ran Y (distance) in Z (time) CLASSIFY AS WORKOUT (1.1)

Ran Y (distance) Z (time) and felt ZZ (mood) CLASSIFY AS WORKOUT (1.2)

Ran Y (distance) in Z (time) and felt ZZ (mood). http * CLASSIFY AS WORKOUT (1.2)

Ran Y (distance) in Z (time) and felt ZZ (mood). * CLASSIFY AS WORKOUT+ (1.2)

Rode Y (distance) Z (time) and felt ZZ (mood) CLASSIFY AS WORKOUT (1.1)
Rode Y (distance) in Z (time) and felt ZZ (mood). http * CLASSIFY AS WORKOUT (1.2)

Rode Y (distance) in Z (time) and felt ZZ (mood). * CLASSIFY AS WORKOUT+ (1.2)

Swam Y (distance) Z (time) and felt ZZ (mood) CLASSIFY AS WORKOUT (1.1)

Swam Y (distance) in Z (time) and felt ZZ (mood). http * CLASSIFY AS WORKOUT (1.2)

Swam Y (distance) in Z (time) and felt ZZ (mood). * CLASSIFY AS WORKOUT+ (1.2)

Walked Y (distance) Z (time) and felt ZZ (mood) CLASSIFY AS WORKOUT (1.1)

Walked Y (distance) in Z (time) and felt ZZ (mood). http * CLASSIFY AS WORKOUT (1.2)

Walked Y (distance) in Z (time) and felt ZZ (mood). * CLASSIFY AS WORKOUT+ (1.2)

**FOR HASHTAG #ENDOMONDO**

For tweets that begin with “Just began a *” CLASSIFY AS CONVERSATION INFORMATION SHARING (3.4)

For tweets that begin with the bullets below, CLASSIFY AS CONVERSATION CORPORATE MARKETING (3.2)

#scsintl
#sclal
#start
#tech*

[ android
endmnd android
endmnd lets
endmnd partners
endmnd raises
endmnd sprts tracker
fitness app developer
health & fitness
just updated Endmnd
social fitness app
sports app endmnd

63
Activity tweets follow structure: CLASSIFY AS ACTIVITY: WORKOUT (1.1)


X could equal:

boxing OR cross training OR cycling OR dancing OR doing aerobics OR doing gymnastics OR doing martial arts OR doing weight training OR exercising OR fencing OR golfing OR hiking OR kayaking OR kite surfing OR mountain biking OR orienteering OR playing American football OR playing badminton OR playing baseball OR playing basketball OR playing hockey OR playing soccer OR playing squash OR playing table tennis OR playing tennis OR playing volleyball OR riding OR roller skiing OR rowing OR running OR sailing OR scuba diving OR skating OR skiing OR snowboarding OR spinning OR swimming OR trekking OR walking OR windsurfing

Y could be km OR mile(s).

FOR HASHTAG #MYFITNESSPAL
If the tweet starts with “completed her” or “completed his” or “posted” THEN CLASSIFY AS INFORMATION SHARING (3.4)

If the tweet starts with “download*” or “free” THEN CLASSIFY AS CORPORATE MARKETING (3.2)

If the tweets start with “Endomondo comes to” or “High Tech” or “HTC” THEN CLASSIFY AS CORPORATE MARKETING (3.2)

If the tweet starts with “lost” THEN CLASSIFY AS STATEMENT OF SUPPORT (3.3)
Activity tweets follow structure: CLASSIFY AS ACTIVITY: WORKOUT (1.1)
Burned X calories doing Y minutes of “Z” #myfitnesspal

EXAMPLE: burned 400 calories doing 30 minutes of "Elliptical Trainer" #myfitnesspal

X = user input number
Y= user input number
Z= an activity that is between quotation marks

FOR HASHTAG #NIKE+
Data Collection and Processing

Nike seems to have a large number of non-English tweets.
If the tweet starts with @nikerun* or @nikestore THEN CLASSIFY AS CONVERSATION:
CORPORATE MARKETING 3.2

If the tweets contain “Nike is donating $1” OR “USOC signs sponsorship” OR “Support the “All for Japan”” OR “If you join the “Every Mile Counts”” OR “If everyone logs 1more mi” OR “better hurry” OR “make your miles count” OR “we did it!” OR “@outcastagency” OR “@runnersworld: First Look.” OR “@tessmith” THEN CLASSIFY AS CONVERSATION: CORPORATE MARKETING 3.2

If the tweet contains @Free_2_Work or @timberland_Jeff or “Nike scores A- “ THEN CLASSIFY AS CONVERSATION: SPAM (2.2)

If the tweet starts with @ozy* OR @paper* OR @Puley* OR @randem* OR @funethechamp OR @runriderun OR @ryfe* OR @sfelgner* OR @skitchey89* OR @sophiehowl27 OR @southleslie OR @stacialynh* OR @stevenshipo* OR @susielauterborn OR @thaREALmeredith* OR @theAmyAnderson OR @thirtydegree OR @TheJimmyRund* OR @thrutheblue OR @toku* OR @tommyguns87* OR @torrybruce OR @wandaswORld2011* OR @wardamndarcie* THEN CLASSIFY AS CONVERSATION: TECHNICAL SUPPORT (3.1)

If the tweet starts with “Check out this Nike” THEN CLASSIFY AS CONVERSATION: INFORMATION SHARING

If the tweet contains “wow” or “congrats” or “yay” THEN CLASSIFY AS CONVERSATION: STATEMENTS OF SUPPORT (3.3)

Activity tweets follow structure: CLASSIFY AS ACTIVITY: WORKOUT (1.1)

I just finished a X (distance) run with a time of Y with Nike+ GPS. #Nike+

X= the distance as a number and could be labelled as km or mi

EXAMPLE: I just finished a 7 km run with a time of 35:07 with Nike+ GPS. #Nike+

If X is less than 0.1 mile or 0.16 km, then classify at BLARNEY: SPAM (2.2)

If the tweet contains http://slowgeek.com, then CLASSIFY AS ACTIVITY: WORKOUT (1.1)

EXAMPLE: ran 7.40 km @ 5’59”/km (4.60 mi @ 9’38”/mi) on Mar.16 14:01, 446 cal.
FOR HASHTAG #RUNKEEPER

Activity tweets follow structure: CLASSIFY AS ACTIVITY: WORKOUT (1.1)

Just completed a X (distance) Y (type) with @runkeeper. Check it out! http://*

X= distance (km or mi)

Y= type (run, bike ride, walk, hike, activity, snowboard ride, skate, ski run, row, swim, chair ride)

EXAMPLE: Just completed a 0.20 mi bike ride with @runkeeper. Check it out!
http://rnkpr.com/aiq05o #RunKeeper

If different than the basic structure above, THEN CLASSIFY AS ACTIVITY: WORKOUT+ (1.2)

EXAMPLE: Just completed a 3.92 mi bike ride - Shorter route today, wanted to take it easy after killing myself ... http://rnkpr.com/agzwi4

If the tweets starts with “#tech” OR “#thenextweb” OR “tweetdingman” OR “tweetsword” OR “Twitter #18” OR “jacobss22” OR “jazzon” OR “Business intelligence” OR “Can The Internet” OR “FitBit Partners” OR “Fitness gets futuristic:” THEN CLASSIFY AS CONVERSATION:
CORPORATE MARKETING (3.2)

If the tweets start with “Achieved a new” OR “Check out my” THEN CLASSIFY AS CONVERSATION: STATEMENT OF SUPPORT (3.3)

If the tweet starts with “Completed a X mi run – Temp:” THEN CLASSIFY AS ACTIVITY: WORKOUT (1.1)

If the tweet starts with “Completed a X km run – Temp:” THEN CLASSIFY AS ACTIVITY: WORKOUT (1.1)

If the tweet starts with “Google ditches” THEN CLASSIFY AS BLARNEY: SPAM (2.2)

If the tweet starts with “I just finished” THEN CLASSIFY AS ACTIVITY: WORKOUT (1.1)

If the tweet starts with “I just joined” THEN CLASSIFY AS CONVERSATION: INFORMATION SHARING (3.4)
If the tweet starts with “I just signed up for @RunKeeper Elite” THEN CLASSIFY AS CONVERSATION: CORPORATE MARKETING (3.2)

If the tweet starts with “I’m training” THEN CLASSIFY AS CONVERSATION: INFORMATION SHARING (3.4)

If the tweet starts with “I’m using the RunKeeper Pro app” THEN CLASSIFY AS CONVERSATION: CORPORATE MARKETING (3.2)

If the tweet starts with “It is clear RunKeeper” THEN CLASSIFY AS CONVERSATION: CORPORATE MARKETING (3.2)

If the tweet starts with “Watch my” THEN CLASSIFY AS CONVERSATION: INFORMATION SHARING (3.4)

If the tweets start with “Zeo and RunKeeper” THEN CLASSIFY AS CONVERSATION: CORPORATE MARKETING (3.2)

3.5 Conclusion
This chapter introduced four elements of this research; the content analysis, the selection criteria for the mobile fitness apps used in this research, the Fitness Tweet Crawler and the Fitness Tweet Classification Model. The creation of the Fitness Tweet Classification Model has established one such set of rules specific to physical activity data collection using Twitter. These models were created to allow future researchers to modify the code to allow for their own data collection, classification, and analysis using this as the foundation.
4 RESULTS ANALYSIS AND INTERPRETATION OF THE FITNESS TWEETS

This chapter discusses the analysis and interpretation of the fitness tweets as categorised by the Fitness Tweet Classification Model. It is through this analysis and interpretation that the context of fitness tweeting from within mobile fitness apps provides insights into what is being shared, by whom and for what reasons. In addition, this chapter discusses the method by which additional demographic characteristics can be extracted from the Fitness Tweet DataSet that can allow for a more in-depth analysis and comparisons between characteristics such as age, location, gender, and online influence.

4.1 Descriptive Analysis of App-shared Tweets

Data collection using TwapperKeeper began on Thursday, April 21, 2011, at 00:00 Greenwich Mean Time and continued until September 21, 2011, at 23:59, for a total collection of Twitter data of 184 days. After review of the collected Twitter data specific to the five mobile fitness app hashtags used for this research (#Nike+, #runkeeper, #dailymile, #myfitnesspal and #endomondo), it was found that 2,856,534 tweets were collected in 23 different languages (Figure 4-1).

---

Results Analysis and Interpretation of the Fitness Tweets

Figure 4-1 - Collected Fitness Tweets by Language (Excluding English)

Figure 4-1 provides insight into the emerging popularity of fitness tweeting from mobile fitness apps from around the world. Of the 23 identified languages from Twitter, English was by far the most represented, with more than 69% of all Tweets. Of note is the popularity of fitness tweeting activities from Asian-speaking users (e.g., Japanese and Indonesian), with the exception of China. The lack of Chinese fitness tweets is related to the Chinese national regulations against the use of Twitter, with the reported 1,402 Chinese tweets most likely being posted from users from outside China. While presenting this research in Shanghai in 2011, it was suggested that mobile fitness app users in China probably fitness tweet in English to avoid punishment by the government. Also of interest is the popularity of fitness tweeting in the Japanese-speaking world. The relatively high ranking of Danish-speaking users versus total population is due to the fact that Endomondo was developed in Denmark.

However, for the purposes of this research, only the English-language tweets were used. Thus, the total number of processed tweets analysed was 1,971,425 (Figure 4-2).
The breakdown of these English-language tweets is presented in Figure 4-3. As the analysis will reveal, users were not limited to only tweeting within one specific area. Therefore, it is very likely that the categories are not mutually exclusive. Since users could post in any or all of the classifications, Unique Users of every single category cannot be added together for a complete Unique User total.
4.2 Analysis of Fitness Tweet Classifications

This section presents the user fitness tweet interactions on Twitter based on the five mobile fitness apps and the Fitness Tweet Classification Model (Figure 4-4). The dataset contained all English-language tweets featuring one of the five mobile fitness app hashtags (#runkeeper, #Nike+, #endomondo, #dailymile or #myfitnesspal) during the specified time period. Of the 1,971,425 total tweets, 1,446,462 (73%) were classified as Activity, 104,360 (5%) as Blarney and 420,603 (18%) as Conversation. The remaining 4% were classified as other; these tweets were from MyFitnessPal and reported daily food journaling. For some calculations within this research, these tweets were included as Conversation and Information Sharing and thus noted as such when appropriate.

Analysis performed on the database determined the daily usage of each mobile fitness app. Since there was an increase in the popularity of mobile fitness apps during this time period, there was a steady increase in the overall daily usage of each mobile fitness app over time. Additional analysis suggested that a weekly decrease occurred toward the end of the work week (Thursday) and continued throughout the weekend, and then increased at the start of the following week. The dates listed on the horizontal axis in Figure 4-4 indicate Thursdays during the selected time period. It is believed that the drastic decrease at the beginning of July was due to the Fourth of July holiday celebrated in the United States.
Results Analysis and Interpretation of the Fitness Tweets

Figure 4-4 - Daily Tweets by Apps
4.3 Sub-classifications of the Fitness Tweet Model

An additional breakdown of each main classification category is presented in Figure 4-5.

![Fitness Tweet Classification](image)

When analysing all five mobile fitness apps as one group, there were a total of 1,446,462 Activity tweets, 1,037,888 (72%) of which were classified as Workout and 408,574 (28%) as Workout+. Of the Activity fitness tweets, more than 76,192,059 minutes of exercise were shared via the five mobile apps via Twitter, equalling more than 145 years of physical activity.

When analysing all five mobile fitness apps as one group, 99% of all Blarney tweets were classified as Pointless Babble and 1% as Spam. Collectively, 7,786 tweets (2%) of all Conversation tweets were classified as Technical Support, 9,429 (2%) as Corporate Marketing, 25,587 (6%) as Statement of Support and 377,801 (90%) as Information Sharing.

After filtering by the specific mobile fitness app, Figure 4-6 shows the breakdown of the number of fitness tweets by major category.
The overall percentages of each mobile fitness app were as follows:

**DailyMile**
- 76.15% Activity
- 0.25% Blarney
- 23.59% Conversation

**Endomondo**
- 77.01% Activity
- 0.09% Blarney
- 22.90% Conversation

**MyFitnessPal**
- 61.67% Activity
- 0.08% Blarney
- 61.67% Conversation

**Nike+**
- 87.72% Activity
Results Analysis and Interpretation of the Fitness Tweets

- 0.27% Blarney
- 12.00% Conversation

RunKeeper
- 72.63% Activity
- 12.54% Blarney
- 14.84% Conversation

The striking volume of Blarney within RunKeeper prompted additional analysis. After a manual review of these tweets, it was determined that one single Twitter user was responsible for over 95% of these Blarney-Spam tweets. It appears that this Twitter account was a Twitter bot that was an automatic reply to a person’s tweets that offered congratulations for a world ranking of RunKeeper participants.

For the major category of Activity, of all the mobile fitness apps, 41% were from RunKeeper, 25% from Nike+, 16% from Endomondo, 13% from DailyMile and 5% from MyFitnessPal (Figure 4-7).

![Classification of Fitness Tweets by Sub-Category](image)

Figure 4-7 - Cross App Comparison of Sub-category

For the major category of Blarney, 98% came from RunKeeper, 1% from DailyMile, 1% from Nike+ and less than 1% from each MyFitnessPal and Endomondo. For the major category of Conversation, 29% of all tweets came from MyFitnessPal, 29% from RunKeeper, 16% from Endomondo, 14% from DailyMile and 12% from Nike+.
4.4 Additional Insights into Activity Tweets

Based on the type of information collected, it can be expected that a majority of the Activities shared using mobile fitness apps through Twitter were of a more structured exercise type, as opposed to continuous monitoring of daily physical activity. This is possibly due to the additional battery drain on the smartphone of the user, which would preclude day-long usage of the app. In addition, the structure of the tweets would also suggest that these activities were measured in terms of duration, suggesting activities such as a run, walk, bike or traditional workout. Other technologies, such as wearables, offer the ability to provide day-long measurement of physical activity that more closely measures the true physical activity of a user. The same sharing of physical activity can be collected from wearables, as shown in this research through mobile fitness apps.

Because of the nature of some of the Activity tweets, it was possible to extract additional information, including the actual type, distance and the amount of time spent on an activity. It was possible for outliers to be present in the database. For example, testing of the mobile fitness app often prompted an Activity tweet with a very short-duration activity (seconds rather than minutes), while very long-duration activities were sometimes recorded for activities when the person did not properly end his or her mobile fitness app activity session. It was possible that some of the longer-duration activities were, in fact, long exercise sessions. For example, a person training for a marathon would track long runs.

From the 165,768 users that posted Activity using a mobile fitness app that was then shared via Twitter, the database included 76,192,059 minutes of activity over the six-month time period. These minutes are equivalent to 52,911 days, 1,738 months or more than 145 years of combined Activity minutes. This equated to an average of 462 minutes of activity per user over the six-month time period or an average of 17 minutes of activity per week. It is understood that users completed physical activity without using their mobile fitness app.

Based on the research data, the number of one-time users of a mobile fitness app that shared their workout using Twitter (Activity tweets) was calculated. A number of reasons could exist for one-time use, including user error, experimentation of sharing functionality or testing by a user choosing a mobile fitness app. While the research cannot determine if a person continued to use a mobile fitness app and decided not to share via Twitter, it was determined that of all users, between 17 and 27% used the sharing to Twitter feature only once (Figure 4-8).
Results Analysis and Interpretation of the Fitness Tweets

Figure 4-8 - Percentage of Users Who only Shared Activity Once via Twitter

4.5 Breakdown of Fitness Tweets from Mobile Fitness Apps

An examination was conducted across each of the five mobile fitness apps to compare and contrast information such as demographics, usage patterns, individual user patterns and individual active user patterns.

4.5.1 DailyMile

With regard to overall fitness tweets from the five selected mobile fitness apps for English-language tweets during the period of this research, DailyMile ranked 4th in total fitness tweets, 4th in Activity tweets (194,401), 3rd in Blarney (644) and 4th in Conversation (60,205) (Figure 4-9).
Specific to Activity tweets for users of DailyMile who shared their workouts via Twitter, there were 8,308 users accounting for a total of 194,401 tweets. To achieve a greater analysis of DailyMile, the top 10% of the most active Activity fitness tweets were also analysed. These 1,355 users accounted for a total of 120,363 tweets. The top 10% of DailyMile users accounted for 61.9% of Activity tweets and made up 16.3% of the total user base. A comparison of all users and the top 10% of all users are presented in Table 4-1.

<table>
<thead>
<tr>
<th></th>
<th>All Users</th>
<th>Top 10%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>23.40</td>
<td>88.82</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>8</td>
<td>75</td>
</tr>
<tr>
<td><strong>Mode</strong></td>
<td>1</td>
<td>58</td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>34.95</td>
<td>40.76</td>
</tr>
<tr>
<td><strong>Minimum</strong></td>
<td>1</td>
<td>48</td>
</tr>
<tr>
<td><strong>Maximum</strong></td>
<td>372</td>
<td>372</td>
</tr>
</tbody>
</table>

Table 4-1 - Statistical Analysis of DailyMile Fitness Tweets

The most popular day for a DailyMile user to fitness tweet was August 22, 2011, with 2,170 tweets. The least popular day for a DailyMile user to fitness tweet was September 11, 2011, with 860 tweets (Figure 4-10).
Regarding the Top 10 users, the @dailymile Twitter account was the 3\textsuperscript{rd} most active. Most of the fitness tweets were either Activity or Conversation (Table 4-2).

\begin{center}
\begin{tabular}{|c|c|c|c|}
\hline
\textbf{User} & \textbf{Total Tweets} & \textbf{Activity} & \textbf{Blarney} & \textbf{Conversation} \\
\hline
DM-User1 & 630 & 265 & 0 & 365 \\
DM-User2 & 540 & 186 & 0 & 354 \\
dailymile & 530 & 0 & 0 & 530 \\
DM-User4 & 476 & 372 & 0 & 104 \\
DM-User5 & 398 & 173 & 0 & 225 \\
DM-User6 & 390 & 151 & 2 & 237 \\
DM-User7 & 380 & 226 & 3 & 151 \\
DM-User8 & 374 & 100 & 1 & 273 \\
DM-User9 & 366 & 236 & 1 & 129 \\
DM-User10 & 361 & 284 & 0 & 77 \\
\hline
\end{tabular}
\end{center}

Table 4-2 - DailyMile Top 10 Users

\subsection*{4.5.2 MyFitnessPal}

With regard to overall fitness tweets from the five selected mobile fitness apps for English-language tweets during the period of this research, MyFitnessPal ranked 5\textsuperscript{th} in total fitness tweets, 5\textsuperscript{th} in Activity tweets (76,278), 5\textsuperscript{th} in Blarney (165) and 1\textsuperscript{st} in Conversation (122,970) (Figure 4-11). It is the opinion of this researcher that the highest ranking in Conversation is due to the design of the sharing options.
within the mobile fitness app, the strong online social network created around the MyFitnessPal app and the sharing of diet and nutrition information along with the sharing of physical activity within the app.

![Breakdown of MyFitnessPal Fitness Tweets](image)

**Figure 4-11 - MyFitnessPal Breakdown of Fitness Tweets**

Specific to Activity tweets for users of MyFitnessPal who shared their workouts via Twitter, there were 4,874 users accounting for a total of 76,285 tweets. To achieve a greater analysis of MyFitnessPal, the top 10% most active Activity fitness users were analysed. There were 1,685 users accounting for a total of 62,066 tweets. The top 10% of MyFitnessPal users accounted for 81.4% of Activity tweets and made up 34.6% of the total user base. A comparison of all users and the top 10% of all users are presented in Table 4-3.

<table>
<thead>
<tr>
<th></th>
<th>All Users</th>
<th>Top 10%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>21.43</td>
<td>36.83</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>7</td>
<td>28</td>
</tr>
<tr>
<td><strong>Mode</strong></td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>15.65</td>
<td>24.89</td>
</tr>
<tr>
<td><strong>Minimum</strong></td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td><strong>Maximum</strong></td>
<td>172</td>
<td>172</td>
</tr>
</tbody>
</table>

Table 4-3 - Statistical Analysis of MyFitnessPal Fitness Tweets

80
The most popular day for a MyFitnessPal user to fitness tweet was September 21, 2011, with 1,624 tweets. The least popular day for a MyFitnessPal user to fitness tweet was July 1, 2011, with 860 tweets (Figure 4-12). It is surmised that the least popular day recorded as July 1st is due to the 4th of July holiday, with July 1st of that year being on a Friday.

![MyFitnessPal Time Period Analysis by Date](image)

Figure 4-12 - MyFitnessPal Time Period Analysis

Regarding the top 10 users, most of the fitness tweets were combined Activity or Conversation tweets (Table 4-4).

<table>
<thead>
<tr>
<th>User</th>
<th>Total Tweets</th>
<th>Activity</th>
<th>Blarney</th>
<th>Conversation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFP-User1</td>
<td>417</td>
<td>0</td>
<td>43</td>
<td>374</td>
</tr>
<tr>
<td>MFP-User2</td>
<td>345</td>
<td>335</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>MFP-User3</td>
<td>329</td>
<td>294</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>MFP-User4</td>
<td>321</td>
<td>311</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>MFP-User5</td>
<td>319</td>
<td>297</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>MFP-User6</td>
<td>314</td>
<td>314</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MFP-User7</td>
<td>313</td>
<td>296</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>MFP-User8</td>
<td>305</td>
<td>305</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MFP-User9</td>
<td>303</td>
<td>262</td>
<td>0</td>
<td>41</td>
</tr>
<tr>
<td>MFP-User10</td>
<td>301</td>
<td>298</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 4-4 - MyFitnessPal Top 10 Users

Note: myfitnesspal had the 765th most total tweets
4.5.3 Endomondo

With regard to overall fitness tweets from the five selected mobile fitness apps for English-language tweets during the period of this research, Endomondo ranked 3\(^{rd}\) in total fitness tweets, 3\(^{rd}\) in Activity tweets (228,801), 4\(^{th}\) in Blarney (267) and 3\(^{rd}\) in Conversation (67,834) (Figure 4-13).

**Figure 4-13 - Endomondo Breakdown of Fitness Tweet**

Specific to Activity tweets for users of Endomondo that shared their workouts via Twitter, there were 14,116 users accounting for a total of 228,081 tweets. To achieve a greater analysis of Endomondo, the top 10\% of the most active Activity fitness users were analysed. There were 1,979 users accounting for a total of 128,341 tweets. The Top 10\% of Endomondo users accounted for 56.3\% of Activity tweets and made up 14.0\% of the total user base. A comparison of all users and the top 10\% of all users are presented in Table 4-5.

<table>
<thead>
<tr>
<th></th>
<th>All Users</th>
<th>Top 10%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>16.16</td>
<td>64.85</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>7</td>
<td>50</td>
</tr>
<tr>
<td><strong>Mode</strong></td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>27.29</td>
<td>46.58</td>
</tr>
<tr>
<td><strong>Minimum</strong></td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td><strong>Maximum</strong></td>
<td>975</td>
<td>975</td>
</tr>
</tbody>
</table>

Table 4-5 - Statistical Analysis of Endomondo Fitness Tweets
Results Analysis and Interpretation of the Fitness Tweets

The most popular day for an Endomondo user to fitness tweet was September 26, 2011, with 2,379 tweets. The least popular day for an Endomondo user to fitness tweet was July 1, 2011, with 754 tweets. Similar to the analysis of MyFitnessPal, it is surmised that the least popular day recorded as July 1st is due to the 4th of July holiday, with July 1st of that year being on a Friday (Figure 4-14).

![Endomondo Time Period Analysis by Date](image)

Figure 4-14 - Endomondo Time Period Analysis by Date

Regarding the top 10 users, the @Endomondo Twitter account was the 2nd most active user. Most of the fitness tweets (other than @Endomondo) were Activity tweets (Table 4-6).

<table>
<thead>
<tr>
<th>User</th>
<th>Total Tweets</th>
<th>Activity</th>
<th>Blarney</th>
<th>Conversation</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN-User1</td>
<td>976</td>
<td>975</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Endomondo</td>
<td>750</td>
<td>0</td>
<td>16</td>
<td>734</td>
</tr>
<tr>
<td>EN-User3</td>
<td>500</td>
<td>500</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EN-User4</td>
<td>364</td>
<td>364</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EN-User5</td>
<td>342</td>
<td>342</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EN-User6</td>
<td>362</td>
<td>332</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>EN-User7</td>
<td>332</td>
<td>332</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EN-User8</td>
<td>501</td>
<td>302</td>
<td>0</td>
<td>199</td>
</tr>
<tr>
<td>EN-User9</td>
<td>302</td>
<td>302</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EN-User10</td>
<td>279</td>
<td>279</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4-6 - Endomondo Top 10 Users
4.5.4 Nike+

With regard to overall fitness tweets from the five selected mobile fitness apps for English-language tweets during the period of this research, Nike+ ranked 2\textsuperscript{nd} in total fitness tweets, 2\textsuperscript{nd} in Activity tweets (355,899), 2\textsuperscript{nd} in Blarney (1,108) and 5\textsuperscript{th} in Conversation (48,701) (Figure 4-15).

![Breakdown of Nike+ Fitness Tweets](image)

Specific to Activity tweets for users of Nike+ that shared their workouts via Twitter, there were 40,524 users accounting for a total of 355,899 tweets. To achieve a greater analysis of Nike+, the top 10\% most active Activity fitness users were analyzed. There were 5,425 users accounting for a total of 195,261 tweets. The top 10\% of Nike+ users accounted for 54.9\% of Activity tweets and made up 13.4\% of the total user base. A comparison of all users and the top 10\% of all users are presented in Table 4-7.

<table>
<thead>
<tr>
<th></th>
<th>All Users</th>
<th>Top 10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>8.78</td>
<td>35.99</td>
</tr>
<tr>
<td>Median</td>
<td>4</td>
<td>29</td>
</tr>
<tr>
<td>Mode</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>13.82</td>
<td>21.14</td>
</tr>
<tr>
<td>Minimum</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>Maximum</td>
<td>380</td>
<td>380</td>
</tr>
</tbody>
</table>

Table 4-7 - Statistical Analysis of Nike+ Fitness Tweets
The most popular day for a Nike+ user to fitness tweet was May 24, 2011, with 3,472 tweets. The least popular day for a Nike+ user to fitness tweet was July 1, 2011, with 1,118 tweets (Figure 4-16).

Regarding the Top 10 users, the @Nike+ Twitter account was the most active. Most of the fitness tweets (other than @Nike+) were of Activity tweets (Table 4-8).

<table>
<thead>
<tr>
<th>User</th>
<th>Total Tweets</th>
<th>Activity</th>
<th>Blarney</th>
<th>Conversation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nike+</td>
<td>6643</td>
<td>0</td>
<td>2</td>
<td>6641</td>
</tr>
<tr>
<td>NI-User2</td>
<td>380</td>
<td>380</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NI-User3</td>
<td>321</td>
<td>321</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NI-User4</td>
<td>303</td>
<td>303</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NI-User5</td>
<td>270</td>
<td>242</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>NI-User6</td>
<td>241</td>
<td>241</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NI-User7</td>
<td>234</td>
<td>234</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NI-User8</td>
<td>229</td>
<td>229</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NI-User9</td>
<td>207</td>
<td>0</td>
<td>0</td>
<td>207</td>
</tr>
<tr>
<td>NI-User10</td>
<td>197</td>
<td>197</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4-8 - Top 10 Nike+ Users

4.5.5 RunKeeper

With regard to overall fitness tweets from the five selected mobile fitness apps for English-language tweets during the period of this research, RunKeeper ranked 1st in total fitness tweets, 1st in Activity tweets (591,803), 1st in Blarney (102,176) and 2nd in Conversation (120,893) (Figure 4-17).
Specific to Activity tweets for users of RunKeeper who shared their workouts via Twitter, there were 48,767 users accounting for a total of 355,899 tweets. To achieve a greater analysis of RunKeeper, the top 10% most active Activity fitness tweets were analysed. There were 7,095 users accounting for a total of 343,592 tweets. The top 10% of RunKeeper users accounted for 58.1% of Activity tweets and made up 14.5% of the total user base. A comparison of all users and the top 10% of all users are presented in Table 4-9.

<table>
<thead>
<tr>
<th></th>
<th>Total Users</th>
<th>Top 10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>16.16</td>
<td>48.43</td>
</tr>
<tr>
<td>Median</td>
<td>5</td>
<td>39</td>
</tr>
<tr>
<td>Mode</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>19.56</td>
<td>29.96</td>
</tr>
<tr>
<td>Minimum</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>Maximum</td>
<td>377</td>
<td>377</td>
</tr>
</tbody>
</table>

Table 4-9 - Statistical Analysis of RunKeeper Fitness Tweets

The most popular day for a RunKeeper user to fitness tweet was August 31, 2011, with 6,103 tweets. The least popular day for a RunKeeper user to fitness tweet was April 21, 2011, with 860 tweets (Figure 4-18).
Regarding the top 10 users, the @runkeeper Twitter account was the 2nd most active. Most of the fitness tweets were from one user (world_rankin) and classified as Spam (Table 4-10).

<table>
<thead>
<tr>
<th>User</th>
<th>Total Tweets</th>
<th>Activity</th>
<th>Blarney</th>
<th>Conversation</th>
</tr>
</thead>
<tbody>
<tr>
<td>world_rankin</td>
<td>100519</td>
<td>0</td>
<td>100514</td>
<td>5</td>
</tr>
<tr>
<td>runkeeper</td>
<td>1360</td>
<td>0</td>
<td>1</td>
<td>1359</td>
</tr>
<tr>
<td>RK-User3</td>
<td>377</td>
<td>377</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RK-User4</td>
<td>701</td>
<td>346</td>
<td>0</td>
<td>355</td>
</tr>
<tr>
<td>RK-User5</td>
<td>417</td>
<td>339</td>
<td>0</td>
<td>78</td>
</tr>
<tr>
<td>RK-User6</td>
<td>336</td>
<td>336</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RK-User7</td>
<td>313</td>
<td>313</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RK-User8</td>
<td>305</td>
<td>305</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RK-User9</td>
<td>301</td>
<td>301</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RK-User10</td>
<td>286</td>
<td>286</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4-10 - Top 10 RunKeeper Users

4.6 Demographic Characteristics of the Workout+ Users

Since the data for this dissertation was from publicly available Twitter data, an additional inference could be made with regards to overall demographics of the data set including, but not limited to, gender, age, marriage status, income level, profession, popular followers, and religion. To infer demographic characteristics of the Twitter users within the data set, an online tool called Demographics Pro was used.
Results Analysis and Interpretation of the Fitness Tweets

This tool uses a series of proprietary algorithms to estimate the likely demographic characteristics of Twitter handles on the basis of Twitter behavior/usage such as age, gender, race/ethnicity, marital status, income and occupation based upon the user’s social media usage. Their methodology is data-centric, relying on multiple data signals from three different areas: language, consumption, and network. The data is filtered through large knowledge bases of established correlations between data points and demographic characteristics and has been used to infer over 300 million characteristics requiring confidence of over 95% to make an estimate (Demographics Pro, 2014).

To descriptively compare the inferred demographics of Twitter users with those who tweeted about their physical activity using a mobile fitness app, a sample of 14,836 randomly selected English-language Twitter users were chosen from the existing Fitness Tweet Data Set of 165,756 fitness tweet users. The random sample was based on a 95% confidence level, with one confidence interval for each of the five mobile fitness apps in relation to the overall data set.

4.6.1 Demographics of tweeters

The inferred characteristics using the Demographics Pro analysis of the random sample of those that tweeted a Workout+ tweet versus the average demographics of Twitter users (as provided by Demographics Pro) is presented in Table 4-11.

Users in the Fitness Tweet Data Set are in their early thirties, typically married with children and have a high income. The group includes a notable concentration in London.

- **Professionally**, people in this group work as programmers, photographers, church leaders, designers, and teachers. The group has a notably high concentration of web developers (within the top 10% of the overall Twitter distribution in this respect).
- **In their spare time**, they particularly enjoy beer, political news, wine, comedy/humor, and cooking. People in this group are charitably generous and particularly health-conscious. Sports that rise most notably above the Twitter norm include cycling, skiing, and golf.
- **On Twitter**, they tweet most often about sports, apps and TV/film. Accounts followed far more than the Twitter average include: @shitmydadsays, @donttrythis, @Seesmic, and @AppStore.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Fitness Twitter Users Average</th>
<th>Twitter Average</th>
<th>% Change from Twitter to Fitness Twitter average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>67.9</td>
<td>46.5</td>
<td>46.2%</td>
</tr>
<tr>
<td>Female</td>
<td>32.1</td>
<td>53.5</td>
<td>-40.1%</td>
</tr>
<tr>
<td>Single</td>
<td>9.9</td>
<td>50.5</td>
<td>-80.4%</td>
</tr>
<tr>
<td>Married</td>
<td>90.1</td>
<td>49.6</td>
<td>81.7%</td>
</tr>
<tr>
<td>Parents</td>
<td>51.2</td>
<td>13.1</td>
<td>292.1%</td>
</tr>
</tbody>
</table>
Results Analysis and Interpretation of the Fitness Tweets

### Age

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Fitness Twitter Users Average</th>
<th>Twitter Average</th>
<th>% Change from Twitter to Fitness Twitter average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 16 and under</td>
<td>0.3</td>
<td>6.2</td>
<td>-95.2%</td>
</tr>
<tr>
<td>Age 17 to 19</td>
<td>1.4</td>
<td>32.8</td>
<td>-95.7%</td>
</tr>
<tr>
<td>Age 20 to 24</td>
<td>8.2</td>
<td>32.3</td>
<td>-74.6%</td>
</tr>
<tr>
<td>Age 25 to 29</td>
<td>23.6</td>
<td>15.6</td>
<td>51.0%</td>
</tr>
<tr>
<td>Age 30 to 34</td>
<td>31.3</td>
<td>5.5</td>
<td>471.2%</td>
</tr>
<tr>
<td>Age 35 to 39</td>
<td>18.9</td>
<td>3.0</td>
<td>525.8%</td>
</tr>
<tr>
<td>Age 40 to 49</td>
<td>13.8</td>
<td>3.4</td>
<td>304.7%</td>
</tr>
<tr>
<td>Age 50 to 59</td>
<td>1.7</td>
<td>0.7</td>
<td>161.5%</td>
</tr>
<tr>
<td>Age 60 and over</td>
<td>0.7</td>
<td>0.5</td>
<td>45.8%</td>
</tr>
</tbody>
</table>

### Income

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Fitness Twitter Users Average</th>
<th>Twitter Average</th>
<th>% Change from Twitter to Fitness Twitter average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under $10,000</td>
<td>5.5</td>
<td>53.5</td>
<td>-89.7%</td>
</tr>
<tr>
<td>$10,000 - $19,999</td>
<td>8.4</td>
<td>21.3</td>
<td>-60.5%</td>
</tr>
<tr>
<td>$20,000 - $29,999</td>
<td>15.4</td>
<td>14.1</td>
<td>9.3%</td>
</tr>
<tr>
<td>$30,000 - $39,999</td>
<td>19.2</td>
<td>5.9</td>
<td>228.2%</td>
</tr>
<tr>
<td>$40,000 - $49,999</td>
<td>21.9</td>
<td>2.9</td>
<td>660.4%</td>
</tr>
<tr>
<td>$50,000 - $74,999</td>
<td>25.7</td>
<td>2.1</td>
<td>1118.0%</td>
</tr>
<tr>
<td>$75,000 - $99,999</td>
<td>3.5</td>
<td>0.2</td>
<td>1844.4%</td>
</tr>
<tr>
<td>Over $100,000</td>
<td>0.5</td>
<td>0.1</td>
<td>400.0%</td>
</tr>
</tbody>
</table>

### Ethnicity

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Fitness Twitter Users Average</th>
<th>Twitter Average</th>
<th>% Change from Twitter to Fitness Twitter average</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>93.1</td>
<td>76.4</td>
<td>21.8%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>4.8</td>
<td>6.7</td>
<td>-28.3%</td>
</tr>
<tr>
<td>African American</td>
<td>1.4</td>
<td>16.3</td>
<td>-91.4%</td>
</tr>
<tr>
<td>Asian</td>
<td>0.7</td>
<td>0.6</td>
<td>16.7%</td>
</tr>
<tr>
<td>Christian</td>
<td>85.2</td>
<td>93.4</td>
<td>-8.8%</td>
</tr>
<tr>
<td>Jewish</td>
<td>11.4</td>
<td>2.8</td>
<td>314.5%</td>
</tr>
<tr>
<td>Muslim</td>
<td>3.4</td>
<td>3.8</td>
<td>-10.8%</td>
</tr>
</tbody>
</table>

Table 4-11 - Demographics of Workout+ Tweeters
Figures 4-19 through 4-23 provide various demographic breakdowns of the Workout+ tweeters. An analysis of these breakdowns appears in the following figures.
Results Analysis and Interpretation of the Fitness Tweets

Figure 4.20 - Relationship Demographics

Figure 4.21 - Age Demographics
This analysis would suggest the profile of the typical mobile fitness app user who shares workouts with comment is male, white, married with children, between the ages of 30 and 49 and has an income over $50,000.
Results Analysis and Interpretation of the Fitness Tweets

When compared to the average Twitter dataset, the following characteristics would suggest higher distributions:

Top 10% of overall Twitter distribution:
- Age 30–34 (31.3%)
- Age 35–39 (18.9%)
- Parents (51.2%)
- Personal Income $30,000–$39,000 (19.2%)
- Personal Income $40,000–$49,000 (21.9%)

Top 20% of overall Twitter distribution:
- Male (67.9%)
- Married (90.1%)
- Age 25–29 (23.6%)
- Personal Income $50,000–$74,000 (25.7%)
- Personal Income $75,000–$99,999 (3.5%)
- Personal Income over $100,000 (0.5%)

Tables 4-12 through 4-18 present the top content followed or discussed by Workout+ tweeters. This additional information about the users of these mobile fitness apps may shed additional insights into the type of individual who shares their physical activity using social networks such as Twitter. By identifying these additional insights, researchers and health promotion practitioners may be in a position to better personalize the proper messages to these individuals in the hopes of making a long-term change in physical activity and wellness.

<table>
<thead>
<tr>
<th>Following</th>
<th>Fitness Twitter Users Average</th>
<th>Twitter Average</th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barack Obama</td>
<td>27.8</td>
<td>20.4</td>
<td>36.5%</td>
</tr>
<tr>
<td>Conan O’Brien</td>
<td>22.6</td>
<td>4.0</td>
<td>469.3%</td>
</tr>
<tr>
<td>Shitmydadssays</td>
<td>20</td>
<td>1.1</td>
<td>1783.4%</td>
</tr>
<tr>
<td>CNN News</td>
<td>19.2</td>
<td>7.1</td>
<td>170.8%</td>
</tr>
<tr>
<td>Jimmy Fallon</td>
<td>16.4</td>
<td>5.3</td>
<td>207.0%</td>
</tr>
<tr>
<td>Pete Cashmore</td>
<td>16.1</td>
<td>2.1</td>
<td>655.4%</td>
</tr>
<tr>
<td>Twitter</td>
<td>16</td>
<td>13.1</td>
<td>22.1%</td>
</tr>
<tr>
<td>Neil Patrick Harris</td>
<td>15.9</td>
<td>3.0</td>
<td>431.3%</td>
</tr>
<tr>
<td>The Onion</td>
<td>15.3</td>
<td>1.8</td>
<td>740.8%</td>
</tr>
</tbody>
</table>

Table 4-12 - Top 10 Twitter Accounts Followed by Workout+ Tweeters
Results Analysis and Interpretation of the Fitness Tweets

<table>
<thead>
<tr>
<th>Topic</th>
<th>Fitness Twitter Users Average</th>
<th>Twitter Average</th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>#RunKeeper</td>
<td>13.7</td>
<td>0.1</td>
<td>18692.9%</td>
</tr>
<tr>
<td>#running</td>
<td>5.2</td>
<td>0.2</td>
<td>3259.2%</td>
</tr>
<tr>
<td>#FitnessAlerts</td>
<td>3.2</td>
<td>0.1</td>
<td>5903.8%</td>
</tr>
<tr>
<td>#fitness</td>
<td>2.9</td>
<td>0.6</td>
<td>354.5%</td>
</tr>
<tr>
<td>#family</td>
<td>2.7</td>
<td>1.7</td>
<td>58.8%</td>
</tr>
<tr>
<td>#thanksgiving</td>
<td>2.4</td>
<td>0.3</td>
<td>855.0%</td>
</tr>
<tr>
<td>#runchat</td>
<td>2.3</td>
<td>0.1</td>
<td>4247.8%</td>
</tr>
<tr>
<td>#USMNT</td>
<td>2.3</td>
<td>0.1</td>
<td>2543.7%</td>
</tr>
<tr>
<td>#GameOf Thrones</td>
<td>2.3</td>
<td>0.5</td>
<td>367.8%</td>
</tr>
<tr>
<td>#art</td>
<td>2.3</td>
<td>1.4</td>
<td>66.8%</td>
</tr>
</tbody>
</table>

Table 4-13 - Top 10 Topics Discussed by Workout+ Tweeters

<table>
<thead>
<tr>
<th>Brand</th>
<th>Fitness Twitter Users Average</th>
<th>Twitter Average</th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starbucks</td>
<td>39.7</td>
<td>13.1</td>
<td>201.9%</td>
</tr>
<tr>
<td>McDonald's</td>
<td>24.6</td>
<td>11.7</td>
<td>110.3%</td>
</tr>
<tr>
<td>Apple Store</td>
<td>23.3</td>
<td>7.4</td>
<td>213.9%</td>
</tr>
<tr>
<td>Runner's World</td>
<td>22.6</td>
<td>0.4</td>
<td>6338.7%</td>
</tr>
<tr>
<td>CNN</td>
<td>22.5</td>
<td>12.6</td>
<td>78.1%</td>
</tr>
<tr>
<td>Google</td>
<td>21.4</td>
<td>10.7</td>
<td>100.1%</td>
</tr>
<tr>
<td>Instagram</td>
<td>21.2</td>
<td>16.5</td>
<td>28.5%</td>
</tr>
<tr>
<td>Walmart</td>
<td>19.2</td>
<td>5.4</td>
<td>253.1%</td>
</tr>
<tr>
<td>BBC</td>
<td>18.2</td>
<td>11.1</td>
<td>64.7%</td>
</tr>
<tr>
<td>Nike</td>
<td>17.8</td>
<td>5.8</td>
<td>207.3%</td>
</tr>
</tbody>
</table>

Table 4-14 - Top 10 Brands Followed by Workout+ Tweeters

<table>
<thead>
<tr>
<th>Likes and Interests</th>
<th>Fitness Twitter Users Average</th>
<th>Twitter Average</th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>News</td>
<td>40.9</td>
<td>29.5</td>
<td>38.8%</td>
</tr>
<tr>
<td>Technology</td>
<td>38</td>
<td>13.7</td>
<td>176.7%</td>
</tr>
<tr>
<td>Comedy</td>
<td>33.7</td>
<td>15.5</td>
<td>117.5%</td>
</tr>
<tr>
<td>Politics</td>
<td>26.4</td>
<td>20.3</td>
<td>30.2%</td>
</tr>
</tbody>
</table>
Results Analysis and Interpretation of the Fitness Tweets

<table>
<thead>
<tr>
<th>Fitness</th>
<th>21.8</th>
<th>1.0</th>
<th>2139.6%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycling</td>
<td>20.2</td>
<td>2.6</td>
<td>686.3%</td>
</tr>
<tr>
<td>Basketball</td>
<td>17.3</td>
<td>11.9</td>
<td>45.6%</td>
</tr>
<tr>
<td>Sports</td>
<td>16.6</td>
<td>9.3</td>
<td>77.9%</td>
</tr>
<tr>
<td>Music</td>
<td>15.1</td>
<td>18.8</td>
<td>-19.8%</td>
</tr>
<tr>
<td>Fashion</td>
<td>14.3</td>
<td>11.8</td>
<td>21.1%</td>
</tr>
</tbody>
</table>

Table 4-15 - Top 10 Likes and Interests of Workout+ Tweeters

<table>
<thead>
<tr>
<th>Profession</th>
<th>Fitness Twitter Users Average</th>
<th>Twitter Average</th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senior managers</td>
<td>8.8</td>
<td>6.4</td>
<td>36.5%</td>
</tr>
<tr>
<td>Web developers</td>
<td>8.4</td>
<td>2.3</td>
<td>258.7%</td>
</tr>
<tr>
<td>Sales/marketing</td>
<td>7.8</td>
<td>4.9</td>
<td>60.4%</td>
</tr>
<tr>
<td>Students</td>
<td>7.6</td>
<td>12.2</td>
<td>-37.6%</td>
</tr>
<tr>
<td>Journalists</td>
<td>6.9</td>
<td>5.0</td>
<td>38.3%</td>
</tr>
<tr>
<td>Consultants</td>
<td>5.2</td>
<td>2.4</td>
<td>115.5%</td>
</tr>
<tr>
<td>Teachers</td>
<td>5.2</td>
<td>3.0</td>
<td>76.2%</td>
</tr>
<tr>
<td>Entrepreneurs</td>
<td>5.1</td>
<td>3.0</td>
<td>71.7%</td>
</tr>
<tr>
<td>Authors/writers</td>
<td>4.6</td>
<td>4.1</td>
<td>12.6%</td>
</tr>
<tr>
<td>Musicians</td>
<td>4.3</td>
<td>8.2</td>
<td>-47.3%</td>
</tr>
</tbody>
</table>

Table 4-16 - Top 10 Professions of Workout+ Tweeters

<table>
<thead>
<tr>
<th>Country</th>
<th>Fitness Twitter Users Average</th>
<th>Twitter Average</th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>58.4</td>
<td>24.5</td>
<td>138.4%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>12.5</td>
<td>2.7</td>
<td>356.6%</td>
</tr>
<tr>
<td>Netherlands</td>
<td>4.3</td>
<td>0.3</td>
<td>1166.9%</td>
</tr>
<tr>
<td>Canada</td>
<td>4</td>
<td>1.1</td>
<td>252.2%</td>
</tr>
<tr>
<td>Japan</td>
<td>3</td>
<td>0.1</td>
<td>2284.7%</td>
</tr>
</tbody>
</table>
Results Analysis and Interpretation of the Fitness Tweets

<table>
<thead>
<tr>
<th>Country</th>
<th>Fitness Twitter Users Average</th>
<th>Twitter Average</th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>2.8</td>
<td>0.4</td>
<td>569.1%</td>
</tr>
<tr>
<td>Sweden</td>
<td>1.4</td>
<td>0.1</td>
<td>1708.8%</td>
</tr>
<tr>
<td>Brazil</td>
<td>1.3</td>
<td>0.7</td>
<td>98.2%</td>
</tr>
<tr>
<td>Spain</td>
<td>1.2</td>
<td>0.3</td>
<td>327.4%</td>
</tr>
<tr>
<td>Mexico</td>
<td>1</td>
<td>0.4</td>
<td>162.7%</td>
</tr>
</tbody>
</table>

Table 4-17 - Top 10 Countries of Origin of Workout+ Tweeters

<table>
<thead>
<tr>
<th>City</th>
<th>Fitness Twitter Users Average</th>
<th>Twitter Average</th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>London</td>
<td>4.7</td>
<td>1.7</td>
<td>177.1%</td>
</tr>
<tr>
<td>New York</td>
<td>3.2</td>
<td>1.9</td>
<td>64.8%</td>
</tr>
<tr>
<td>Tokyo</td>
<td>2.3</td>
<td>0.1</td>
<td>2950.4%</td>
</tr>
<tr>
<td>Chicago</td>
<td>2.1</td>
<td>0.5</td>
<td>315.8%</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>1.7</td>
<td>1.5</td>
<td>14.0%</td>
</tr>
<tr>
<td>Stockholm</td>
<td>1.6</td>
<td>0.1</td>
<td>2447.8%</td>
</tr>
<tr>
<td>Boston</td>
<td>1.5</td>
<td>0.2</td>
<td>590.0%</td>
</tr>
<tr>
<td>Toronto</td>
<td>1.5</td>
<td>0.3</td>
<td>360.4%</td>
</tr>
<tr>
<td>San Francisco</td>
<td>1.3</td>
<td>0.2</td>
<td>762.1%</td>
</tr>
<tr>
<td>Atlanta</td>
<td>1.3</td>
<td>0.4</td>
<td>262.7%</td>
</tr>
</tbody>
</table>

Table 4-18 - Top 10 Cities of Origin of Workout+ Tweeters

4.7 Conclusion

This chapter presented the analysis and interpretation of the fitness tweets from the five mobile fitness apps used in this dissertation research. An analysis of the demographics of mobile fitness app users and of the fitness tweet classifications, as well as a comparison of mobile fitness app users to the average Twitter user, was provided. In addition, insights into Activity fitness tweets, analysis of each of the five mobile fitness app social were introduced. Finally, insights into the general demographics of Workout+ Users were presented, giving the ability to generalize and compare between different segments of the Twitter population, as well as the general population.
5 EXPERIMENTAL APPROACHES IN USING FITNESS TWITTER DATA

This chapter presents three experimental approaches used from the collected fitness tweets, and by so doing introduces future research opportunities resulting from this academic work. Each experiment used different subsets of the total fitness tweet database, thus the total number of tweets analyzed was dependent upon the research criteria.

These experiments include:

1. Analysis of fitness tweets to determine the impact of influence and gender
2. Sentiment analysis of the Workout+ sub-classification of fitness tweets
3. Analysis of fitness tweets to examine the differences in data collection used for Healthy People 2020 and to suggest a new and innovative way to collect such data

Each experimental approach has been published verbatim as follows:

Section 5.1 has been published as: Vickey, T., & Breslin, J. (2017). Online Influence and Sentiment of Fitness Tweets: Analysis of Two Million Fitness Tweets. JMIR Public Health Surveillance, 3(4), e82. https://doi.org/10.2196/publichealth.8507

Section 5.2 has been published as: Vickey, T., & Breslin, J. (2017). Online Influence and Sentiment of Fitness Tweets: Analysis of Two Million Fitness Tweets. JMIR Public Health Surveillance, 3(4), e82. https://doi.org/10.2196/publichealth.8507

Section 5.3 has been published as: Vickey, T., & Breslin, J. (2015). Do As I Tweet, Not As I Do: Comparing Physical Activity Data Between Fitness Tweets and Healthy People 2020. mHealth, 1(9), 1–7. https://doi.org/10.3978/j.issn.2306-9740.2015.11.01
5.1 Online Influence and Sentiment of Fitness Tweets: Analysis of Two Million Fitness Tweets

**Background:** Publicly available fitness tweets may provide useful and in-depth insights into the real-time sentiment of a person’s physical activity and provide motivation to others through online influence.

**Objective:** The goal of this experimental approach using the Fitness Twitter Dataset is two-fold: (1) to determine if there is a correlation between the type of activity tweet (either workout or workout+), calculated gender, and one’s online influence as measured by Klout Score and (2) to examine the sentiment of the activity-coded fitness tweets by looking at real-time shared thoughts via Twitter regarding their experiences with physical activity and the associated mobile fitness app.

**Methods:** The Fitness Tweet Dataset includes demographic and activity data points, including minutes of activity, Klout Score, classification of each fitness tweet, the first name of each fitness tweet user, and the tweet itself. Gender for each fitness tweet user was determined by a first name comparison with the US Social Security Administration database of first names and gender.

**Results:** Over 184 days, 2,856,534 tweets were collected in 23 different languages. However, for the purposes of this study, only the English-language tweets were analyzed from the activity tweets, resulting in a total of 583,252 tweets.

After assigning gender to Twitter usernames based on the Social Security Administration database of first names, analysis of minutes of activity by both gender and Klout influence was determined. The mean Klout Score for those who shared their workout data from within four mobile apps was 20.50 (13.78 SD), less than the general Klout mean of 40, as was the Klout Score at the 95th percentile (40 vs. 63). As Klout Score increased, there was a decrease in the number of overall workout+ tweets. With regards to sentiment, fitness-related tweets identified as workout+ reflected a positive sentiment toward physical activity by a ratio of 4 to 1.

**Conclusions:** The results of this research suggest that the users of mobile fitness apps who share their workouts via Twitter have a lower Klout Score than the general Twitter user and that users who chose to share additional insights into their workouts are more positive in sentiment than negative. We present a novel perspective into the physical activity messaging from within mobile fitness apps that are then shared over Twitter. By moving beyond the numbers and evaluating both the Twitter user and the emotions tied to physical activity, researchers are able to determine relationships between the user’s online influence, the enjoyment of physical activity, and the likelihood of continued use of a fitness app.
5.1.1 Introduction

Physical activity can reduce the risk for many different types of chronic diseases and can help people maintain a healthy weight. Although this knowledge is widely known, adults and children in many countries do not get recommended amounts of physical activity (Lewis, Napolitano, Buman, Williams, & Nigg, 2017). Recent advances in physical activity monitoring now provide researchers with unparalleled opportunities to increase and improve our understanding of the health benefits of physical activity by assessing daily quantities of activity, patterns, and trends (Schrack et al., 2016), as well as the real-time sentiment of physical activity. Research suggests that technology is one factor that has contributed to the increase in sedentary behavior and decrease in physical activity, but it has also led to a number of innovative physical activity interventions (Lewis et al., 2017).

One such innovation is through the use of mobile fitness apps and the sharing of one’s workout through a social network. This paper will focus on the collection of self-reported fitness data through a mobile fitness app that is then shared with one’s social network via Twitter. The dataset of these tweets along with other connected datasets of demographic information allows for a number of analyses, including but not limited to the potential influence of such tweets and the sentiment of these tweets. By combining the digital traces as people interact through mobile phones and emerging technology may now provide novel methods to assess a range of factors objectively and with minimal expense and burden to participants (Schootman et al., 2016). This paper will review both the potential online influence and the sentiment of the shared fitness tweets.

Social media has changed the way society is exposed to information (Garimella, Weber, Garimella, & Weber, 2017). Social networking sites such as Twitter have developed into increasingly useful platforms for the general public to share thoughts, ideas, and opinions. Twitter is a free social networking platform that is widely used around the world by businesses and individuals and is considered one of the most widely used microblogging platforms with 328 million monthly active users with more than 1 billion unique monthly visits to sites with embedded tweets with a mission to “To give everyone the power to create and share ideas and information instantly, without barriers” (Twitter, 2017b). Twitter users can rapidly and directly share with and respond to a massive audience, using messages of 140 characters or less. With the creation and introduction of newly developing technologies such as Twitter, new opportunities to obtain global health data that may circumvent the limitations of traditional data sources used in population health and physical activity research are now available (Schootman et al., 2016).

At the same time, these publicly shared data are resulting in vast and growing user-contributed repositories of data (Arias, Arratia, & Xuriguera, 2013). Twitter provides user-generated data that can be collected and analyzed to examine opinions around health-related foci, including discussions about physical activity, alcohol and marijuana use, depression, and suicide (Schootman et al., 2016). From a
health-promotion standpoint, these data can be useful in to measure participants’ dependence on social
support, given that exercisers today are just as, if not more, likely to seek motivation and validation
from social media—in particular, Twitter—than their in-person friends and family members (Pagoto et al., 2014). Because it is possible to glean precise information from tweets, including the time of the
tweet and location of the user, this suggests that the 140-character messages could be predictive in other
areas, such as the types of physical activity that users engage in and where and when they engage in
these activities.

Using Twitter integration with mobile fitness apps can be a helpful tool for obtaining descriptive and
predictive real-time shared health information in a non-invasive way. New and innovative cloud-based
data collection and analysis tools may aid research efforts because they can yield a large collection of
tweets in a short period of time. They may also be useful for longitudinal data collection (Driscoll &
Walker, 2014). The link between publicly available health and fitness data sources is made possible as
more users publicly share their self-collected data from devices and apps through social media services
such as Twitter (Wang, Weber, & Mitra, 2016). An enhanced understanding of mobile fitness apps and
the sharing of physical activity through one’s social network, the different types of measurement
properties, and the subsequently generated data are critical to furthering our understanding of daily
physical activity.

Sentiment analysis is a classification process, the primary focus of which is to predict the polarity of
words and to then classify these words as positive, negative, or neutral with the aim of identifying
attitude and opinions (Khan, Bashir, & Qamar, 2014). Specific to Twitter, sentiment analysis is the task
of automatically identifying and extracting subjective information from tweets. This method of data
analysis has received increasing attention from the Web-mining community (Bravo-Marquez,
Mendoza, & Poblete, 2014). Although Twitter provides extremely valuable insight into publicly shared
opinions, it also provides new big data challenges, including the processing of massive volumes of data
and the identification of human expressiveness within short text messages (Bravo-Marquez et al., 2014).
Much of the existing research on textual information processing has been focused on the mining and
retrieval of factual information, with little research on the processing of opinions (Liu, 2010).

The mining of Twitter for data provides a rich database of information on people’s thoughts and
sentiments about a myriad of health topics, including physical activity. Analysis of social networks data
using Twitter has become a powerful tool that is currently being used to answer research questions
across the health spectrum, including local and national flu surveillance (Broniatowski, Paul, & Dredze,
2013), the sharing of information between cancer patients (Tsuya, Sugawara, Tanaka, & Narimatsu,
2014), marijuana usage among teens (Cavazos-Rehg et al., 2015), and drug safety surveillance (Freifeld
et al., 2014). This paper represents, to the best of our knowledge, the first analysis of shared tweets from
mobile fitness apps specific to physical activity. A significant proportion of tweets contained nonneutral sentiments regarding the shared physical activity of the four mobile apps featured in this research.

The ability to evaluate the sentiment of an individual immediately after a bout of physical activity has been completed can be powerful. A typical tweet might include the type of exercise performed, the duration and intensity of that exercise, and how the person felt during and after the activity. If the sentiment is negative (e.g., “Just hiked to the top of Mt. Pisgah. Took me 2 hours and I’m completely exhausted. Don’t think I’ll do that again! #myfitnesspal”), a coach or trainer can intervene and modify the activity accordingly. Finding exercise that is enjoyable and of the appropriate intensity is an important precursor to long-term adherence. Behavioral researchers suggest that one’s emotions can profoundly affect individual behavior and decision making (Bollen, Mao, & Zeng, 2011). Simply stated, a tweet can be a window into real emotion provided in real time.

Other research reported that when fitness promoters initiated a #PlankADay challenge on Twitter—which was designed to encourage core-strengthening exercise—72% of users participated for at least 30 days straight and at the end of the challenge reported an increased enjoyment of the activity and expressed interest in continuing to do abdominal exercise (Pagoto, Schneider, Oleski, Smith, & Bauman, 2014). This indicates that Twitter and other social networks can be useful in spreading exercise awareness and encouraging positive exercise behaviors. Together, this information can facilitate research on how technology can be used to monitor and motivate physical activity and how online social networks may play a role in physical activity promotion and adherence. Identifying the types of people who use mobile fitness apps and finding ways to track what they do and motivate them to continue to engage in physical activity is a form of data mining for this “customer base.”

5.1.2 Collection of Tweets

After a review of online tools that could collect and manage tweets, an open-source program called TwapperKeeper was deemed appropriate as the Twitter data-collection tool. TwapperKeeper is a Web app designed to collect social media data via Twitter for long-term archival and analysis. The app uses a Twitter-enabled application program interface (API) that acts as an interface between the Twitter search function and a cloud database for tweet storage (Vickey, Breslin, & Tsai, 2011).

For this research, we chose four mobile fitness apps based on their availability on iPhone, the ability of the mobile fitness app to share workout information through Twitter, and the fact that they targeted beginner versus experienced exercisers. The research team used these criteria to narrow possible choices and reviewed additional academic research for previously used apps, researched publicly available reviews on different mobile fitness apps, interviewed both developers and users of mobile fitness apps to obtain their input, and met as a group to finalize the selected mobile fitness apps to study (Vickey et al., 2013).
Experimental Approaches in Using Fitness Twitter Data

The four apps chosen were Endomondo, Nike+, RunKeeper, and DailyMile. Tweets were then collected from the mobile fitness apps using the following hashtags: #endomondo, #nikeplus, #runkeeper, and #dailymile. These were used because these apps automatically attach these hashtags to a tweet to indicate it has come from that particular mobile fitness app. It is through these hashtags that common themes or information can be grouped within Twitter.

Data collection using TwapperKeeper continued for 184 days. During this period, 2,856,534 user-generated mobile fitness app tweets were collected in 23 different languages. The Twitter data in this study was public, and the research was deemed exempt from human subjects’ review. This research was approved by the National University of Ireland Galway in Galway Ireland Research Ethics Committee.

Two analyses were completed on a dataset of collected tweets from four mobile fitness apps. The first was to measure the online influence of Twitter users through their Klout Score. The second was to measure the sentiment of physical activity related tweets. The Twitter data in this research were public, and our research was deemed exempt from human subjects’ review by the National University of Ireland at Galway Research Ethics Committee.

5.1.3 Measuring Online Influence

One important factor to consider when analyzing tweets to report physical activity is the credibility and authority of the person sending the tweets. Previous data collectors have looked at a Twitter user’s number of followers, although researchers discovered that monitoring retweets and the messages themselves are a better predictive tool (Hansen, Shneiderman, & Smith, 2009).

Websites such as Klout have developed the means to determine a user’s reach or influence on social media. The Klout Score is the measurement of a person’s overall online influence, with scores ranging from 1 to 100; higher scores represent a wider and stronger sphere of influence. Scores greater than 50 are rare (Lassen & Brown, 2011). A Klout Score places less emphasis on a user’s number of followers and number of tweets but rather measures the extent to which the user’s content is retweeted (Quercia et al., 2011). One’s influence on Twitter can be difficult to measure accurately. Klout uses more than 3600 features that capture the online social network activity of the user to conduct the influence analysis and calculate the Klout Score (Zong, 2016). The Klout Score allows for tailored statistical analysis of social media usage and is tangible proof of the effect of the Internet on a person’s lifestyle (Barbieri et al., 2017). With regards to influence, Internet users perceived a mock Twitter page with a high Klout Score as more credible than the same page with a moderate or low Klout Score (Edwards, Spence, Gentile, Edwards, & Edwards, 2013).

Online influence services such as Klout are in the process of scoring millions, eventually billions, of people on their level of influence. To proponents, the measurement of online influence is an inspiring
tool that encourages the democratization of influence, where one no longer must be a celebrity, politician, or media personality to be considered influential (Rosenbloom, 2011).

For this experimental approach, the user’s Klout Score—a measure of their online influence—was used to compare shared physical activity levels from mobile fitness apps.

In this experiment, we examined the sharing of fitness tweets from within mobile fitness apps (Nike+, RunKeeper, DailyMile, and Endomondo) and analyzed the data based on the participant’s gender and online influence, as measured by their Klout Score. We identified two types of activity tweets from dataset: workout tweets, which included what was generated by the mobile fitness app, and workout+ tweets, which included the same information as a workout tweet but also contained user-created communication. We hypothesized that those with a higher Klout Score would share fewer minutes of activity and more overall workout+ tweets. We also hypothesized that across both genders, the higher the Klout Score, the lower the minutes of shared physical activity.

The data for this research was drawn from an existing dataset of fitness tweets from mobile fitness app users who shared their physical activity and, in some cases additional conversation, over Twitter. Over 184 days, 2,856,534 tweets were collected in 23 different languages. However, for the purposes of this study, only the English-language tweets were analyzed from the activity tweets, resulting in a total of 583,252 tweets.

The Fitness Tweet Classification Model (Vickey et al., 2013) was used to classify each tweet into main categories of activity, blarney, and conversation and then into subcategories as shown in Figure
The different types of collected information from the mobile fitness apps and each corresponding Twitter account provided a number of different and unique data points to review. For this experiment, those data points included the activity tweets, the user’s gender, the minutes of physical activity, and the user’s Klout Score. The statistical analysis of physical activity on Twitter from the four selected mobile fitness apps was performed in SAS 9.3, a software suite developed by the SAS Institute for advanced analytics, business intelligence, and predictive analytics, using two key datasets: (1) the first dataset included all user information about Twitter users who sent tweets relating to workout and workout+ and (2) the second dataset contained all the actual tweets sent by each user.

5.2 Sentiment Analysis
Of the activity tweets, there were a total of 408,574 workout+ tweets. From this total, a random sample of 23,391 was created. These tweets were user-generated, where the end user was providing additional text to a workout tweet (i.e., the user provided supplementary information beyond that which is created by the app itself). Tweets were then grouped by mobile fitness app using the corresponding hashtags. There were no significant numbers of emojis available in the fitness tweets for use in the sentiment analysis.
Experimental Approaches in Using Fitness Twitter Data

The AYLIEN Text Analysis for Google Sheet add-on was utilized for the analysis of the sentiment for each collected information-sharing conversation tweet as filtered by the Fitness Tweet Classification Model.

The AYLIEN Tweet Sentiment Analysis function is a three-step process:

1. Pre-processing: tweets are normalized and reformatted, and the parts that are considered irrelevant to the sentiment are stripped;
2. Parsing: tweets are parsed, and their structure, tags, and negations are extracted;
3. Classification: tweets are classified as positive, negative, or neutral by a pre-trained classifier, assisted by a lexicon-based approach as a second judge.

For this experiment, the sentiment analysis tool that analyzed each tweet and returned the value of positive, neutral, or negative was used for classification. These data were saved into an Excel spreadsheet for additional data processing by converting the text value to a numerical value (positive=1, neutral=0, and negative=−1).

5.2.1 Analysis 1: Measuring Online Influence

Twitter does not collect the gender of users. To be able to compare across genders, a means of identifying the possible gender of the Twitter users was needed. To accomplish this, we used the US Social Security Administration’s name database to match English names with gender. The name database from the Social Security Administration website included popular names ranked by gender since 1880.

The first gender-match calculation between the first names in the collected Twitter demographic database (the Twitter user’s full name was one of the many demographic characteristics collected from Twitter) and the Social Security Administration database eliminated names that were used fewer than 200 times because many such names were much more popular among one gender than another (e.g., girls were named Aaron <0.5% of the time). The assumption was that this adjustment eliminated a vast majority of gender confusion among names. Once this was completed, names were matched to genders using the VLOOKUP function in Excel.

A second gender-match calculation was performed for those Twitter users with names that appeared less than 200 times, in which we attempted to assign gender to the remaining names that did not match in the first round. Usernames that did not match either gender (<2%) were not included in the analysis.

After gender assignment, a descriptive statistical analysis was performed to compute the frequency of the following: (1) total minutes by gender, (2) total minutes by Klout Score, (3) total minutes by gender and Klout Score, (4) total number of tweets, (5) minutes exercised per tweet, and (6) total number of workout and workout+ tweets (separately).
5.2.1.1 Determination of Klout Quartiles

To examine the distribution of tweets, minutes of exercise described by said tweets and the categories mentioned in each tweet (workout or workout+), it was necessary to separate the users’ Klout Scores into quartiles. We used the quartile method of data classification to create categories with a rank-ordered dataset split into four equal parts.

This was done through a two-step process in SAS. First, the distribution of Klout Scores was examined using the univariate procedure in SAS (PROC UNIVARIATE) and assigned quartiles based on that distribution. Second, using a data step, values of 1, 2, 3, and 4 were assigned to observations within the first, second, third, and fourth quartiles, respectively (Table 5-1). The maximum of any Klout Score is 100, and the minimum is 1. It was determined that the median Klout Score from the collected dataset was 20.50. As reported by Klout, the mean Klout Score is 40, with users with a score of 63 ranked in the 95th percentile (Klout, 2013).

<table>
<thead>
<tr>
<th>Quartile</th>
<th>Klout Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Maximum</td>
<td>100.00</td>
</tr>
<tr>
<td>99%</td>
<td>56.59</td>
</tr>
<tr>
<td>95%</td>
<td>49.03</td>
</tr>
<tr>
<td>90%</td>
<td>44.09</td>
</tr>
<tr>
<td>75% Q3</td>
<td>35.65</td>
</tr>
<tr>
<td>50% Median</td>
<td>20.50</td>
</tr>
<tr>
<td>25% Q1</td>
<td>11.92</td>
</tr>
<tr>
<td>10%</td>
<td>10.10</td>
</tr>
<tr>
<td>5%</td>
<td>10.00</td>
</tr>
<tr>
<td>1%</td>
<td>10.00</td>
</tr>
<tr>
<td>0% Minimum</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 5-1 - Klout Score quartiles

5.2.1.2 Number of Activity Tweets (Male Versus Female):

The descriptive statistical analysis found that males produced 57.9% of the total of activity tweets, whereas females produced 42.1%. This difference was consistent across Klout quartiles (Table 5-2).

<table>
<thead>
<tr>
<th>Quartile and Klout Score</th>
<th>Activity tweets, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male (n=336,109)</td>
</tr>
<tr>
<td>1: ≤11.92 (n=179,831)</td>
<td>102,007 (56.7)</td>
</tr>
<tr>
<td>2: &gt;11.93 and &lt;=20.50 (n=154,669)</td>
<td>89,822 (58.1)</td>
</tr>
<tr>
<td>3: &gt;20.51 and &lt;=35.65 (n=125,096)</td>
<td>73,394 (58.7)</td>
</tr>
<tr>
<td>4: &gt;35.65 (n=123,656)</td>
<td>70,886 (57.3)</td>
</tr>
</tbody>
</table>

Table 5-2 - Klout Score by activity tweet and gender (N=583,252)

5.2.1.3 Number of Tweets (Male/Female Among Workout Groups):

The descriptive analysis was expanded to compare males and females in the activity category. It was found that both genders tweeted far more among the workout group than the workout+ group (72.01%, vs. 27.99%) in the lowest Klout quartile. This trend decreased slightly through the second and third
Klout quartiles and then dramatically among the highest quartile of Klout Scores. In that quartile, the number of tweets varied much less (56.79% vs. 43.21%).

**Mean Minutes Per Tweet (Males Versus Females):**

The ANOVA procedure (PROC ANOVA) within SAS was used to compare the mean number of minutes tweeted by each gender using gender in the class statement and setting the model as minutes=gender. It was found that, overall, the mean number of minutes tweeted did not vary significantly between males and females. However, the mean number of minutes tweeted was almost double among females of the lowest Klout Score quartile (those with Klout ≤11.92).

**5.2.1.4 Determination of Activity Tweets by Klout Quartile**

After assigning quartiles, we examined the frequency of observations within each stratum of Klout Scores using PROC FREQ in SAS for the following (Table 5-3): (1) minutes by Klout Score quartile and (2) exercise types by Klout Score quartile.

<table>
<thead>
<tr>
<th>Quartile and Klout Score</th>
<th>Workout tweets (n=420,010)</th>
<th>Workout+ tweets (n=163,242)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tweets, n</td>
<td>Minutes (total)</td>
</tr>
<tr>
<td>1. ≤11.92</td>
<td>143,552</td>
<td>6,320,924</td>
</tr>
<tr>
<td>2. &gt;11.93 and ≤20.50</td>
<td>118,047</td>
<td>5,125,345</td>
</tr>
<tr>
<td>3. &gt;20.51 and &lt;35.65</td>
<td>88,182</td>
<td>4,348,112</td>
</tr>
<tr>
<td>4. &gt;35.65</td>
<td>70,229</td>
<td>2,897,436</td>
</tr>
</tbody>
</table>

*Table 5-3: Workout and workout+ tweets by Klout quartile*

**5.2.1.5 Tests of Significance Between Groups**

**Minutes Tweeted Between Workout Categories**

Also using the ANOVA procedure within SAS, analysis compared the total number of minutes tweeted among workout groups (workout vs. workout+) and found a statistically significant difference \( P=.01 \) (Table 5-4).
Table 5.4: Minutes exercised by gender and Klout Score among workout+ group.

There was no significant difference between males and females in the number of tweets for workouts (P=0.64).

There was no significant difference between mal- and females in the number of tweets for workout+ (P=0.55).

<table>
<thead>
<tr>
<th>Quartile and Klout Score</th>
<th>Male Tweets (total)</th>
<th>Minutes (total)</th>
<th>Minutes per tweet, mean (SD)</th>
<th>Female Tweets (total)</th>
<th>Minutes (total)</th>
<th>Minutes per tweet, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workout a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1: ≤ 11.92</td>
<td>81,503 (33.78)</td>
<td>3,528,992</td>
<td>43.33 (48.10)</td>
<td>62,049 (34.71)</td>
<td>2,791,932</td>
<td>45.00 (137.26)</td>
</tr>
<tr>
<td>2: &gt; 11.93 and ≤ 20.50</td>
<td>67,666 (28.05)</td>
<td>2,942,049</td>
<td>43.48 (56.45)</td>
<td>50,381 (28.18)</td>
<td>2,183,296</td>
<td>43.34 (76.06)</td>
</tr>
<tr>
<td>3: &gt; 20.51 and ≤ 35.65</td>
<td>51,863 (21.50)</td>
<td>2,811,512</td>
<td>54.21 (420.74)</td>
<td>36,319 (20.32)</td>
<td>1,536,600</td>
<td>42.33 (51.54)</td>
</tr>
<tr>
<td>4: &gt; 35.65</td>
<td>40,222 (16.67)</td>
<td>1,652,786</td>
<td>41.09 (49.08)</td>
<td>30,007 (16.79)</td>
<td>1,224,650</td>
<td>41.50 (61.62)</td>
</tr>
<tr>
<td>Workout+ b</td>
<td>94,855 (35.57)</td>
<td>4,437,573</td>
<td>46.79 (234.49)</td>
<td>68,387</td>
<td>3,220,919</td>
<td>47.10 (114.44)</td>
</tr>
<tr>
<td>1: ≤ 11.92</td>
<td>20,504 (21.62)</td>
<td>952,567</td>
<td>46.46 (114.94)</td>
<td>15,775 (23.07)</td>
<td>793,154</td>
<td>50.28 (144.89)</td>
</tr>
<tr>
<td>2: &gt; 11.93 and ≤ 20.50</td>
<td>22,156 (23.36)</td>
<td>1,002,024</td>
<td>45.24 (85.01)</td>
<td>14,466 (21.15)</td>
<td>664,973</td>
<td>45.97 (101.02)</td>
</tr>
<tr>
<td>3: &gt; 20.51 and ≤ 35.65</td>
<td>21,531 (22.70)</td>
<td>983,395</td>
<td>45.67 (112.10)</td>
<td>15,383 (22.49)</td>
<td>711,416</td>
<td>46.25 (98.06)</td>
</tr>
<tr>
<td>4: &gt; 35.65</td>
<td>30,664 (32.33)</td>
<td>1,499,587</td>
<td>48.90 (362.80)</td>
<td>22,763 (33.29)</td>
<td>1,051,375</td>
<td>46.19 (117.88)</td>
</tr>
</tbody>
</table>
5.2.2 Analysis 2: Sentiment Analysis of Workout+ Tweets

In total, there were 23,391 unique tweets within the original dataset that fit the filtering criteria from this random sample. Four of the mobile fitness apps were used in this analysis: DailyMile, Endomondo, Nike+, and RunKeeper. The overall sentiment of all mobile fitness apps suggests that half of these workout+ activity tweets were neutral in nature. In addition, there were four times as many positive tweets than negative. There were four times as many positive tweets than negative.

![Overall Percentage of Sentiment Analysis for All Mobile Fitness Apps](image)

Figure 5-2 - Overall percentage of sentiment by all mobile fitness apps
5.2.2.1 Overall Percentage of Sentiment Analysis for All Mobile Fitness Apps

The breakdown of sentiment analysis for negative, neutral, and positive sentiment by mobile fitness apps is as follows:

<table>
<thead>
<tr>
<th></th>
<th>DailyMile</th>
<th>Endomondo</th>
<th>Nike+</th>
<th>RunKeeper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Number of Tweets</td>
<td>9,298</td>
<td>820</td>
<td>3999</td>
<td>9284</td>
</tr>
<tr>
<td>Positive Sentiment</td>
<td>7097 (76.41%)</td>
<td>211 (25.73%)</td>
<td>418 (10.45%)</td>
<td>1,663 (17.91%)</td>
</tr>
<tr>
<td>Negative Sentiment</td>
<td>1392 (14.99%)</td>
<td>51 (6.22%)</td>
<td>350 (8.75%)</td>
<td>549 (5.91%)</td>
</tr>
<tr>
<td>Neutral Sentiment</td>
<td>799 (8.60%)</td>
<td>558 (68.05%)</td>
<td>3,231 (80.80%)</td>
<td>7,072 (76.17%)</td>
</tr>
</tbody>
</table>

Table 5-5 - Total number of tweets by sentiment and app

Figure 5-3 - Percentage of sentiment by mobile fitness app
5.2.2.2 The Sentiment of Fitness Tweets by Mobile Fitness App

In addition, the most common words in each mobile fitness workout+ tweet were reviewed. Sample word clouds for the combined mobile fitness apps are shown in Figure 5-4. The larger and bolder a word appears, the more it was used in the associated tweets. In addition to words denoting duration, distance, and intensity (e.g., hours, km, miles, and burned calories), users often used words such as “awesome,” “love,” and “yay” as indicators of how they felt following a workout, which would confirm the sentiment analysis.

Figure 5-4 - Word cloud of Workout+ Tweets

5.2.3 Analysis 1: Measuring Online Influence

This study further explored a novel approach to classify fitness tweets through Klout influence score. The study further stratified by gender through the use of a validated government database, which was probability matched to our data using exact matching procedures. This gender validation allowed for additional analysis of the gender breakdown of the existing dataset. The data were filtered through the matching criteria twice to improve precision, resulting in a 97% gender match. Although we gender-
matched twice, the process used to gender match could still be missing a few names that appear more often today than they did even a few years ago. Because popular names can change with high frequency, some gender matching in this study may not be valid within several years.

Based on the current database of collected fitness tweets from five mobile fitness apps, the highest Klout quartile included those individuals with a Klout Score of 35.65 or greater. Klout Scores can reach 100; therefore, our highest score tier may not capture an accurate representation of the most influential people on the Twitter platform. Additional insights from this research are described subsequently.

5.2.3.1 Men Share Their Physical Activity from Mobile Fitness Apps Via Twitter More Often Than Women

Based on this research, men share their workouts using Twitter and mobile fitness apps more often than women (54.35% vs. 45.65%). Although we believe this to be the first gender analysis of the sharing of physical activity from mobile fitness apps using Twitter, previous research on the overall gender use of Twitter suggests that more women than men use Twitter (Duggan, Greendwood, & Perrin, 2016) with some nonacademic research suggesting that 40 million more women use Twitter on a monthly basis, that 62% of Twitter users are women (McCandless, 2012), and those with a higher Klout Score tend to be women. Additional research regarding gender suggests that women are likely to be more active on Twitter as opposed to men, with women tweeting once every 20 hours versus men tweeting once every 26 hours (Hargittai & Litt, 2011).

However, additional research into our dataset using a third-party software called Demographics Pro suggests that the average mobile fitness app user in the fitness tweet dataset is a male in his early thirties, typically married with children and having a high income. Additional insights into the users of mobile fitness apps who also tweet their physical activity include that this group’s most common professions are programmers, photographers, church leaders, designers, and teachers. The group has a notably high concentration of Web developers (within the top 10% of the overall Twitter distribution in this respect). In their spare time, they particularly enjoy beer, political news, wine, comedy/humor, and cooking. People in this group are charitably generous and particularly health conscious. Sports that rise most notably above the Twitter norm include cycling, skiing, and golf. As a consumer, this group is relatively affluent, with spending focused most strongly on technology, wining/dining, and health/fitness. Their strongest brand affiliations include Apple Store, Trader Joe’s, CrossFit, Trek Bicycle, and MyFitnessPal.

5.2.3.2 The Design of the Mobile Fitness App and the Sharing of Physical Activity Data to Social Networking Sites Matters

The sharing of workout+ tweets is dramatically enhanced by the user interface of the mobile fitness app. When comparing the four mobile fitness apps for the total number of activity tweets (workout tweets plus workout+ tweets), the most popular mobile fitness app was Endomondo (211,240 tweets),
followed by NikePlus (203,991 tweets), DailyMile (183,732 tweets), and MyFitnessPal (70,723 tweets). The same usage ranking order was seen with men and women (men: 123,482 for Endomondo, 116,388 for NikePlus, 106,846 for DailyMile, and 70,723 for MyFitnessPal; women: 87,758 for Endomondo, 87,603 for NikePlus, 76,886 for DailyMile, and 30,233 for MyFitnessPal). However, there was a large difference when reviewing the workout+ tweets with 97.67% of all workout+ tweets from DailyMile, 1.7% from NikePlus, 0.4% from Endomondo, and no workout+ tweets from MyFitnessPal. In reviewing the user interface for all four mobile fitness apps, it is evident that the design of DailyMile made it much easier to share not only the workout but also additional information about the workout when compared to the other three mobile fitness apps. Also, during the evaluation time period for the activity tweets, Endomondo used a third-party service called @addthis to share workout+ tweets. With no workout+ tweets from MyFitnessPal, we determined that the app made a design decision not to allow users to share additional information regarding their physical activity workouts.

5.2.3.3 There is Brand Loyalty Regarding Mobile Fitness App Usage and the Sharing of Physical Activity Data Using Twitter

Of the 113,340 overall users in the dataset, 97.21% (110,186 users) tweeted their physical activity from just one mobile fitness app, 3,105 (2.74%) used two mobile fitness apps, with 101 (0.09%) users sharing from three mobile fitness apps and just one user (0.0009%) sharing from four mobile fitness apps. We base this on the analysis of tweets per users and cannot determine the actual usage of the app, only the sharing of physical activity data from the apps. We surmise one reason that more than 97% used just one app could be loyalty, but other reasons such as poor user interface and difficulty in connecting one’s Twitter account to the mobile fitness app may account for other reasons.

5.2.4 Analysis 2: Sentiment Analysis

A better understanding of the online influence of those who are sharing their fitness tweets may lead to new and innovative ways to encourage their followers to be more physically active through peer-to-peer influence, similar to programs created by marketing agencies to influence consumer behavior. Analogous to the other health-related research, physical activity researchers can monitor and attempt to influence physical activity Twitter chatter sent by influential Twitter users who are physically active and popular among various demographic groups and age ranges (Cavazos-Rehg et al., 2015). The findings can be used to inform online and offline efforts that work to target individuals who are most at risk for the harms associated with a lack of physical activity.

The relatively high number of neutral tweets was expected because each of the mobile fitness apps had a predetermined structure that limited additional information that could be included by the user. There also is the fact that a majority of the tweets simply did not contain words or phrases that could be classified as either positive or negative. Additional insights from this research are described subsequently.
5.2.4.1 The Real-Time Shared Sentiment of the Physical Activity Can Provide Additional Insights to Physical Activity

We believe that the sharing of one’s physical activity with additional commentary (for the purposes of this research called workout+ tweets) from mobile fitness apps can provide researchers with new insights that in the past may have been difficult to measure. The design of many of the mobile fitness apps allows for the user to share characteristics such as who they were with, the type of weather, the location of the physical activity, and their immediate thoughts regarding the physical activity. These and other insights will allow physical activity researchers to have a greater understanding of the real-time reasons, thoughts, and sentiment of how and perhaps why a person partakes in physical activity. This data will enable a greater understanding surrounding the complexities of physical activity, which can then be used for an enhanced design of mobile fitness apps as a potential tool in the decrease in physical inactivity.

5.2.4.2 Most Shared Mobile Fitness App Physical Activity Is of a Structured Exercise Type

It is through the analysis and interpretation that the context of fitness tweeting from within mobile fitness apps provides insights into what is being shared, by whom, and for what reasons. Based on the type of information collected, it can be expected that a majority of the activities shared using mobile fitness apps through Twitter were of a more structured exercise type, as opposed to continuous monitoring of daily physical activity. This is possibly due to the additional battery drain on the mobile phone of the user, which would preclude daylong usage of the app. In addition, the structure of the tweets would also suggest that these activities were measured in terms of duration, suggesting activities such as a run, walk, bike, or traditional workout. Because of the nature of some of the activity tweets, it was possible to extract additional information, including the actual type, distance, and the amount of time spent on an activity. It was possible for outliers to be present in the database. For example, the first use of a mobile fitness app could be the user testing the mobile fitness app that may have prompted an activity tweet with a very short-duration activity (seconds rather than minutes), whereas very long-duration activities were sometimes recorded for activities when the person did not properly end his or her mobile fitness app activity session. It was possible that some of the longer-duration activities were, in fact, long exercise sessions. For example, a person training for a marathon would track long runs.

5.2.4.3 A Significant Majority of Users from Each App Used the App More Than Once

Based on the research data, the number of one-time users of a mobile fitness app that shared their workout using Twitter (activity tweets) was calculated. Although the research cannot determine if a person continued to use a mobile fitness app and decided not to share via Twitter, it was determined that of all users, between 17% and 27% used the sharing to Twitter feature only once depending on the app. A number of reasons could exist for one-time use, including user error, experimentation of sharing functionality, or testing by a user choosing a mobile fitness app. From the 165,768 users that posted activity using a mobile fitness app that was then shared via Twitter, the database included 76,192,059
minutes of activity over the 6-month time period. These minutes are equivalent to 52,911 days, 1738 months, or more than 145 years of combined activity minutes. We cannot determine if this physical activity was the only performed physical activity by each user during the time period because it is understood that users may have completed physical activity without using their mobile fitness app.

These findings and interpretations should be regarded as exploratory and speculative because they represent what can be potentially done in a short development time and with ease of use for non-computer programming health-promotion researchers.

5.2.5 Limitations

There are a number of limitations to this research study. Utilizing outside data, in this case, the US government, to determine each user’s gender leaves room for error.

This research was conducted using the Twitter firehose, which allows for the collection of all publicly available tweets. Although we are confident in this data-collection process, there is no way to verify it without a financial expense to purchase all tweets. There also remains a challenge in the extraction of useful data from these repositories through data mining and knowledge discovery (Arias et al., 2013) due to a rapidly evolving explosion of data services and tools that can be used for analysis. This is due in large part to commercial pressures and the potential for using social networking data for computational research (Batrinca & Treleaven, 2014). To minimize this limitation, we were able to link different datasets using the user’s Twitter name as the unique identifier through free publicly available data. Future work could enhance our model by purchasing commercially available datasets for analysis.

There has been a steady growth in social media usage, from 5% of the US population in 2005 to close to 70% in 2015. As more Americans have adopted social media, the user base has also grown more representative of the broader population; however, it is still mostly used by younger age groups (Pew Research Center, 2017).

5.2.6 Comparison with Prior Work

The use of social media and emerging technologies to study physical activity and the possible lack thereof continues to increase with the development of such technologies. This type of physical activity research remains popular in the Quantified Self movement, in which life activities are tracked using on-body sensors in the hopes of a better understanding of human performance in various tasks (Gurrin, Smeaton, & Doherty, 2014). Previous research has shown an interest in specific characteristics of the social environments adversely affecting health outcomes (Schootman et al., 2016). Other research has studied the use of wearables and other smart devices to quantify various different health conditions with the self-reported data being shared on social networks, such as Facebook and Twitter (Wang, Weber, & Mitra, 2016), and have suggested that the adoption of such emerging technology to monitor physical
activity has created new research opportunities to observe, quantify, and define physical activity in the real-world setting (Schrack et al., 2016). Our research continues to build on these previous studies by providing researchers with other options for data collection and different objectives to consider.

Previous work regarding the role of technology in physical activity through social media includes a dearth of studies that have studied various aspects of the impact of social media on physical activity. Some research has focused on the behavior change challenges that include self-monitoring, goal setting, and problem-solving strategies (King et al., 2016). Other research has suggested a change in how we think about physical activity and sedentary behavior measurement, a research topic that includes the use of mobile fitness apps and social networks that can collect large amounts of real-time data that previously would have been difficult to collect (Kelly, Fitzsimons, & Baker, 2016). Research by Tsoh explores contextual and psychological factors that may underlie the observed low physical activity levels among mobile fitness app users. Our research is more closely related to that of Grundy et al. on the network analysis of prominent health and fitness apps and work by Haddadi et al. on the integration of shared health and fitness data from mobile fitness apps that are shared over social networks. Although these works are highly relevant to the research presented in this paper, we expand the research by carrying out data analysis including gender and online influence.

Similar approaches to inferring gender include works using a gender-based dictionary (Liu & Ruths, 2013), through profile picture and background inference (Alowibdi, Buy, & Yu, 2013), and a third-party Web service that can often reveal gender through proprietary algorithms (An & Weber, 2016). Specific research on using social media networks and physical activity include work by Althoff et al. on the influence of Pokémon Go, the tweeting of physical activity as a possible method to increase physical activity by Tsoh, and work by Liu & Young on using social media data analysis for physical activity surveillance.

5.2.7 Future Work

We created a very powerful tool for conducting large-scale research by collecting physical activity data from Twitter, but the demographics used in this research could suggest a bias regarding the breakdown of mobile fitness app users and thus underrepresent certain groups. If researchers wish to use Twitter and mobile fitness apps for physical activity research, additional steps would need to be taken to ensure that all groups are represented in the data samples collected. Apart from technical limitations, there could be ethical challenges that are equally as challenging. Although tweets are considered public, they may contain information that many would consider “private” due to the possible misconception of the perceived audience (a user’s Twitter followers) versus the actual audience (data researchers) (Wang et al., 2016). To expand on this work, additional investigation could address possible trends specific to forms of physical activity per gender that could constitute a higher Klout Score.
The popularity of consumer-facing health wearables (e.g., Fitbit, Garmin) that also share physical activity data with online social networks would be a topic worthy of future research. By using these tracking devices, which monitor physical activity on an ongoing basis, a more inclusive picture of daylong physical activity can be achieved. This contrasts with mobile fitness app data, which is typically collected and shared following a traditional “workout” (e.g., a walk, run, bike). The same data collection and classification model presented in this paper can be used with minimal changes. With regards to online influence, other work could use an alternate measure of online influence rather than Klout.

**Conclusion**

This research analyzed publicly shared physical activity data collected via Twitter from five different mobile fitness apps. From this dataset, two analysis of the data was conducted to highlight the unique ability to use this type of data within the study of physical activity. The first analysis categorized the users into four quartiles that represented their online influence as calculated by Klout as well as a method to assign gender to each Twitter user. The analysis suggests that men share their workout tweets more than women, that there is more basic sharing of physical activity data (workout tweets) when compared to tweets that also contain commentary by the user (workout + tweets) and that there is no significant difference in the tweeting of men and women. The second analysis was conducted with workout+ tweets and showed across all apps, most of the shared tweets were neutral, but that of those with a sentiment, there were four times as many positive tweets as negative.
5.3 Do as I Tweet, Not as I Do: Comparing Physical Activity Data between Fitness Tweets and Healthy People 2020

**Background:** The goal of this research was to compare the self-reported estimates of daily physical-activity data provided to the *Healthy People 2020* research team via a telephone survey to the mobile fitness app real-time reporting of physical activity using Twitter.

**Methods:** The Fitness Tweet Classification Data Set was collected from mobile fitness app users who shared their physical activity over Twitter. Over 184 days, 2,856,534 tweets were collected in 23 different languages. However, for the purposes of this study, only the English-language tweets were analyzed, resulting in a total of 1,982,653 tweets by 165,768 unique users. The information and data gleaned from this dataset, which reflected 184 days of continuous data collection, were compared to the results from the *Healthy People* survey, which were compiled using telephone interviews of self-reported physical activity from the previous week.

**Results:** The data collected from fitness tweets using the five mobile fitness apps suggest lower percentages of people achieving both the 150 to 300 minute and 300+ minute levels than is reflected in the *Healthy People* survey results. While employing Twitter and other social media as data-collection tools could help researchers obtain information that users might not remember or be willing to disclose face-to-face or over the telephone, further research is needed to determine the cause of the lower percentages found in this study.

**Conclusion:** Though some challenges remain in using social media like Twitter to glean physical-activity data from the public, this approach holds promise for yielding valuable information and improving outcomes.

5.3.1 Introduction

The promotion and monitoring of physical activity have been a focus of public health efforts in recent years. However, objectively measuring population-level physical activity is challenging because it requires tracking a large number of people using expensive devices and imposing strict data-collection protocols (Lim, Wyker, Bartley & Eisenhower, 2015). That said, emerging technology can provide reliable and valid alternative surveillance tools for self-reported measures of physical activity (Lim et

---

*Parts of this chapter were published as: Vickey, T., & Breslin, J. (2015). Do As I Tweet, Not As I Do: Comparing Physical Activity Data Between Fitness Tweets and Healthy People 2020. *mHealth*, 1(9), 1–7. https://doi.org/10.3978/j.issn.2306-9740.2015.11.01*
Experimental Approaches in Using Fitness Twitter Data

al., 2015). Although there is an increase in the number of studies using integrated sensor technology to collect physical activity data on a population level, there is little technical guidance for researchers who want to use this technology within their research (Hurvitz, Moudon, Kang, Saelens & Duncan, 2014).

According to Graham and Hipp (2014), “Physical activity measurement research is achieving greater ease of use, precision, and scope by incorporating emerging technologies. These emerging technologies are noteworthy because they can: (1) greatly increase external validity of measures and findings through ease of use and transferability; (2) significantly increase the ability to analyze patterns; (3) improve the ongoing, systematic collection and analysis of public health surveillance due to real-time capabilities; and (4) address the need for research about the cyberinfrastructure required to cope with big data.”

In 2010, the U.S. Department of Health and Human Services published the fifth installment of the national report on health and wellness, reflecting the strong state of the science supporting the health benefits of regular physical activity based on the accomplishments of previous Healthy People initiatives (Healthy People, 2010). The report, entitled Healthy People 2020, introduced new 10-year objectives for health promotion and disease prevention. New to the objectives is “myHealthyPeople,” a challenge for technology application developers. The research discussed here reflects an attempt to meet that challenge. The use of the Fitness Tweet Classification Model which was developed for this study enables researchers to collect ongoing data in real time, which is a sharp contrast to phone interviews that rely on participant recall. The biggest challenge in using technology to track physical activity lies in accounting for the fact that many users are inconsistent in their use of the tracking devices.

One component of Healthy People 2020 involves physical activity, suggesting that Americans should engage in at least 150 minutes per week of moderate-intensity physical activity to obtain substantial health benefits and more than 300 minutes per week to obtain more extensive health benefits.

Current baseline and targets are presented in Table 5-6.

<table>
<thead>
<tr>
<th>Baseline</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduce the proportion of adults who engage in no leisure-time physical activity</td>
<td>36.2%</td>
</tr>
</tbody>
</table>

119
Experimental Approaches in Using Fitness Twitter Data

| Increase the proportion of adults who engage in aerobic physical activity of at least moderate intensity for at least 150 minutes/week | 43.5% | 47.9% |
| Increase the proportion of adults who engage in aerobic physical activity of at least moderate intensity for more than 300 minutes/week | 28.4% | 31.3% |

Table 5-6 - Healthy People 2020 Baseline and Targets

In 2008, when the goals and objectives for Healthy People 2020 were first developed, 43.5% of American adults met the goal of 150 minutes per week of moderate-intensity physical activity, with only 28.4% reaching 300 minutes per week (Carlson et al., 2010).

5.3.2 Methods

For this research project, a comparison between the collected physical-activity data provided in the Healthy People 2020 report and physical-activity data collected from five mobile fitness apps (Nike+, DailyMile, MyFitnessPal, Endomondo, and RunKeeper) as publicly shared over Twitter was conducted.

Each mobile fitness app used in this research had a standard word phrase for the automatic sharing of physical activity using fitness tweets that include time and/or distance of the physical activity. In addition, some mobile fitness apps included in the standard word phrase a shortened URL that directed back to the mobile fitness app’s user page.

On that page, additional information not included in the fitness tweet could be collected (Figure 5-5). A data-scraping script was written to collect this information. Once the data were collected, a data cleaning removed low totals (less than 15 minutes of reported physical activity over 28 weeks) and high totals (more than 30,000 minutes of reported physical activity over 28 weeks) to account for one-time users or user error and invalid results stemming from technology-related issues (e.g., a fitness app being left open after a workout is completed, which would inflate the numbers and skew the data).
Data for this research was from two data sets:

1. Healthy People 2020
2. Fitness Tweet Classification Data Set

Results from the Healthy People survey were compiled using telephone interviews of self-reported physical activity from the previous week. There are considerable concerns about this methodology, as physical-activity questionnaires show limited reliability and validity (Shephard, 2003). Even so, they have long been considered the only feasible means of collecting data in large populations, even though researchers know that responses can be influenced by cultural factors, language barriers and recall accuracy, particularly in older populations (Shepard, 2003). One aim of this research study is to explore the use of Twitter as a more reliable and valid alternative.
The Fitness Tweet Classification Data Set was collected from mobile fitness app users who shared their physical activity over Twitter. Over 184 days, 2,856,534 tweets were collected in 23 different languages. However, for the purposes of this study, only the English-language tweets were analyzed, resulting in a total of 1,982,653 tweets by 165,768 unique users.

The Fitness Tweet Classification Model (Vickey et al., 2013) was used to classify each tweet into main categories of Activity, Blarney, and Conversation and then into subcategories as shown in Figure 5-6.

![Fitness Tweet Classification Model](image)

**Figure 5-6 - Fitness Tweet Classification Model**

### 5.3.4 Results

In total, 102,544 users mentioned workout duration in their tweets, accounting for 2.4 million minutes of physical activity. The addition of workout type, duration and distance allowed additional analysis to be conducted (Figure 5-7). Physical activity is a sporadic and complex behavior to measure, but previous research suggests that three days of accelerometer data, four days of pedometer data or four days of physical-activity logs are needed to reliably measure physical-activity levels in older adults (Hart & Swartz, 2011).
As demonstrated in Figure 5-7, the data collected from fitness tweets using the five mobile fitness apps suggest lower percentages of people achieving both the 150 to 300 minute and 300+ minute levels. The lower percentage for the 150 to 300-minute range was expected, as it is difficult to know whether a person used their mobile fitness app during every workout session. What was a bit more surprising was the lower percentages for the 300+ minute levels, as the more active population, might be expected to be more dedicated users of their wearable devices and mobile fitness apps.

Consider, for example, users of Nike+, which could be considered the most physically active overall group, as Nike targets athletic shoe buyers through social media channels. Analysis of the fitness tweet conversations also indicated a higher number of mobile fitness app users using Nike+ to train for 5K, 10K, half-marathon and full-marathon events. That being the case, the hypothesis was that of the five mobile fitness apps, Nike+ would be one of the more used mobile fitness apps within the 300+ minute category due to the daily training regimens of the participants. Figure 5-8 highlights the data analysis suggesting that weekly Nike+ user’s fitness tweet an average of 81 minutes per week of physical activity. In fact, RunKeeper, which is also geared toward runners, reported the second lowest average weekly minutes of physical activity, with just over 104 minutes per week.
The overall variance in the data derived from those who completed the Healthy People 2020 survey and mobile fitness app fitness tweeters could be due to users not sharing all their physical activity via Twitter and/or an overestimation of weekly minutes of exercise collected during the phone surveys for Healthy People 2020. Table 5-8 shows how one aspect of physical activity data collected from Twitter can be presented. To maintain the confidentiality of the users, Twitter usernames were replaced with generic ‘User xxx’ labels. It is important to determine the user’s first user date of the mobile fitness app within the data-collection period, as the data-collection timeframe is just a snapshot over time. A user could have already been using the mobile fitness app and sharing the data before the start of the data-collection period. Cells that contain the label “X” indicate that the first use date of the mobile fitness app by the user occurred after the week header. For example, the first use date for User 13 occurred sometime in Week 4. It is also important to be able to determine gaps of weekly usage over time, showing that a user is not consistent in the sharing of physical-activity data from mobile fitness apps using Twitter, or that the user simply did not exercise for a time due to injury, illness, vacation, etc.

<table>
<thead>
<tr>
<th>FromUser</th>
<th>FirstUseDate</th>
<th>TotalMin</th>
<th>Week1</th>
<th>Week2</th>
<th>Week3</th>
<th>Week4</th>
</tr>
</thead>
<tbody>
<tr>
<td>User 2345</td>
<td>20/07/2011</td>
<td>1408</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>User 5678</td>
<td>31/05/2011</td>
<td>926</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
5.3.5 Discussion

This case study presents a comparison between weekly minutes of physical activity derived from *Healthy People 2020* survey results and fitness activity tweets of mobile fitness app users and provides physical-activity researchers an alternative method of data collection that could be more reliable than self-reported physical-activity survey data. The issue of why this research yielded lower percentages of physical activity than the *Healthy People* phone survey remains unaddressed. Is it possible that the Twitter data is more accurate, and that people are over-reporting their activity levels over the phone? Can further research derive a means of accounting for any under-reporting that is taking place via Twitter? Recall bias is a considerable issue in phone surveys, as people tend to overestimate their physical activity and underestimate their sedentary time; thus, researchers have developed ways to account for this bias when analyzing the resulting data (Salas et al., 2012). This needs to be done for Twitter-based data as well, but ongoing, real-time data analysis is an invaluable resource for researchers that should eventually prove to be more reliable than recall-based phone surveys.

This active data collection could provide numerous benefits when compared to passive data collection. For example, some evidence already suggests that the knowledge that their activities are being monitored could impact participants’ weekly minutes of physical activity (Hart & Swartz, 2011). While this may be problematic in a research setting, as described above, it can lead to true lifestyle change in individuals who use social media to motivate themselves to stay on track.

Obtaining information from social media allows for crowdsourced participation, which can provide much more data diversity in terms of greater range of age, geography, and ethnicity of users. Moylan, Derr, and Lindhorst (2015) found that mobile technology was especially useful in reaching out to participants who were previously inaccessible due to geography or physical disability. Employing

<table>
<thead>
<tr>
<th>User</th>
<th>First Use Date</th>
<th>Total Minutes</th>
<th>Weekly Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>7/5/2011</td>
<td>834</td>
<td>X X X</td>
</tr>
<tr>
<td>445</td>
<td>19/08/2011</td>
<td>280</td>
<td>X X X</td>
</tr>
<tr>
<td>613</td>
<td>21/04/2011</td>
<td>183</td>
<td>X 84 22 30</td>
</tr>
<tr>
<td>6969</td>
<td>8/6/2011</td>
<td>167</td>
<td>X X</td>
</tr>
<tr>
<td>8675</td>
<td>24/04/2011</td>
<td>157</td>
<td>X X 0 0</td>
</tr>
<tr>
<td>5688</td>
<td>29/07/2011</td>
<td>140</td>
<td>X X X</td>
</tr>
<tr>
<td>8791</td>
<td>16/06/2011</td>
<td>134</td>
<td>X X X</td>
</tr>
<tr>
<td>415</td>
<td>29/04/2011</td>
<td>126</td>
<td>X X</td>
</tr>
</tbody>
</table>

Table 5-7 - User Tables with First Use Date, Total Minutes and Weekly Status

Experimental Approaches in Using Fitness Twitter Data
Twitter and other social media as data-collection tools could help researchers obtain information that users might not remember or be willing to disclose face-to-face or over the telephone. Ahlwardt et al. (2014) found that patients are often willing to reveal information about their personal healthcare experiences on Twitter, allowing healthcare providers to glean insight on how to improve communication with patients and treat them more effectively. Because users are often relaying information in real time, some researchers posit that the personal details users share may be more accurate than data collected by traditional methods.

Social media data collection also provides the added benefit of allowing researchers to access more people in their target population over a shorter amount of time (Rothman, Gnanaskathy, Wicks & Papadopoulos, 2015). Furthermore, Casler and College (2013) discovered that participants who signed up for studies online performed behavioral tasks just as well as people who participated face-to-face or over the phone. Information collected over social media has also provided additional useful healthcare data, from the presentation of menopause symptoms in women to the prevalence of children with ulcerative colitis. Healthcare practitioners can also access additional information through these methods, including demographics, current medication lists, and potential diagnoses. This could not only influence future research but also lead to improved patient experiences in real time (Rothman et al., 2015).

Limitations

A common limitation in this type of research lies in the fact that adolescents and adults do not always accurately report physical-activity levels (Chinapaw, Slootmaker, Schuit, Van & Mechelen, 2009), with many underreporting sedentary behaviors and over-reporting exercise (Salas et al., 2012).

As this research used all Twitter data (i.e., users were not assigned a specific mobile fitness app to share physical-activity data), one cannot assume that when a user reported zero minutes of weekly activity, this means that the user actually performed no physical activity during that period. There could have been any number of user, device, data collection or Twitter errors. Other than sending a tweet to each individual user, it would be difficult to determine the reason for the lack of data.

An additional limitation is the actual definition of physical activity. During the original data collection for the Healthy People project, depending on how the question was phrased, the respondent may have answered in one of two ways. First, he or she could have provided the number of minutes he or she performed traditional physical activity by going to the gym or going outside for a run, for example. Second, the respondent could have included all physical activity, including non-structured activity such as walking in a mall. The data set of the Fitness Tweets would suggest that this data is a collection of exercise-type activity rather than ongoing measurement, as such measurement would be difficult due to battery issues throughout the day.
While we are confident that during the data-collection process we had access to the Twitter firehose allowing for the collection of all publicly available tweets, there is no way to verify this without actually purchasing all of the tweets. There remains a challenge in the extraction of useful data within these repositories through data mining and knowledge discovery (Arias et al., 2013). While there has been an explosion of data services and tools for analysis, this research area is undergoing rapid change and evolution due to commercial pressures and the potential for using social networking data for computational research (Batrinca & Treleaven, 2014). To minimize this limitation, different data sets were linked utilizing the user’s Twitter name as the unique identifier through free publicly available data. Researchers could enhance our model by purchasing commercially available data sets for analysis in future studies.

While we created a very potent tool for large-scale research by collecting physical-activity data from Twitter, the demographics used in this research could suggest a bias in terms of the users of the mobile fitness apps and thus under-represent certain groups. If researchers wish to use Twitter and mobile fitness apps for physical-activity research, additional steps would need to be taken to ensure that all groups are represented in the data samples collected from Twitter.

These findings and interpretations should be regarded as exploratory and speculative, as they represent what can be potentially done in a short development time and with ease of use for non–computer programming health-promotion researchers.

**Future Work**

Advancements in technology design for both smartphones and wearables allow for continuous monitoring of physical activity without a drain in battery life. Depending on the sharing ability, physical activity could be measured by hour or even by minute, thus providing even greater detail of recorded physical activity.

One benefit of using the Fitness Tweet Classification Model was that the database included 184 days of continuous data collection, which stands in stark contrast to the one-week recall used in the *Healthy People* project. While not every subject had daily physical-activity measures, the same is true with the survey respondents in the *Healthy People* project. One future area of work could be the determination of how many days’ worth of fitness tweets would be needed to reliably measure physical activity.

Future research could also involve a study that uses fitness tweeting as a more effective data-collection tool, with participants understanding what is being measured and the need to share all physical-activity sessions—as opposed to passive data collection. Knowing they are being monitored could impact participants’ weekly minutes of physical activity, but perhaps not in longer-term studies. Because this type of research can be conducted on an ongoing basis, the phenomenon of study participants
outperforming their usual activity levels should dissipate over time as they return to their usual behavioral patterns.

5.3.6 Conclusion
Technologies currently used in other fields could be adopted for physical-activity measurement (Graham & Hipp, 2014). This research used one such technology—Twitter—and created a method to collect physical-activity data from publicly available tweets. The precise measurement of physical activity, including type, amount, context and place is essential for increasing physical activity (Hurvitz et al., 2014). While this approach shows promise in data collection, future research on how to account for user inconsistency in terms of reporting physical activity is needed before Twitter-based data can be considered truly reliable, but it is clear that Twitter, other forms of social media and smartphone apps are here to stay. Health and fitness professionals and researchers in this area would benefit from leveraging the ever-growing population of users in their work.

5.4 Summary
This chapter introduced three different uses of the Fitness Tweet Data Set as three separate experimental approaches.

The first experimental approach investigated the online influence of a user and the number of minutes of shared physical activity. The second approach was a sentiment analysis of workout tweets, giving a unique insight into the thoughts of the user immediately after completing his or her workout. The third approach considered the possibility of collecting physical-activity data, with a comparison of that data with data from the Healthy People project. These are just a few of many types of experimental approaches that can be conducted, not only from the existing Fitness Tweet Data Set but also future research into physical activity, wellness, and health promotion.
6 DISCUSSION

This chapter provides a summary of the research outcomes of this work. The chapter begins with a review of the three research questions. The chapter then discusses the results of the conducted experiments based on the Fitness Tweet Data Set. Finally, the chapter gives an overview of the contributions of this work from two perspectives: contributions to theory and implications for practice. This discussion focuses on the novelty of the Fitness Tweet Data Collection Tool and the Fitness Tweet Classification Model. This chapter also outlines the limitations of this research and provides an overview of future work.

This dissertation began with an introduction to the health aspects of physical activity in Chapter 1, and the technological aspects of mobile technology and social networks in Chapter 2. Chapter 3 reviewed related work with the use of social networks to share health information, online influence, the measurement of physical activity and the mining and text classification of data collected from Twitter. Chapter 4 introduced the determination of major classifications, the selection criteria used to determine the five mobile fitness apps to include in this research and the data-collection and data-processing tools created for this research. Chapter 5 presented the results of analysis from the Fitness Tweet Data Set and the interpretation of these results. Chapter 6 presented three different and unique experimental approaches to the use of the Fitness Tweet Data Set.

The following sections discuss the findings of this research regarding the three research questions posed:

RQ1: Can an automated data collector accurately identify fitness tweets shared from one’s mobile fitness app from the Twitter stream?
   - If so, how can these fitness tweets be collected and processed?
   - What are the limitations in the data processing of these fitness tweets?

RQ2: Can an automated fitness tweet classification model quantify minutes of physical activity?
   - If so, what is the process?
   - What are the limitations in the collection of physical-activity minutes using Twitter?

RS3: Can additional demographic information from those who share their workouts online using Twitter be generated to give additional insights into the character of the type of person whom fitness tweets?

6.1 Research Questions discussion

The first question (RS1) posed if an automated data collector accurately identify fitness tweets shared from one’s mobile fitness app from the Twitter stream, and if so how can these fitness tweets be
collected and processed and what are the limitations in the data processing of these fitness tweets? Chapter 4 begins to answer this question with the creation of the fitness tweet crawler and the classification model. Chapter 5 provides a result analysis and interpretation of the data showing that an automated data collector can accurately identify fitness tweets.

The second question (RS2) asked if the classification model could quantify minutes of physical activity. As seen in Chapter 5, it is possible for the classification model to pull the physical activity minutes of exercise from each fitness tweet but with additional coding, data mining from the user’s corresponding mobile fitness app website can provide even more data in addition to minutes. Future research can include multiple analysis from many different collected measurements including but not limited to time, heart rate, distance, and cadence.

The third and final research question (RS3) posed if additional information from those that shared their workouts online could be generated. As shown in Chapter 6, peer-reviewed published papers have shown that it is possible and of value to collect additional demographic information from users. This data, when used appropriately, can provide insights into the behaviors of active persons that will allow researchers to work together with app developers and government agencies to create more effective intervention strategies and allow future funding to be allocated to different research areas of wellness. The type of data collected can be a key asset to create and strengthen leadership, governance, multisectoral partnership and advocacy groups to achieve excellence in resource mobilization and the implementation of coordinated action plans around the world to both increase physical activity and reduce sedentary behaviors (World Health Organization, 2018).

6.2 Hypotheses discussion
Tackling the increase in physical inactivity continues to be a daunting and complicated task, and it remains a fundamental problem. Neither the wearing of a fitness sensor nor the tracking of one’s physical activity alone will change behavior. However, what can be measured can be changed. Mobile fitness apps and the sharing of data can effectively measure physical activity, allowing for the possible motivation for a change in behavior.

It was the hypothesis of this dissertation that the results of this research would allow for the extraction of valuable fitness data generated by mobile fitness apps through the use of Twitter. This dissertation has established that this hypothesis was true. Not only was an abundance of information available from a single 140-character tweet, but also through each user’s mobile fitness app website and his or her connections within an online social network.

The practical application of the tools developed as part of this research was explored in relation to the Healthy People 2020 report, the data for which was gathered via a telephone survey that relied on participant recall. To test whether data gathered in real time from Twitter and mobile fitness apps might
be an option for future large-scale physical-activity data collection, publicly available physical-activity data was collected via Twitter over 184 consecutive days after being shared from mobile fitness apps. This data-collection process created in 2011 for the sharing of workouts from mobile fitness apps can be adjusted to the real-time physical-activity-tracking wearable devices and apps that measure not only one’s traditional exercise sessions but also the physical activities of daily life like walking to work, gardening and other leisure-type activities. In so doing, this data-collection model can include, where applicable, 24/7 monitoring of not just physical activity, but also other health-related measurements that are currently being recorded by consumer-facing digital health technologies that include, but are not limited to, blood pressure, heart rate, and sleep.

The findings suggest that the users of mobile fitness apps that share their physical activity over Twitter show lower percentages of people achieving physical-activity recommendations than was reflected in the Healthy People data. While employing Twitter and other social media as data-collection tools could help researchers obtain information that users might not remember or be willing to disclose face-to-face or over the telephone, further research is needed to determine the cause of the lower percentages found in this study. Though some challenges remain, it is clear that this approach holds promise for yielding valuable information for improving behavior-change outcomes.

Another potential application of this research is in the area of sentiment analysis, which evaluates real-time shared thoughts on users’ experiences with the mobile fitness apps and associated physical activity. It was discovered that fitness-related tweets reflected a positive sentiment toward physical activity: for every one negative sentiment, there were four positive sentiments. This is particularly important for health and fitness professionals, who can gain important insight into the mindset of their clients before, during and after a workout or bout of physical activity. Imagine a personal trainer working with a client who completes a 6-mile strenuous hike and then tweets about how exhausted she is, saying that she will need to take a few days off to recover. The trainer can use this information to modify the client’s regimen to create a more positive experience and enhance motivation and adherence to the program.

On a larger scale, by moving beyond the numbers and evaluating the emotions tied to certain types of physical activity as performed by individual app users, researchers may be able to determine relationships between the enjoyment of physical activity and the likelihood of continued use of a wearable device and its associated fitness app.

Of particular interest, this research found that:

- The use of mobile fitness apps continued to gain popularity worldwide during the data-collection time period, with most users writing in English, followed by Indonesian, Japanese and Spanish. The small representation of Chinese fitness tweets is most likely due to Twitter being blocked by the government.
Discussion

- Mobile fitness app users who share fitness tweets are loyal to the mobile fitness app and rarely post to multiple mobile fitness apps.

- One in five users of mobile fitness apps and fitness tweeting only fitness tweet once.

- One in five users of mobile fitness apps and fitness tweeting do so more than 21 times, with an average of 16 tweets.

- Mobile fitness apps can make an impact on the perceived health of their users by providing fitness challenges for family, friends, and co-workers.

At the time of this research, more than 14,000 mobile health apps were available on the iTunes store alone. The developers in this space struggle for market share, user engagement, and financial reward. Mobile fitness app developers that understand the importance of optimizing user engagement by adapting products and services based on users’ responsiveness, sound behavior-change models, and perceived needs will likely flourish.

6.3 Summary of Hypotheses
The following section evaluates the original hypothesis from Chapter 1. The proposed hypothesis were as follows:

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: The higher the influence via Klout score, the lower the number of fitness tweets</td>
<td>Accepted</td>
</tr>
<tr>
<td>H2: Women will report more physical activity than men using sharing via Twitter from mobile fitness apps</td>
<td>Rejected</td>
</tr>
<tr>
<td>H3: There will be more fitness tweets from the Workout+ group that are positive versus negative.</td>
<td>Accepted</td>
</tr>
<tr>
<td>H4: Fitness tweet physical activity will be higher than self-reported measures from Healthy People 2020</td>
<td>Rejected</td>
</tr>
</tbody>
</table>

Table 6-1 - Summary of Hypotheses

**H1: The higher the influence via Klout score, the lower the number of fitness tweets**

Outcome – Accepted

It is the hypothesis of this researcher that those with a higher Klout score will have a lower number of fitness tweets. To maintain a high Klout score, a person must spend a larger amount of time on various online social media, and would, therefore, have fewer perceived opportunities to fitness tweet. As the Klout score passes 40, it becomes harder to maintain and increase one’s score (Martin, 2014). That said, a savvy user of fitness tweets might take the opportunity to maximize his or her Klout score by sharing workouts online.
This research found that overall, the higher the Klout score, the fewer fitness tweets.

![Number of Activity Tweets per Klout Quartile](image)

**Figure 6-1 - Number of Activity Tweets per Klout Quartile**

Of the total 583,252 Activity tweets, 30.8% were from the lowest Klout quartile, followed by a gradual decrease over the next three quartiles (26.5%, 21.4%, and 21.2%). A similar decrease in Klout by gender was observed, with a slight increase from the third to fourth quartile for women.
For each gender, percentages from low Klout to high Klout quartiles for Activity tweets were as follows:

Men – Q1 30.3%, Q2 26.7%, Q3 21.8% and Q4 21.2%

Women – Q1 31.5%, Q2 26.2%, Q3 20.9% and Q4 21.4%

However, when analyzed by the sub-classifications of Workout and Workout+, the same decrease was noted.
Figure 6-3 - Number of Workout Tweets per Klout Quartile by Gender

For each gender, percentages from low Klout to high Klout quartiles for Activity tweets were as follows:

Men – Q1 33.8%, Q2 28.0%, Q3 21.5% and Q4 16.7%

Women – Q1 34.7%, Q2 28.2%, Q3 20.3% and Q4 16.8%
For each gender, percentages from low Klout to high Klout quartiles for Activity tweets were as follows:

**Men**
- Q1: 21.6%
- Q2: 23.4%
- Q3: 22.7%
- Q4: 32.3%

**Women**
- Q1: 23.17%
- Q2: 21.2%
- Q3: 22.5%
- Q4: 33.3%

The finding of most interest in this hypothesis is from the Workout+ gender analysis, where it was calculated that women with higher Klout scores tweet more about their workouts from quartile 4 to then quartile 1 to then quartile 3 then to quartile 2. A similar but greater increase in men occurs from quartile 3 to 4.

Previous research by Shapp (2014) suggests a difference between the usage of hashtags by men and women, in that women, are more expressive than men are in their tweets and in their use of hashtags. In addition, the hashtag has become a way to show emphasis, with little to do with its original purpose to join a public conversation on a given topic (Julien, 2015), in this case, the sharing of their workout using a mobile fitness app. Users of Twitter have found new ways of using hashtags for interpersonal communication that also may build community (Shapp, 2014). Additional research into multiple uses of hashtags within each fitness tweet will provide insights into this phenomenon within fitness tweeting of physical activity within communities. As this research viewed only the hashtags created by the
mobile fitness app, the actual use of the hashtag was predetermined. However, a closer look at the actual words and phrases used by the breakdown of men and women from the fitness tweets database, specifically in the Workout+ classification, may give additional insights into the differences between men and women and fitness tweeting of their physical activity. This is an area for future research.

Additional future research can focus on overall online social influence as defined as a measure of how people, directly or indirectly, affect the thoughts, feelings, and actions of others (Smailovic, Striga, Mamic & Podobnik, 2014). A better understanding of the online influence of those who are active on Twitter and sharing their fitness tweets may lead to new and innovative ways to influence their followers to be more physically active through peer-to-peer influence, similar to programs created for marketing agencies to influence consumer behavior. Similar to the research by Cavazos-Rehg et al. (2015), physical-activity researchers can monitor and attempt to influence physical activity Twitter chatter sent by influential Twitter users who tend to encourage physical activity and popular among various demographic groups and age ranges. The findings can be used to inform online and offline efforts that work to target individuals who are most at risk for harms associated with a lack of physical activity.

H2: Women will report more physical activity than men using sharing via Twitter from mobile fitness apps

Outcome – Rejected

While equal proportions of men and women use Twitter in the United States, women appear to trail male users in terms of online influence (Julien, 2015). Research is divided regarding which gender is more physically active, with some researchers suggesting that women are less physically active than men (in this case, in Brazil) (Azevedo et al., 2007) and others suggesting there is no difference between the physical activity of genders (Bauman et al., 2011). However, it is the hypothesis of this researcher that the desire of women to be more expressive and share their physical activities online will create the opportunity for them to share more on Twitter than men. Therefore, they will report more physical activity.

This research found that within the subset of the Fitness Tweet Data Set that contained tweets with minutes reported, there were a total of 670,380 Activity tweets. Of those Activity tweets, 57.88% were from men, and 42.12% were from women. The percentages for the subcategory of Workout were 57.75% men and 42.12% women, and for Workout+ they were 58.23% men and 41.77% women. For all three categories, men had more fitness tweets than women.

With regards to minutes of physical activity tweeted, men had a higher percentage of minutes (56.66% to 43.34%), Workout tweets (58.65% to 41.33%) and Workout+ tweets (51.97% to 48.03%). Stated simply, in terms of both the number of Activity tweets (Activity total, Workout total, and Workout+
total) and minutes of physical activity (Activity total, Workout total, and Workout+ total), men outpaced
women.

This research can also be used to do further research into the physical-activity habits of those who use
Twitter to share self-reported physical activity, not only through mobile fitness apps but also through
the use of consumer-facing health technology wearables (like Fitbit). Through the procedures described
in this dissertation, gender can be inferred up to a possible 90% accuracy, thus allowing the creation of
a new fitness tweet database that can help support this type of research. Overall demographic
characteristics can also be implied through data analysis through services like DemographicPro and
include additional data points including, but not limited to, age, income level, and ethnic background.

**H3: There will be more fitness tweets from the Workout+ group that are positive versus negative.**

**Outcome – Accepted**

The hypothesis that there will be more fitness tweets from the Workout+ group that are positive versus
negative is based on the assumption that people tend to feel better after a workout, thus at the time of
their sending the Workout+ tweet, they will be in a positive state of mind. From the Fitness Tweet Data
Set and including the online sentiment analysis service (Aylien), the research was able to determine the
sentiment of the Workout+ tweets from four mobile fitness apps (Nike+, Endomondo, Runkeeper and
DailyMile). As these tweets were likely completed immediately after the user finished his or her
workout, this method of collecting the user’s perceptions and thoughts about the workout provides
unique, real-time insight. Previous research by Paul and Dredze (2011), Cavazos-Rehg et al. (2015)
and Pagoto, Schneider, Evans et al. (2014) used Twitter to determine different characteristics of
participants who used Twitter to share health-related information, including sentiment.

This research hypothesis suggests that overall, most of the Workout+ tweets were classified by the
algorithm as neutral (50%), with 40% classified as positive and 10% negative. However, a deeper
analysis of the sentiment from these four mobile fitness apps shows that these ratios change depending
on what mobile fitness app is being used to share the fitness tweet. Most surprising is that for DailyMile,
76% of the fitness tweets were classified as positive. The other three apps still had a majority of positive
to negative tweets; Endomondo (26% positive to 6% negative), Runkeeper (18% positive to 6%
negative) and Nike+ (10% positive to 9% negative).

When removing the neutral classification, the overall positive sentiment score increased, with the
overall score for all apps at 80% positive. Of interest is that Nike+ users tweeted 54.4% positive, which
is considerably less than the other three apps. The researcher theorizes that this is due to a perceived
high number of more elite athletes who use Nike+ for the tracking of their competitive workouts. These
users may share negative feelings after a more difficult run or bike might have a higher rate of injury and could be sharing results of competitions where they did not perform as expected.

This analysis would suggest that the sentiment score of Workout+ tweets varies depending on the choice of mobile fitness app used. Future work into this hypothesis would include a more in-depth breakdown of other demographic information such as age, gender, location and income level, as well as analysis of consumer-facing health-tracking wearables such as Fitbit or Garmin.

**H4: Fitness tweet physical activity will be higher than self-reported measures from Healthy People 2020**

**Outcome – Rejected**

The fourth hypothesis in this research was that the collected information about those who shared their workouts online in tweets that also include minutes of activity would be higher than data collected for the Healthy People 2020 report. The assumption is that since those that use mobile fitness apps are more interested in their physical activity, as evidenced in their use of the mobile fitness app, the resulting data will show higher minutes of physical activity than those surveyed for Healthy People 2020. This hypothesis was rejected as highlighted in Figure 6-6.
Across all mobile fitness apps and the physical activity shared from each, this research analysis would suggest that the reported physical activity is less than what was collected for the Healthy People 2020 report. In the United States, population data regarding physical activity is often collected and measured as part of an overall national health survey through self-reports (Troiano, Macera & Ballard-Barbash, 2001). Objective measurement devices that measure number of steps (pedometers) and movement intensity (accelerometers) offer a potential solution to problems with self-reported data (Troiano, 2006). While this hypothesis was rejected, it does not indicate a failure in the data-collection method. The advancement of these types of technologies to measure physical activity also presents new methodical challenges with regards to the type and amount of data that can be collected (Hurvitz et al., 2014). In addition, as suggested in previous research, recall bias could account for the higher than actual measurement of physical activity seen in the Healthy People 2020 report (Prince et al., 2008).

6.4 Challenges

One challenge within the fitness technology space is the lack of sharing of one’s health data. Similar to many database structures, health data silos are prevalent, with the database containing one’s physical activity typically not linked with the database containing one’s health records or daily food journaling. To achieve the collective goal of better health through technology, companies and developers must tear down these walls of data isolation and allow users to view their own linked personal health, fitness, and
Discussion

medical data. By so doing, a true digital picture of health can be established. This wellness digital picture will allow for a more effective and personalized journey toward health.

Another challenge in this emerging fitness technology space is the lack of oversight and regulation needed to protect the general public from fraudulent health claims. Like the personal-training industry in years past, any person or company can claim to be an expert in exercise and health. The Federal Drug Administration (FDA) in 2012 published draft guidance for mobile medical apps, with thousands of apps subject to FDA evaluation (Dolan, 2012). One concern with non-qualified developers who lack the basic foundation of exercise physiology and the understanding of exercise psychology is the potential for harm and even death.

Using the data-collection and data-processing tools described in this paper, this research created a growing data set of information that people publicly share from their smartphones and other devices, via Twitter, about their workout activities. This information includes data collected by the app itself—such as exercise type, duration, day of the week, mood, geographical location and time—as well as data on how people use fitness apps to share information and engage in social networking regarding their fitness activities. When looked at collectively, this information can facilitate research on how technology can be used to monitor and motivate participation in physical activity and how online social networks may play a role in physical activity promotion and adherence.

In addition, this research has provided preliminary data on how people are engaging with their online social communities to share information on their fitness activities. While there is a substantial amount of information being shared via Twitter regarding actual workouts (i.e., Activity tweets), there appears to be a lesser amount of conversation between the users of these mobile fitness apps (i.e., Conversation tweets). Unfortunately, this is where the established health and wellness benefits of social networking may take place. This is an area of future research using the data-collection model developed for this research. It is also unclear why people would decide to share their workouts on a social networking service such as Twitter. A future research hypothesis might be that there are possible relationships among mobile fitness users, online influence and the motivation to start or continue physical activity. In addition, it is important to understand why a person shares workout information on Twitter and what benefits are gained by doing so. Perhaps if users were encouraged to share their workouts via incentives or were presented with evidence that doing so would improve their chances for long-term success, an increase in sharing may take place. The author intends to explore these hypotheses in future work.

This dissertation highlights the importance of interdisciplinary research between exercise science and technology and the continued support and funding between these two areas of research. While each may have their own processes, procedures, and methods, they can share a common goal in the global fight to increase health and wellness through an increase in physical activity.
6.5 Future Research

Using the Fitness Tweet Classification Model, future research can be conducted with a better understanding of the associations and relationships among the use of mobile fitness apps, the sharing of this information with a social network and the possible long-term implications with regards to technology and physical activity.

The advancements of wearable technologies that collect physical-activity data throughout the day (such as Fitbit, Garmin or the Microsoft Band) will allow researcher to expand the data collection from not only the actual exercise as tracked using mobile fitness apps but also daily physical activity, thus giving a more accurate picture of one’s physical activity. The same data-collection and data-processing models can be used for data collection from wearables with small changes in coding of the Fitness Tweet Classifier.

The lack of meaningful and engaged conversation between mobile fitness app users is an important issue to be addressed in future research. If it is the intent of mobile fitness app developers to have their users increase physical-activity behavior by strengthening social support via one’s social network, then having a true understanding of why conversation is not occurring is critical. These findings would indicate similar usage patterns for general Twitter usage, a one-way, one-to-many publishing service as is seen in a two-way, peer-to-peer communication network (Heil & Piskorski, 2009).

This paper did not address the relationship between the social network (Twitter) and possible social influences on the type of fitness tweet posted by users. However, the data-collection tools and analysis models may allow for future research to address these types of relationships. If research has suggested a positive relationship between social support and increased physical activity and that the social support is related to social contacts (i.e., Twitter), then there could be a possible association between mobile fitness apps that track a person’s physical activity when shared with a social network and one’s physical activity, thus making Twitter a behavioural change tool. Future research could include a study that highlights fitness tweeting as an effective data-collection tool, enabling participants to fully grasp what the app measures and understand the benefits of sharing their physical-activity sessions.

This research was based on publicly available Twitter data specific to physical activity and the use of five different mobile fitness apps over a 184-day period. Thus, a more robust evaluation of fitness-related tweets over a longer time period and expanding to not only additional mobile fitness apps, but also to those tweets that represent physical activity, would be more comprehensive. This could be done with the creation of a list of words synonymous with physical activity and by using a tool like the Fitness Tweet Crawler to collect the data.

The evaluation of additional languages other than English would also allow for cross-cultural analysis of the sentiments of shared physical activity around the world. In addition, other social networks where
this data is being shared, like Facebook and the social networks of the mobile fitness apps themselves, could provide insights into the sharing of this type of physical activity information.
7 Conclusion

The results of this research include methodological and practical contributions. First, this research extends the body of literature related to consumer-facing health technologies through a smartphone, the use of online social networks and the sharing of self-reported health, wellness and fitness information through the network. Secondly, from a methodological approach, this dissertation proposes model criteria for inclusion of mobile fitness apps for research, a novel methodology in the collection of health, wellness and fitness information through an online social network based on previously established research principles, a system to assign gender for Twitter users, a more in-depth analysis of collected data, a procedure to assign and include additional demographic characteristics of Twitter users and a data-classification tool that can be modified by future researchers in their classification efforts. Finally, this research presents three practical contributions in the analysis and presentation of information from the Fitness Tweet Data Set including online influence, sentiment analysis and a comparison model to self-reported physical activity levels and Healthy People 2020.

7.1 Methodological Contributions

This dissertation explores and examines the use of technology by using mobile fitness applications, the sharing of fitness tweets within one’s social network and the outcomes of such sharing activity. The main contributions of this research are as follows:

**The Fitness Tweet Data Collection Tool** – Given the tremendous amount of data generated by Twitter, researchers need tools to manage and analyze these data in order to address research questions regarding the use of technology and social networks to promote health behavior change. The Fitness Tweet Data Collection Tool was created to assist in the collection of data (tweets) for this research. In addition to the collection of each tweet, other information such as Twitter user demographics and social influence via a Klout score was collected. As a result, additional researchers have adopted and modified the basis of the data-collection tool for their own research, outside of the fitness and health research areas.

**The Fitness Tweet Classification Model** – The Fitness Tweet Classification Model was created to allow the researcher to understand better and to classify each collected tweet. By so doing, researchers can have a better understanding of how Twitter users are engaging in conversation about their own health and fitness. This classification model was based on existing data-classification models specific to social network and Twitter research but is the first Twitter classification model for mobile fitness app tweets.

**Research data-collection using Twitter** – As highlighted in this dissertation, a new and innovative data-collection tool for physical activity is possible using Twitter. As suggested by the studies presented in this research, self-reported physical-activity measures can often underestimate the true levels of
physical activity. By constructing a physical-activity reporting system using Twitter and mobile fitness apps and/or sensors, more precise and accurate results can be obtained. An excellent example of this is the wide discrepancy between the Healthy People 2020 physical-activity levels and the measured physical-activity levels as reported via mobile fitness apps and Twitter.

7.2 Practical Contributions

This research provides three practical contributions related to experimental approaches from the collected data in the Fitness Tweet Data Set, demonstrating the possibility and utilization of such information in future research as follows:

The Reporting of Physical Activity on Twitter from Mobile Fitness Apps and Klout: The goal of this experimental approach is to use the Fitness Tweet Data Set and determine if there is a correlation between the type of Activity tweet (either Workout or Workout+), gender and one’s online influence as measured by Klout score.

Sentiment Analysis of Activity Fitness Tweets: The goal of this experimental approach is to examine the sentiment of publicly available fitness tweets from four specific mobile fitness apps by looking at users’ real-time shared thoughts via Twitter regarding their experience with the mobile fitness app and associated physical activity.

The Use of Twitter to Collect Real-time Physical Activity—Comparing Data with Healthy People 2020: The goal of this research was to compare the self-reported estimates of physical-activity data provided to the Healthy People 2020 research team via a telephone survey to the mobile fitness app real-time reporting of physical activity using Twitter.

7.3 Limitations

As with many research topics that include technology, the advancement of the technologies often outpaces the advancement of research. This research is no different. At the time this research began, mobile fitness apps were considered new and innovative tools for tracking many types of physical activity. In the years since the start of this research, a change in the landscape of the fitness-tracker industry has occurred, from the addition of thousands of different types of mobile health and fitness apps to the introduction of a new line of consumer-facing fitness wearable technology (Fitbit, Microsoft Band, and Jawbone). While the mobile fitness apps used in this research are still functional and operational, consolidation in the marketplace has occurred with two of the apps, MyFitnessPal and Endomondo, being acquired by Under Armour. As with many consolidations within an industry, positive consequences can be expected, including, but limited to, an increase in financial capital, greater efficiencies, new approaches to innovation, increased awareness of products and services and a greater understanding of privacy and security concerns (Berger, Demetz & Strahan, 1999).
 Conclusion

A second limitation of this research was the use of just one online social network. It was established in this research that the sharing of fitness information within social networks from a mobile fitness app happens mostly in three different social networks, the one created by the mobile fitness app company, on Facebook, and on Twitter. Since Twitter was the only online social network that was public and thus able to be data mined, it was the social network of choice. Now that the concept of data mining of physical activity through an online social network has been established, it is the hope of this researcher that future research into the private data collections on Facebook and the mobile fitness apps can take place.

A third limitation involves limiting this research to the guidelines of the National University of Ireland at Galway and the School of Engineering. While this was designed as a multidisciplinary research topic between computer science and health promotion, the primary research focused on the creation of tools for the data collection, analysis and classification and not the implications around behavior change. That said, with these available tools, behavioural researchers can now examine a wide range of questions, such as how the use of mobile fitness apps and the sharing of workout information using Twitter is related to possible exercise motivation within one’s social network, how mobile fitness apps are related to the possible influence of social support by using Twitter and which mobile fitness app sharing features are most appropriate with regards to using technology to impact physical inactivity. Addressing these issues will lay the foundation for understanding and potentially improving the role that technology and social networking can play in improving health and fitness behavior. While personal tracking has proven an effective tool in the fitness arena, the tracking itself may not facilitate long-term adherence to continued physical activity. West et al. (2012) encourage researchers to develop more applications that target public health behaviors, such as healthy eating and exercise. Although there are many apps on the market that deal with personal health, diet and physical activity, the need persists for tools that include reinforcing factors such as encouragement, evaluation, and user interaction.

7.4 Conclusion

This research presents a novel perspective into the shared fitness tweets of the users of mobile fitness apps. The research has created a data-collection model for fitness tweets (Fitness Tweet Crawler) as well as a data-classification model (Fitness Tweet Classification Model). These tools were created to allow other researchers to modify these tools for their own use in other fields of health promotion, physical activity, wellness or other Twitter-related data-gathering research. Three experimental approaches using the Fitness Tweet Data Set show novel ways to analyze and present self-reported physical-activity data using an online social network. Wearable devices and their associated smartphone apps continue to grow in consumer popularity. It should be the goal of everyone involved, from the device and app developers to the personal trainers
and their clients, to maximize the effectiveness of using these exciting tools. This dissertation has provided tools for advancing research on mobile fitness app use, social networking and physical activity.
8 REFERENCES


148


References


References


Golder, S., Ahmed, S., Norman, G., & Booth, A. (2017, June 6). Attitudes toward the ethics of research using social media: A systematic review. *Journal of Medical Internet Research. JMI... Publica...ns Inc. https://doi.org/10.2196/jmir.7082*


References


References

https://doi.org/10.1016/S1048-9843(97)90028-X


References


References


References


References


158
References


References


9 Appendix

9.1 Appendix A – Directions Used for Human Model Fitness Tweet Classification

Instructions:

- The Excel sheet again has 500 tweets (100 random tweets from the 5 different hashtags)
- All of the tweets have been placed into one sheet and sorted for you
- In the CLASSIFICATION column, please pick the ONE classification from the list below that you think best fits the tweet. Use the number listed before the classification type. **Put only the number into the spreadsheet:**
  1. Activity – Workout
  2. Activity – Workout+
  3. Blarney – Pointless Babble
  4. Blarney – SPAM
  5. Conversation – Technical Support
  6. Conversation – Corporate Marketing
  7. Conversation – Statement of Support
  8. Conversation – Information Sharing
- Only classify tweets that are in English. Non-English tweets should be classified as #3.

Classification descriptions:

1. **Activity – Workout**
   - Use this classification if the tweet contains ONLY information about the workout (e.g., distance and time).
2. **Activity – Workout+**
   - Use this classification if the tweet contains information about the workout AND contains any additional information about the workout (e.g., the person adds his or her own commentary about the workout).
3. **Blarney – Pointless Babble**
   - Use this classification if the tweet is relevant to the hashtag, but seems to be pointless babble (e.g., our first set of tweets included a tweet about looking at a chicken). If the tweet is NOT in English, mark it as Blarney – Pointless Babble.
4. **Blarney – SPAM**
   - Use this classification if the tweet is SPAM.
5. **Conversation – Technical Support**
   - Use this classification if the tweet is of some sort of technical support, either by the company or by other users of the hashtag.
6. **Conversation – Corporate Marketing**
   - Use this classification if the tweet is any sort of marketing or promotion.
7. **Conversation – Statement of Support**
   - Use this classification if the tweet is any sort of support from either the company or other members of the community.
8. **Conversation – Information Sharing**
   - Use this classification if the tweet involves the sharing of any information (that is not already covered above). For example, if the tweet is a marketing message, classification #6 of Corporate Marketing would take priority.
#tweet
What do you do when you aren't burning Dailymile donuts? Full time public accounting major. Mr... http://dailymile.com/0QBDL
@andygaunt watching your live broadcast on runkeeper. nice going buddy :)
@bamacmac That is why I started using myfitnesspal.com so that I could monitor my food intake & weight loss.
@bnyalaanya well I didn’t have a bike to cycle so... ;). But other than the exercise lessons I think it might be better than RunKeeper
@britishbullblog Of course, that email still doesn't take away the fact that we were all "early adopters" of @Dailymile. :) @andyo22
@chapree This is a more complex question than you think! The answer depends, among other things, on whether your quickstart run
@Endomondo seems like the "New Workout from Route" has gone bonkers!! http://bit.ly/eP8g9i
@Endomondo Trance #music #endomondo
@EventJoe stay away from the butter #myfitnesspal
@JoseMataGomez no hay que intentarlo, hay que hacerlo, instalaren endomondo
@kejeduk rutenya ruwet brader, cari yg jalanya turun http://bit.ly/h27Hzh
@la.stik da Ñятандартно по каналу http://www.endomondo.com/workouts/ql_s9wNKRuA
@miraxtk #nikeplus ½ì, ¼° ¼ ^ ^
@pauloamagrela tá o q vc queria. RT @BetoRochaCosta: Was out mountain biking 22.37 km with #Endomondo. See it here: http://bit.ly/g2H69d
@StuMclaren @RunKeeper Oh, just walking... ;)... I see! Yeah, pretty good I guess.... ;)
@sunbug Thanks :) After putting on 2lb last week I was nervous. I've been using myfitnesspal.com app, it works very well.
@taichioutwindow I lost a s-tone using myfitnesspal app. I out on some just married podge. Cycle everywhere now.
@tarun_ The SportWatch GPS can save one interval scheme at a time, but it's very easy to edit the interval scheme between workouts.
@testpattern oh oops. Ipod not iPhone. Nevermind about runkeeper. #brainfail
@theashbed burned 389 calories doing 24 minutes of "Jillian Michaels 30 Day Shred (Level 1)" #myfitnesspal http://amzn.to/fQw=Gw8Tvz
@TheBeerRunner @DailyMile Very funny! But true. Gonna run a 1/2 marathon in June with a beer tied to a stick on my hat to lead me on!
@ZeroK66 I wish you would tweet more than just your myfitnesspal updates :-( #boring
@ZeroK66 I wish you would tweet more than just your myfitnesspal updates :-( #boring
128kilocaloríes consumed. #Endomondo
A little ski racing =) Layering under lycra is never flattering haha http://dailymile.com/e/Oo8s1
Achei de terminar uma corrida de 10.0 km em 1:21:04 com o Nike+ GPS. #nikeplus
Achei de terminar uma corrida de 16.6 km em 2:06:30 com o Nike+ GPS. #nikeplus
Achei de terminar uma corrida de 1.05 km em 01:16:34 com o Nike+ GPS. #nikeplus
Achei de terminar uma corrida de 1.34 km em 09:16 com o Nike+ GPS. #nikeplus
Achei de terminar uma corrida de 1.92 mi com um tempo de 32:45 com Nike+ GPS. #nikeplus
Achei de terminar uma corrida de 12.3 km com um tempo de 1:14:37 com Nike+ GPS. #nikeplus
Achei de terminar uma corrida de 4.35 km com um tempo de 30:45 com Nike+ GPS. #nikeplus
Achei de terminar uma corrida de 4.55 km com um tempo de 20:00 com Nike+ GPS. #nikeplus
Achei de terminar uma corrida de 6.03 km com um tempo de 38:10 com Nike+ GPS. #nikeplus
Achieved a new personal record with @RunKeeper: Farthest distance http://bit.ly/gleMgY #FitnessAlerts
Achieved a new personal record with @RunKeeper: Farthest distance http://bit.ly/gsAYpg #FitnessAlerts
Achieved a new personal record with @RunKeeper: Farthest distance in a week... http://bit.ly/hKFBcs #FitnessAlerts
Achieved a new personal record with @RunKeeper: Fastest average pace... http://bit.ly/gY70OI #FitnessAlerts
Ah, that felt great. Just posted a 19.52 km run with @runkeeper. Check it out! http://rnkpr.com/ahmcig #RunKeeper
Anyone out there have the myfitnesspal app? If so, let's lose weight together! ;}
Appendix

Arvo swim http://dailymile.com/e/PbUJ

As you probably already know, I am raising money to #beatcancer. Please check out my latest blog post! You’… http://dailymile.com/e/S4SK

Blog post for my dailymile friend @reallynotarunnr http://the-walking-nun.blogspot.com/2011/03/childrens-hos… http://dailymile.com/e/OjZL


burned 1026 calories doing 106 minutes of cardio exercises, including "Circuit training, general" #myfitnesspal
burned 1027 calories doing 92 minutes of "Elliptical Trainer" #myfitnesspal
burned 103 calories doing 30 minutes of "Pilates" #myfitnesspal
burned 1050 calories doing 68 minutes of cardio exercises, including "Elliptical Trainer" #myfitnesspal
burned 128 calories doing 30 minutes of "Walking, 2.5 mph, leisurely pace" #myfitnesspal
burned 139 calories doing 15 minutes of "Walking, 4.0 mph. very brisk pace" #myfitnesspal
burned 148 calories doing 10 minutes of "Stationary bike, moderate effort (bicycling, cycling, biking)" #myfitnesspal
burned 166 calories doing 13 minutes of "Basketball, game" #myfitnesspal
burned 196 calories doing 20 minutes of "Stationary bike, light effort (bicycling, cycling, biking)" #myfitnesspal
burned 204 calories doing 35 minutes of "Walking, 3.0 mph. mod. pace, walking dog" #myfitnesspal
burned 222 calories doing 45 minutes of "Walking, 3.0 mph. mod. pace, walking dog" #myfitnesspal
burned 26 calories doing 5 minutes of "Mini Trampoline" #myfitnesspal
burned 277 calories doing 90 minutes of "Housework" #myfitnesspal
burned 310 calories doing 90 minutes of "Walking, 2.0 mph, slow pace" #myfitnesspal
burned 350 calories doing 30 minutes of "Fire 30" #myfitnesspal
burned 369 calories doing 51 minutes of "Dancing, general" #myfitnesspal
burned 379 calories doing 36 minutes of "Stationary bike, moderate effort (bicycling, cycling, biking)" #myfitnesspal
burned 379 calories doing 45 minutes of "Aerobics, step, with 6-8 inch step" #myfitnesspal
burned 416 calories doing 120 minutes of "Cleaning, light, moderate effort" #myfitnesspal
burned 470 calories doing 90 minutes of cardio exercises, including "Cleaning, light, moderate effort" #myfitnesspal
burned 48 calories doing 10 minutes of "Walking, 3.0 mph. mod. pace, walking dog" #myfitnesspal
burned 489 calories doing 45 minutes of "Elliptical Trainer" #myfitnesspal
burned 503 calories doing 128 minutes of "Yoga" #myfitnesspal
burned 505 calories doing 30 minutes of "Stair-treadmill ergometer, general" #myfitnesspal
burned 506 calories doing 40 minutes of "Rd 2 Day 5 Insanity" #myfitnesspal
burned 528 calories doing 45 minutes of cardio exercises, including "Stair-treadmill ergometer, general" #myfitnesspal
burned 532 calories doing 40 minutes of cardio exercises, including "Fire30" #myfitnesspal
burned 590 calories doing 60 minutes of "BODYPUMP" #myfitnesspal
burned 609 calories doing 60 minutes of "Running (jogging), 5 mph (12 min mile)?" #myfitnesspal
burned 63 calories doing 6 minutes of "Wii Gold's Gym Cardio Workout - Shape Boxing - Beginner - Warm Up - Beginner Combo" #myfitnesspal
burned 708 calories doing 140 minutes of "Stationary bike, moderate effort (bicycling, cycling, biking)" #myfitnesspal
burned 71 calories doing 8 minutes of "Wii baseball" #myfitnesspal
burned 718 calories doing 60 minutes of cardio exercises, including "Stair-treadmill ergometer, general" #myfitnesspal
burned 75 calories doing 27 minutes of "Peddler" #myfitnesspal
burned 771 calories doing 58 minutes of "#P90X Kenpo X" #myfitnesspal
burned 776 calories doing 120 minutes of "Walking, 2.0 mph, slow pace" #myfitnesspal
burned 783 calories doing 45 minutes of "Calisthenics (pushups, sit-ups), vigorous effort" #myfitnesspal
burned 783 calories doing 50 minutes of "Turbo kick" #myfitnesspal
burned 98 calories doing 45 minutes of "Pilates" #myfitnesspal

Calories in Dunkin' Donuts Hazelnut Coffee - Calories and... Calories in Dunkin' Donuts Hazelnut Coffee. Find... http://bit.ly/fz2U8D

completed her food and exercise diary for 03/10/2011 and was under her calorie goal #myfitnesspal
completed her food and exercise diary for 03/10/2011 and was under her calorie goal #myfitnesspal
completed her food and exercise diary for 03/11/2011 and was under her calorie goal http://bit.ly/ARIr#myfitnesspal
completed her food and exercise diary for 03/17/2011 #myfitnesspal
completed her food and exercise diary for 03/26/2011 and was under her calorie goal http://bit.ly/ekN6oe #myfitnesspal
completed her food and exercise diary for 03/28/2011 and was under her calorie goal #myfitnesspal
completed her food and exercise diary for 03/29/2011 and was under her calorie goal #myfitnesspal
completed her food and exercise diary for 03/29/2011 and was under her calorie goal http://bit.ly/eq70b #myfitnesspal
completed her food and exercise diary for 03/30/2011 and was under her calorie goal #myfitnesspal
completed her food and exercise diary for 04/02/2011 and was under her calorie goal #myfitnesspal
completed her food and exercise diary for 04/04/2011 and was under her calorie goal #myfitnesspal
completed her food and exercise diary for 04/09/2011 and was under her calorie goal #myfitnesspal
completed her food and exercise diary for 04/09/2011 and was under her calorie goal http://bit.ly/gHbRFP #myfitnesspal
completed her food and exercise diary for 04/10/2011 and was under her calorie goal #myfitnesspal
completed her food and exercise diary for 04/11/2011 and was under her calorie goal #myfitnesspal
completed her food and exercise diary for 04/12/2011 and was under her calorie goal #myfitnesspal
completed her food and exercise diary for 04/21/2011 #myfitnesspal
completed her food and exercise diary for 04/23/2011 #myfitnesspal
completed his food and exercise diary for 03/10/2011 and was under his calorie goal #myfitnesspal
completed his food and exercise diary for 03/11/2011 #myfitnesspal
completed his food and exercise diary for 03/13/2011 #myfitnesspal
completed his food and exercise diary for 03/21/2011 #myfitnesspal
completed his food and exercise diary for 03/21/2011 and was under his calorie goal #myfitnesspal
completed his food and exercise diary for 03/22/2011 #myfitnesspal
completed his food and exercise diary for 03/26/2011 and was under his calorie goal #myfitnesspal
completed his food and exercise diary for 04/01/2011 http://bit.ly/0wOM #myfitnesspal
completed his food and exercise diary for 04/06/2011 and was under his calorie goal http://bit.ly/KTLv #myfitnesspal
completed his food and exercise diary for 04/08/2011 #myfitnesspal
completed his food and exercise diary for 04/09/2011 and was under his calorie goal #myfitnesspal
completed his food and exercise diary for 04/12/2011 and was under his calorie goal #myfitnesspal
completed his food and exercise diary for 04/23/2011 and was under his calorie goal #myfitnesspal
completed his food and exercise diary for 04/23/2011 and was under his calorie goal #myfitnesspal
completed his food and exercise diary for 04/23/2011 and was under his calorie goal #myfitnesspal
completed his food and exercise diary for 05/06/2011 and was under his calorie goal http://bit.ly/IKTLv #myfitnesspal
completed his food and exercise diary for 04/08/2011 #myfitnesspal
completed his food and exercise diary for 04/09/2011 and was under his calorie goal #myfitnesspal
completed his food and exercise diary for 04/12/2011 and was under his calorie goal #myfitnesspal
completed his food and exercise diary for 04/23/2011 and was under his calorie goal #myfitnesspal

Cool new app added to my iPad myfitnesspal night all

correu 15.61 km no dia 22/4/2011 at 9:13 AM a uma velocidade de 4'26"/km http://go.nike.com/2tf4qa2
corri 3.91 km el 22/4/2011 at 7:31 PM con un ritmo de 7'27"/km http://go.nike.com/6k2hlv7

Dear Christopher, Congratulations! You are now registered for the 2011 Dodge Rock 'n' Roll San Diego Marathon... http://dailymile.com/e/OcS0
Decided to have a "rest day" and only swim this afternoon. Planning on 90 miles on the bike followed by a 10... http://dailymile.com/e/QC1W
Did a fitness workout 2.2 miles in 25 mins and felt great. http://dailymile.com/e/Puys
Did a fitness workout 2.5 kilometers in 18 mins and felt good. I did a pyramid, which is run 1, walk 1, run 2... http://dailymile.com/e/Qihj
Did a fitness workout for 27 mins and felt great. Warm up, stretch, cool down http://dailymile.com/e/QiijD
Did a fitness workout for 35 mins and felt good. Onto a new cycle, this was mostly upper body today. http://dailymile.com/e/Plos
Did a fitness workout for 70 mins. For a Monday morning, not too bad! (Thanks to an early bedtime on Sun. ni... http://dailymile.com/e/PyVc
Did a fitness workout for 90 mins and felt alright. meh http://dailymile.com/e/OgW9
Did a spinning workout and felt great. ....so tough today. Renee's class was a series of workouts that reali... http://dailymile.com/e/QyWh
Did a weights workout and felt good. Don't get bored with me. I Alternate between the dame upper body lower b... http://dailymile.com/e/Qox9
Did a weights workout for 1 hour and felt great. I love this workout, makes me feel so strong!!!!!! http://dailymile.com/e/QPXM

Did a weights workout for 35 mins and felt good. Ran first so weights felt heavy at first... Then got lighter! http://dailymile.com/e/RwMQ

Did a yoga workout for 60 mins and felt good. I used quite light weights, at points I could have gone heavier. http://dailymile.com/e/RyNR

Did a yoga workout for 1 hour and 30 mins and felt great. Tried out a new Yoga Class w @ekshee, great way to ... http://dailymile.com/e/Pe9Z

Did a yoga workout for 45 mins and felt good. My first morning at Pregnancy PT - as being a participant and n... http://dailymile.com/e/PpyR

Endomondo Cycle Workout http://t.co/AkUQlEA

Fitbit Partners With RunKeeper, Microsoft, About.Me And Others With New API http://sns.mx/Kqc5y8


For any of you in the UK (I think its only in the uk... unsure) I got introduced to Spotify this afternoon af... http://dailymile.com/e/QiYZ

Going to Joggers and Lagers tonight so postponed morning run. There is a reason I workout in the morning... it'... http://dailymile.com/e/Ru78

Great app @runkeeper really fills a need I had. Thanks for the recommendation @TripleDutyMomma

Holy shitballz. http://dailymile.com/e/Qesp

How come I can go nearly 4 weeks pre-race with no alcohol, but can't go 24hrs without chocolate. Prohibition ... http://dailymile.com/e/QV6T

I am getting tees and possibly limited edition screen prints made to raise money for the relief effort in Jap... http://dailymile.com/e/Obzs

I installed the @MyFitnessPal app for my BB but it doesn't show my wall just food diary etc. :

I just finished a 0.01 mi run with a time of 00:28 with Nike+ GPS. #nikeplus

I just finished a 0.02 mi run with a time of 06:53 with Nike+ GPS. #nikeplus

I just finished a 1.02 km run with a time of 05:57 with Nike+ GPS. #nikeplus

I just finished a 1.01 mi run with a time of 11:22 with Nike+ GPS. #nikeplus

I just finished a 1.05 mi run with a time of 11:30 with Nike+ GPS. #nikeplus

I just finished a 1.23 mi run with a time of 13:05 with Nike+ GPS. #nikeplus

I just finished a 1.28 mi run with a time of 11:26 with Nike+ GPS. #nikeplus

I just finished a 1.59 km run with a time of 19:18 with Nike+ GPS. #nikeplus

I just finished a 1.70 mi run with a time of 13:52 with Nike+ GPS. #nikeplus

I just finished a 1.85 mi run with a time of 18:29 with Nike+ GPS. #nikeplus

I just finished a 10.0 mi run with a time of 1:31:44 with Nike+ GPS. #nikeplus

I just finished a 11.6 km run with a time of 1:00:07 with Nike+ GPS. #nikeplus

I just finished a 11.9 km run with a time of 09:04 with Nike+ GPS. #nikeplus

I just finished a 12.0 km run with a time of 59:20 with Nike+ GPS. #nikeplus

I just finished a 2.00 mi run with a time of 24:44 with Nike+ GPS. #nikeplus

I just finished a 2.26 km run with a time of 1:08:24 with Nike+ GPS. #nikeplus

I just finished a 2.44 mi run with a time of 35:38 with Nike+ GPS. #nikeplus

I just finished a 2.66 km run with a time of 22:51 with Nike+ GPS. #nikeplus

I just finished a 2.73 mi run with a time of 46:20 with Nike+ GPS. #nikeplus

I just finished a 2.78 mi run with a time of 35:34 with Nike+ GPS. #nikeplus

I just finished a 3.00 km run with a time of 17:02 with Nike+ GPS. #nikeplus

I just finished a 3.90 km run with a time of 29:35 with Nike+ GPS. #nikeplus

I just finished a 3.12 mi run with a time of 28:51 with Nike+ GPS. #nikeplus

I just finished a 3.46 mi run with a time of 51:46 with Nike+ GPS. #nikeplus

I just finished a 3.79 km run with a time of 40:39 with Nike+ GPS. #nikeplus

I just finished a 4.49 km run with a time of 43:28 with Nike+ GPS. #nikeplus

I just finished a 4.00 mi run with a time of 30:23 with Nike+ GPS. #nikeplus

I just finished a 4.00 mi run with a time of 32:11 with Nike+ GPS. #nikeplus

I just finished a 4.01 km run with a time of 27:04 with Nike+ GPS. #nikeplus

I just finished a 4.01 mi run with a time of 39:00 with Nike+ GPS. #nikeplus
I just finished a 4.05 mi run with a time of 51:13 with Nike+ GPS. #nikeplus

I just finished a 4.33 km run with a time of 42:32 with Nike+ GPS. #nikeplus

I just finished a 5.01 mi run with a time of 50:30 with Nike+ GPS. #nikeplus

I just finished a 5.03 mi run with a time of 47:20 with Nike+ GPS. #nikeplus

I just finished a 5.37 km run with a time of 40:08 with Nike+ GPS. #nikeplus

I just finished a 5.97 mi run with a time of 54:17 with Nike+ GPS. #nikeplus

I just finished a 6.31 km run with a time of 39:44 with Nike+ GPS. #nikeplus

I just finished a 6.0 mi run with a time of 50:30 with Nike+ GPS. #nikeplus

I just finished a 6.11 mi run with a time of 57:08 with Nike+ GPS. #nikeplus

I just finished a 6.13 km run with a time of 32:42 with Nike+ GPS. #nikeplus

I just finished a 6.20 mi run with a time of 1:14:08 with Nike+ GPS. #nikeplus

I just finished a 6.31 km run with a time of 1:08:23 with Nike+ GPS. #nikeplus

I just finished a 6.67 mi run with a time of 1:03:42 with Nike+ GPS. #nikeplus

I just finished a 7.11 km run with a time of 58:32 with Nike+ GPS. #nikeplus

I just finished a 7.21 km run with a time of 33:34 with Nike+ GPS. #nikeplus

I just finished a 8.02 km run with a time of 42:33 with Nike+ GPS. #nikeplus

I just finished a 9.01 mi run with a time of 1:24:46 with Nike+ GPS. #nikeplus

I might miss my routine in DC, but I am not throwing myself a pity party anymore. mythousandmileyear.com #dailymile

I think part of my consistency problem is that I just really don't like running. I've been hoping that the C2... http://dailymile.com/e/P7T3

Ich habe gerade einen Lauf über 1,85 km in einer Zeit von 12:11 mit Nike+ GPS beendet. #nikeplus

Ich habe gerade einen Lauf über 10,0 km in einer Zeit von 59:14 mit Nike+ GPS beendet. #nikeplus

Ich habe gerade einen Lauf über 4,11 km in einer Zeit von 22:49 mit Nike+ GPS beendet. #nikeplus

Ich habe gerade einen Lauf über 6.12 km in einer Zeit von 43:58 mit Nike+ GPS beendet. #nikeplus

J'ai terminé une course de 3.02 km en 21:35 avec Nike+ GPS. #nikeplus

Jancik opo iki rek #ENVY RT @cacinkholiday: Was out cycling 6.00 km with #Endomondo. See it here: http://bit.ly/hH80JK

Just began a cycling workout using #Endomondo. Follow me live: http://bit.ly/e2v9hL


Just began a cycling workout using #Endomondo. Follow me live: http://bit.ly/izr2hWg

Just began a cycling workout using #Endomondo. Follow me live: http://bit.ly/bbYy3wJ

Just began a riding workout using #Endomondo. Follow me live: http://bit.ly/IMq1Uj


Just began a running workout using #Endomondo. Follow me live: http://bit.ly/PvUYw

Just began a running workout using #Endomondo. Follow me live: http://bit.ly/5gcdAX


Just began a walking workout using #Endomondo. Follow me live: http://bit.ly/hhCIoN

Just completed a 0.68 mi walk with @runkeeper. Check it out! http://rnkpr.com/agquyz #RunKeeper

Just completed a 0.81 mi walk - with sis's pups. http://rnkpr.com/a/74sa #RunKeeper

Just completed a 0.96 mi walk with @runkeeper. Check it out! http://rnkpr.com/ahbyrv #RunKeeper

Just completed a 1.01 mi run with @runkeeper. Check it out! http://rnkpr.com/aht3cp #RunKeeper

Just completed a 1.06 mi run with @runkeeper. Check it out! http://rnkpr.com/airkcg #RunKeeper

Just completed a 1.25 km ski run with @runkeeper. Check it out! http://rnkpr.com/ah62np #RunKeeper

Just completed a 1.45 mi run - Taking steps backwards WTF. http://rnkpr.com/aighle #RunKeeper

Just completed a 1.74 km walk with @runkeeper. Check it out! http://rnkpr.com/agvevz #RunKeeper

Just completed a 1.88 mi walk with @runkeeper. Check it out! http://rnkpr.com/4w6s #RunKeeper
Appendix

Just completed a 1.90 mi run - 2nd run with the fivefingers and this time was a lot harder need to get used to them. http://rnkpr.com/ah89es #RunKeeper

Just completed a 1.91 km walk with @runkeeper. Check it out! http://rnkpr.com/ahI3y #RunKeeper

Just completed a 10.02 km run with @runkeeper. #NikePlus is so far behind! http://rnkpr.com/ai7e6u #RunKeeper

Just completed a 10.11 km walk with @runkeeper. Check it out! http://rnkpr.com/aiujm #RunKeeper

Just completed a 10.37 km run - One of my favorite runs with one of my favorite peoples... @gary_Hardw ... http://rnkpr.com/aiag0h #RunKeeper

Just completed a 11.18 km bike ride with @runkeeper. Check it out! http://rnkpr.com/ahe4jp #RunKeeper

Just completed a 11.35 km bike ride - Coaching. http://rnkpr.com/aihnjg #RunKeeper

Just completed a 14.33 km run with @runkeeper. Check it out! http://rnkpr.com/aihnju #RunKeeper

Just completed a 17.57 mi bike ride with @runkeeper. Check it out! http://rnkpr.com/ai332a #RunKeeper

Just completed a 2.00 mi run with @runkeeper. Check it out! http://rnkpr.com/ah43x #RunKeeper

Just completed a 2.04 mi run with @runkeeper. Check it out! http://rnkpr.com/ahkwku #RunKeeper

Just completed a 2.10 km run with @runkeeper. Check it out! http://rnkpr.com/aiikug #RunKeeper

Just completed a 2.72 km run - Muito bom! http://rnkpr.com/agwv5y #RunKeeper

Just completed a 2.81 km walk - Yeah that was even lekker! http://rnkpr.com/ah61r #RunKeeper

Just completed a 22.49 km bike ride with @runkeeper. Check it out! http://rnkpr.com/ai24 #RunKeeper

Just completed a 24.31 km bike ride - Gemiddeld 25 km/u verloren bij twee stoplichten. Muts verloren ... http://rnkpr.com/ai324 #RunKeeper

Just completed a 26.96 km bike ride - ...). http://rnkpr.com/ahswo #RunKeeper

Just completed a 29.69 km bike ride with @runkeeper. Check it out! http://rnkpr.com/ai23 #RunKeeper

Just completed a 3.06 mi run with @runkeeper. Check it out! http://rnkpr.com/ai52m #RunKeeper

Just completed a 3.07 mi run with @runkeeper. Check it out! http://rnkpr.com/ai5shn #RunKeeper

Just completed a 3.09 km bike ride - boy that was a short ride. http://rnkpr.com/ahqgr6 #RunKeeper

Just completed a 3.15 mi bike ride - Struggling to pick up the pace a bit. http://rnkpr.com/ai6aq #RunKeeper

Just completed a 3.31 km run - Que floja!! http://rnkpr.com/ahs6j4 #RunKeeper

Just completed a 3.31 mi run - That hurt. Wind sucks . http://rnkpr.com/ah6j1 #RunKeeper

Just completed a 3.32 mi walk - Dog walk. http://rnkpr.com/aijkian #RunKeeper

Just completed a 3.40 km run with @runkeeper. Check it out! http://rnkpr.com/ai25q #RunKeeper

Just completed a 3.53 mi walk - wwwp5k walk. http://rnkpr.com/ai5m8 #RunKeeper

Just completed a 3.61 mi run with @runkeeper. Check it out! http://rnkpr.com/ai5rhb #RunKeeper

Just completed a 3.74 mi run with @runkeeper. Check it out! http://rnkpr.com/aihn8k #RunKeeper

Just completed a 3.85 km run - Last night's vodka-redbull definitely did not give me wings. http://rnkpr.com/ai5ayv #RunKeeper

Just completed a 3.85 mi run with @runkeeper. Check it out! http://rnkpr.com/ai5df #RunKeeper

Just completed a 3.96 mi run - Slow and steady wins the race ya? http://rnkpr.com/aih5m #RunKeeper

Just completed a 3.97 km walk with @runkeeper. Check it out! http://rnkpr.com/ai5br #RunKeeper

Just completed a 35.55 km bike ride with @runkeeper. Check it out! http://rnkpr.com/aiel13 #RunKeeper

Just completed a 4.00 mi run - @lochary was awesome!!! Today was an easy run! http://rnkpr.com/aihn0h #RunKeeper

Just completed a 4.06 km bike with @runkeeper. Check it out! http://rnkpr.com/aih9ln #RunKeeper

Just completed a 4.11 mi run - Woo hoo - feeling fantastic! This week's goal smashed thanks to my run ... http://rnkpr.com/aih4wis #RunKeeper

Just completed a 5.01 km run with @runkeeper. Check it out! http://rnkpr.com/aij52q #RunKeeper

Just completed a 5.02 km walk with @runkeeper. Check it out! http://rnkpr.com/aih9wm #RunKeeper

Just completed a 5.10 km run with @runkeeper. Check it out! http://rnkpr.com/aihp8 #RunKeeper

Just completed a 5.70 km bike ride with @runkeeper. Check it out! http://rnkpr.com/aiheut #RunKeeper

Just completed a 5.82 km run with @runkeeper. Check it out! http://rnkpr.com/ai5hu #RunKeeper

Just completed a 5.84 km run - Ittapaivarallit metikossa. http://rnkpr.com/aielt4 #RunKeeper

Just completed a 6.14 km bike ride - 😊-http://rnkpr.com/aiy8x #RunKeeper

Just completed a 6.25 mi run with @runkeeper. Check it out! http://rnkpr.com/ahlhfw #RunKeeper

Just completed a 6.84 km walk with @runkeeper. Check it out! http://rnkpr.com/ahwyv #RunKeeper
Just completed a 6.92 mi bike ride with @runkeeper. Check it out! http://rnkpr.com/ah8r7 #RunKeeper

Just completed a 7.06 mi run - Run with @Ivanhoe Thompson who was trying to scam me into believing th... http://rnkpr.com/a9i3s #RunKeeper

Just completed a 7.11 km run with @runkeeper. Check it out! http://rnkpr.com/as5tt #RunKeeper

Just completed a 7.39 km bike ride with @runkeeper. Check it out! http://rnkpr.com/a0i3s #RunKeeper

Just completed a 7.66 km run with @runkeeper. Check it out! http://rnkpr.com/a7j4t #RunKeeper

Just completed a 8.04 km run - hoy ha estado bien. Buen promedio. http://rnkpr.com/aj1kc1 #RunKeeper

Just posted a 0.44 mi elliptical workout - Warming up. http://rnkpr.com/ajb063 #RunKeeper

Just posted a 0.87 mi swim with @runkeeper. Check it out! http://rnkpr.com/ajb06 #RunKeeper

Just posted a 1.20 mi bike ride - Race Attack Program. It Burn Good. http://rnkpr.com/ajb06 #RunKeeper

Just posted a 3.00 mi run - indoor track wake-up jog battling for lanes with the ROTCs; a little jell... http://rnkpr.com/ajb06 #RunKeeper

Just posted a 3.57 mi bike ride with @runkeeper. Check it out! http://rnkpr.com/as5tt #RunKeeper

Just posted a 4.22 mi run with @runkeeper. Check it out! http://rnkpr.com/ah8r7 #RunKeeper


Just posted a 4.80 km run with @runkeeper. Check it out! http://rnkpr.com/ah7b0 #RunKeeper

Just posted a 5.00 mi run with @runkeeper. Check it out! http://rnkpr.com/ajkje #RunKeeper

Just posted a 5.76 mi run with @runkeeper. Check it out! http://rnkpr.com/aknpk #RunKeeper

Just posted a 6.00 mi run with @runkeeper. Check it out! http://rnkpr.com/agzrus #RunKeeper

Just posted a 6.68 km run - Op deze 1e paasdag toch maar even gaan hardlopen. Heerlijk! http://rnkpr.com/ahtl0b #RunKeeper

Just posted a 7.00 mi run - 1548 vertical. http://rnkpr.com/ahtl0b #RunKeeper


lost 0.6 pounds since her last weigh-in! She's lost 1.7 pounds so far. #myfitnesspal

lost 0.6 pounds since her last weigh-in! She's lost 5.2 pounds so far. #myfitnesspal

lost 0.7 pounds since his last weigh-in! He's lost 11.7 pounds so far. #myfitnesspal

lost 1 pound since her last weigh-in! She's lost 12 pounds so far. #myfitnesspal

lost 1 pound since her last weigh-in! She's lost 26 pounds so far. #myfitnesspal

lost 1 pound since his last weigh-in! He's lost 20.5 pounds so far. #myfitnesspal

lost 14 pounds since her last weigh-in! #myfitnesspal

lost 2 pounds since his last weigh-in! #myfitnesspal

lost 3 pounds since his last weigh-in! She's lost 55.7 pounds so far. #myfitnesspal

lost 3 pounds since her last weigh-in! She's lost 6 pounds so far. #myfitnesspal

lost 4 pounds since her last weigh-in! She's lost 16 pounds so far. #myfitnesspal

lost 4 pounds since his last weigh-in! He's lost 80.6 pounds so far. #myfitnesspal

lost 5 pounds since her last weigh-in! #myfitnesspal

Mapped my ride for tomorrow :) http://t.co/WgP8CS9

Need to break things up a bit. How about a laugh? Sure. I could use one. How about a banana. Ok, I'll ta... http://dailymile.com/e/R2OY

Nike+ GPSã" ¤ã'㏦』0.68 kmã''09:32㏧今走り終ｄ㏾ã— ã€' #nikeplus

Nike+ GPSã" ¤ã'㏦』0.88 kmã''06:54㏧今走り終دعو #nikeplus

Nike+ GPSã" ¤ã'㏦』1.05 kmã''08:04㏧今走り終دعو #nikeplus

Nike+ GPSã" ¤ã'㏦』1.10 kmã''07:51㏧今走り終دعو #nikeplus

Nike+ GPSã" ¤ã'㏦』10.0 kmã''59:08ã— ã€' #nikeplus

Nike+ GPSã" ¤ã'㏦』10.0 kmã''59:49ã— ã€' #nikeplus
OK, figured it out. The route appears when you click on the description field, but there’s a catch: only the... http://dailymile.com/e/R19H

One of those days when I fight the depression of life and wonder ... What Would Una Do? Maybe I’ll get out of... http://dailymile.com/e/Q19u

Out of curiosity, I want to track calories in addition to my WW points for a while. MyFitnessPal vs. Lose It, thoughts?

Ran 0.5 kilometers 50 sec and felt good. http://dailymile.com/e/POrp

Ran 1 mile in 8 mins and 30 secs. http://dailymile.com/e/QVlp

Ran 1.75 miles in 26 mins and felt alright. Definitely not a record-setting night for C+C. Kinda just went ou... http://dailymile.com/e/QEsC

Ran 2 miles in 17 mins and felt good. Another injury recovery run... something's working. Body felt more rea... http://dailymile.com/e/Qh3g

Ran 2.01 miles in 23 mins and felt great. TNT Hill Training! My very first experience running hills. What a c... http://dailymile.com/e/OrtP

Ran 2.9 miles in 29 mins. I real slow recovery run and the legs felt good. http://dailymile.com/e/PQCi

Ran 3.1 miles in 33 mins and felt good. http://dailymile.com/e/R5jq

Ran 3.22 miles in 40 mins and felt good. a littl... http://dailymile.com/e/Qh6f

Ran 3.25 miles in 30 mins. sometimes those "F*# it all" runs are really fast http://dailymile.com/e/Ov4Y

Ran 4 miles in 40 mins and felt good. Squeezed in a run after work on a sunny early evening. I love dayli... http://dailymile.com/e/PRUz

Ran 4.05 miles in 38 mins and felt great. It was a great run this morning, but I was ready for my water at th... http://dailymile.com/e/POaE
<table>
<thead>
<tr>
<th>Time</th>
<th>Distance</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4:25</td>
<td>4 miles</td>
<td>Ran 4.25 miles in 40 mins and felt good. Needed to run off that obese amount of trail mix I ate today... <a href="http://dailymile.com/e/S2rJ">http://dailymile.com/e/S2rJ</a></td>
</tr>
<tr>
<td>4:37</td>
<td>4 miles</td>
<td>Ran 4.37 miles in 40 mins and felt good. Felt pretty good after the 8 miles yesterday. Ran an easy 1st mile... <a href="http://dailymile.com/e/PDvV">http://dailymile.com/e/PDvV</a></td>
</tr>
<tr>
<td>4:37</td>
<td>4 miles</td>
<td>Ran 4.37 miles in 43 mins and felt good. Needed to run for 3 days hoping that the rest would get rid of my achill... <a href="http://dailymile.com/e/Orz0">http://dailymile.com/e/Orz0</a></td>
</tr>
<tr>
<td>4:37</td>
<td>5 miles</td>
<td>Ran 5 miles in 50 mins and felt alright. Speed work indoor this am. Hot. Work out within wu and cd = 5x3' far... <a href="http://dailymile.com/e/RzLb">http://dailymile.com/e/RzLb</a></td>
</tr>
<tr>
<td>5:18</td>
<td>6 miles</td>
<td>Ran 5.18 miles in 60 mins and felt alright. #soreheads #momrunning 5:30am, 5-mile run with a friend. We too... <a href="http://dailymile.com/e/Qhr0">http://dailymile.com/e/Qhr0</a></td>
</tr>
<tr>
<td>5:32</td>
<td>34 mins</td>
<td>Ran 5.32 kilometers in 34 mins and felt good. Dungeons isem måa-la pocit, Âš/mi ne to nebĂďÂšĂďÂš ani z kopec. V pĂľĂšcre js... <a href="http://dailymile.com/e/Qqpw">http://dailymile.com/e/Qqpw</a></td>
</tr>
<tr>
<td>5:37</td>
<td>44 mins</td>
<td>Ran 5.37 miles in 44 mins and felt good. After sitting on my tush all day discussing ebooks and libraries, 1... <a href="http://dailymile.com/e/QZXt">http://dailymile.com/e/QZXt</a></td>
</tr>
<tr>
<td>5:48</td>
<td>41 mins</td>
<td>Ran 5.48 miles in 41 mins and felt good. After pounding the snooze button for 30 minutes, i finally rolled ou... <a href="http://dailymile.com/e/On67">http://dailymile.com/e/On67</a></td>
</tr>
<tr>
<td>6</td>
<td>52 mins</td>
<td>Ran 6 miles in 52 mins and felt great. Weather: Partly Clouds. 30c/86f <a href="http://dailymile.com/e/Quhp">http://dailymile.com/e/Quhp</a></td>
</tr>
<tr>
<td>6.25</td>
<td>3 miles</td>
<td>Ran 6.25 miles in 58 mins and felt good. YAY! All the boyz took off and I was resigned to running alone...un... <a href="http://dailymile.com/e/Pxji">http://dailymile.com/e/Pxji</a></td>
</tr>
<tr>
<td>6.45</td>
<td>1 hour</td>
<td>Ran 6.45 miles in 1 hour and 23 sec and felt alright. good hilly run. only a slight pain in my hip...but hydrati... <a href="http://dailymile.com/e/QjS">http://dailymile.com/e/QjS</a></td>
</tr>
<tr>
<td>7.51</td>
<td>1 hour</td>
<td>Ran 7 miles in 1 hour and 4 mins and 38 secs and felt great. It was good to run outside! <a href="http://dailymile.com/e/P71q">http://dailymile.com/e/P71q</a></td>
</tr>
<tr>
<td>7.51</td>
<td>1 hour</td>
<td>Ran 7.51 kilometers in 1 hour and felt alright. Walked more than I wanted today. Perhaps because I didn't ha... <a href="http://dailymile.com/e/QQSh">http://dailymile.com/e/QQSh</a></td>
</tr>
<tr>
<td>8</td>
<td>2 hours</td>
<td>Ran 8 miles and felt good. <a href="http://dailymile.com/e/OdBY">http://dailymile.com/e/OdBY</a></td>
</tr>
<tr>
<td>8.01</td>
<td>1 hour</td>
<td>Ran 8.01 miles in 1 hour and 5 mins. went very well. felt like i could have kept going for a while. looks li... <a href="http://dailymile.com/e/Okgg">http://dailymile.com/e/Okgg</a></td>
</tr>
<tr>
<td>8.1</td>
<td>40 mins</td>
<td>Ran 8.1 miles in 1 hour and 47 mins and felt alright. It felt so much harder than last week, but I came in ar... <a href="http://dailymile.com/e/Qrac">http://dailymile.com/e/Qrac</a></td>
</tr>
<tr>
<td>9</td>
<td>2 hours</td>
<td>Ran 9 miles. Great workout with Hedda B. She killed it, lots of fun! <a href="http://dailymile.com/e/Onrd">http://dailymile.com/e/Onrd</a></td>
</tr>
<tr>
<td>Registered for the Corporate Run 5k! @Nemours <a href="http://www.dailymile.com/events/42997-corporate-run-5k">http://www.dailymile.com/events/42997-corporate-run-5k</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>5 miles</td>
<td>Rode 1.4 miles and felt great. <a href="http://dailymile.com/e/Pptf">http://dailymile.com/e/Pptf</a></td>
</tr>
<tr>
<td>11.64</td>
<td>49 mins</td>
<td>Rode 11.64 kilometers in 49 mins and felt good. Short Ride From NewLuchan to MinesView and Back via Bonifaci... <a href="http://dailymile.com/e/PRmp">http://dailymile.com/e/PRmp</a></td>
</tr>
<tr>
<td>13</td>
<td>50 mins</td>
<td>Rode 13 miles in 50 mins and felt great. Legs were still tired from Sunday, but did a strong workout with &quot;hi... <a href="http://dailymile.com/e/RFcj">http://dailymile.com/e/RFcj</a></td>
</tr>
<tr>
<td>20</td>
<td>45 mins</td>
<td>Rode 20 kilometers in 45 mins and felt good. <a href="http://dailymile.com/e/QsO">http://dailymile.com/e/QsO</a></td>
</tr>
<tr>
<td>52</td>
<td>2 hours</td>
<td>Rode 52 kilometers in 2 hours and 10 mins and felt alright. The New year ride! Yeeoor Hill repeats X5. <a href="http://dailymile.com/e/QP9H">http://dailymile.com/e/QP9H</a></td>
</tr>
<tr>
<td>52</td>
<td>2 hours</td>
<td>Rode and felt great. <a href="http://dailymile.com/e/Rz2A">http://dailymile.com/e/Rz2A</a></td>
</tr>
<tr>
<td>RT @aline_mart: Endomondo Running Workout <a href="http://t.co/o1RMaEyi">http://t.co/o1RMaEyi</a> via @AddThis - AĂŠĂŠĂŠ!!!!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT @dailymile: There is a new free gps iPhone app that integrates with dailymile! Kinetic Lite GPS <a href="http://bit.ly/fAnudF">http://bit.ly/fAnudF</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT @henskobben: Just completed a 15.62 km bike ride. - Zo eerste 'officiële' tijdwaarneming. Na 100km lekker vertrouwd... <a href="http://rmkrp.com/ahpw59">http://rmkrp.com/ahpw59</a> #RunKeeper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT @RoadID: The @dailymile put Road ID through the paces. Find out what 10 reviewers had to say. <a href="http://bit.ly/h8kgen">http://bit.ly/h8kgen</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT @seeksboston26mi: Finally! My 2011 #BostonMarathon Race blog. This marathon sport can be a hu... <a href="http://dailymile.com/e/PeYh">http://dailymile.com/e/PeYh</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scored Running Times and @dailymile stickers in the mail today! Merry Christmas to me!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT @aline_mart: Endomondo Running Workout <a href="http://t.co/o1RMaEyi">http://t.co/o1RMaEyi</a> via @AddThis - AĂŠĂŠĂŠ!!!!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fi...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ced in cime... <a href="http://dailymile.com/e/qWjpF">http://dailymile.com/e/qWjpF</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sim, acho muito legal pra registrar o pedal... na corrida, uso o #nikeplus ;) RT: @aline_mart: @lcrsil_bra vc tb usa Endomondo! !!!!!!!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rode and felt great. <a href="http://dailymile.com/e/Rz2A">http://dailymile.com/e/Rz2A</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rode and felt great. <a href="http://dailymile.com/e/Rz2A">http://dailymile.com/e/Rz2A</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rode and felt great. <a href="http://dailymile.com/e/Rz2A">http://dailymile.com/e/Rz2A</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rode and felt great. <a href="http://dailymile.com/e/Rz2A">http://dailymile.com/e/Rz2A</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT @aline_mart: Endomondo Running Workout <a href="http://t.co/o1RMaEyi">http://t.co/o1RMaEyi</a> via @AddThis - AĂŠĂŠĂŠ!!!!!!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT @dailymile: There is a new free gps iPhone app that integrates with dailymile! Kinetic Lite GPS <a href="http://bit.ly/fAnudF">http://bit.ly/fAnudF</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT @henskobben: Just completed a 15.62 km bike ride. - Zo eerste 'officiële' tijdwaarneming. Na 100km lekker vertrouwd... <a href="http://rmkrp.com/ahpw59">http://rmkrp.com/ahpw59</a> #RunKeeper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT @RoadID: The @dailymile put Road ID through the paces. Find out what 10 reviewers had to say. <a href="http://bit.ly/h8kgen">http://bit.ly/h8kgen</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT @seeksboston26mi: Finally! My 2011 #BostonMarathon Race blog. This marathon sport can be a hu... <a href="http://dailymile.com/e/PeYh">http://dailymile.com/e/PeYh</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td>Distance</td>
<td>Link</td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>Was out cycling 1.31 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/gnB5w8">http://bit.ly/gnB5w8</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 1.44 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/eRQw1">http://bit.ly/eRQw1</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 1.68 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/ICV7AA">http://bit.ly/ICV7AA</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 11.19 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/gTMVDA">http://bit.ly/gTMVDA</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 13.05 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/hZdLs">http://bit.ly/hZdLs</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 13.20 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/fmxJaB">http://bit.ly/fmxJaB</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 14.03 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/hcXFI7">http://bit.ly/hcXFI7</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 14.11 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/h1OGo">http://bit.ly/h1OGo</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 14.61 miles with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/hv1Goi">http://bit.ly/hv1Goi</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 16.31 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/hCcil6F">http://bit.ly/hCcil6F</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 16.73 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/xy5yGA">http://bit.ly/xy5yGA</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 17.45 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/f1Bx76">http://bit.ly/f1Bx76</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 18.31 miles with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/IBQX7">http://bit.ly/IBQX7</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 21.24 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/PTw17">http://bit.ly/PTw17</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 23.97 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/hn259i">http://bit.ly/hn259i</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 26.17 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/gITuK">http://bit.ly/gITuK</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 3.70 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/PTw0bs">http://bit.ly/PTw0bs</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 3.58 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/9dEIbd">http://bit.ly/9dEIbd</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 3.22 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/gM3Ok">http://bit.ly/gM3Ok</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 40.25 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/hKZC1W">http://bit.ly/hKZC1W</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 41.68 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/fJRCZ">http://bit.ly/fJRCZ</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 7.11 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/eYhAPL">http://bit.ly/eYhAPL</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 7.85 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/1Zuc1">http://bit.ly/1Zuc1</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 8.46 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/ed4Bv">http://bit.ly/ed4Bv</a></td>
<td></td>
</tr>
<tr>
<td>Was out cycling 9.75 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/gXgXz">http://bit.ly/gXgXz</a></td>
<td></td>
</tr>
<tr>
<td>Was out exercising 37.04 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/fOYy8">http://bit.ly/fOYy8</a></td>
<td></td>
</tr>
<tr>
<td>Was out hiking 3.34 miles with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/6FLpq">http://bit.ly/6FLpq</a></td>
<td></td>
</tr>
<tr>
<td>Was out mountain biking 16.00 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/e2KNEf">http://bit.ly/e2KNEf</a></td>
<td></td>
</tr>
<tr>
<td>Was out mountain biking 20.47 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/tEXRk">http://bit.ly/tEXRk</a></td>
<td></td>
</tr>
<tr>
<td>Was out mountain biking 23.35 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/xxdLi5">http://bit.ly/xxdLi5</a></td>
<td></td>
</tr>
<tr>
<td>Was out mountain biking 23.35 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/xxdLi5">http://bit.ly/xxdLi5</a></td>
<td></td>
</tr>
<tr>
<td>Was out mountain biking 25.23 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/hw6O6">http://bit.ly/hw6O6</a></td>
<td></td>
</tr>
<tr>
<td>Was out running 0.00 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/fhTB2x5">http://bit.ly/fhTB2x5</a></td>
<td></td>
</tr>
<tr>
<td>Was out running 0.00 miles with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/g9XJxw">http://bit.ly/g9XJxw</a></td>
<td></td>
</tr>
<tr>
<td>Was out running 1.83 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/ha4Oua">http://bit.ly/ha4Oua</a></td>
<td></td>
</tr>
<tr>
<td>Was out running 10.03 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/g5S8Vf">http://bit.ly/g5S8Vf</a></td>
<td></td>
</tr>
<tr>
<td>Was out running 10.28 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/8ZEEAY">http://bit.ly/8ZEEAY</a></td>
<td></td>
</tr>
<tr>
<td>Was out running 11.01 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/gQ57FO">http://bit.ly/gQ57FO</a></td>
<td></td>
</tr>
<tr>
<td>Was out running 13.00 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/gFEP1E">http://bit.ly/gFEP1E</a></td>
<td></td>
</tr>
<tr>
<td>Was out running 14.64 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/L2ZKUB">http://bit.ly/L2ZKUB</a></td>
<td></td>
</tr>
<tr>
<td>Was out running 15.01 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/eRyR1a">http://bit.ly/eRyR1a</a></td>
<td></td>
</tr>
<tr>
<td>Was out running 16.12 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/Eh451">http://bit.ly/Eh451</a></td>
<td></td>
</tr>
<tr>
<td>Was out running 2.42 miles with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/f95c0d6">http://bit.ly/f95c0d6</a></td>
<td></td>
</tr>
<tr>
<td>Was out running 2.87 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/hW8uw7">http://bit.ly/hW8uw7</a></td>
<td></td>
</tr>
<tr>
<td>Was out running 20.25 km with #Endomondo.</td>
<td>See it here: <a href="http://bit.ly/ceCNBk0">http://bit.ly/ceCNBk0</a></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>Distance</td>
<td>Link</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>Running 3.64 miles</td>
<td>3.64</td>
<td><a href="http://bit.ly/lydlyR1">http://bit.ly/lydlyR1</a></td>
</tr>
<tr>
<td>Running 3.70 miles</td>
<td>3.70</td>
<td><a href="http://bit.ly/0MIjaPd">http://bit.ly/0MIjaPd</a></td>
</tr>
<tr>
<td>Running 3.80 km</td>
<td>3.80</td>
<td><a href="http://bit.ly/6oOHH">http://bit.ly/6oOHH</a></td>
</tr>
<tr>
<td>Running 4.13 km</td>
<td>4.13</td>
<td><a href="http://bit.ly/eCUc0f">http://bit.ly/eCUc0f</a></td>
</tr>
<tr>
<td>Running 5.01 miles</td>
<td>5.01</td>
<td><a href="http://bit.ly/oKXEaD">http://bit.ly/oKXEaD</a></td>
</tr>
<tr>
<td>Running 5.21 km</td>
<td>5.21</td>
<td><a href="http://bit.ly/gWVoPx">http://bit.ly/gWVoPx</a></td>
</tr>
<tr>
<td>Running 5.77 km</td>
<td>5.77</td>
<td><a href="http://bit.ly/ehPB50">http://bit.ly/ehPB50</a></td>
</tr>
<tr>
<td>Running 6.97 km</td>
<td>6.97</td>
<td><a href="http://bit.ly/dMm1D">http://bit.ly/dMm1D</a></td>
</tr>
<tr>
<td>Running 7.01 km</td>
<td>7.01</td>
<td><a href="http://bit.ly/i4VwPW">http://bit.ly/i4VwPW</a></td>
</tr>
<tr>
<td>Running 7.35 km</td>
<td>7.35</td>
<td><a href="http://bit.ly/dsU3h7">http://bit.ly/dsU3h7</a></td>
</tr>
<tr>
<td>Skiing 2.58 km</td>
<td>2.58</td>
<td><a href="http://bit.ly/eESCOG">http://bit.ly/eESCOG</a></td>
</tr>
<tr>
<td>Swimming 1.00 km</td>
<td>1.00</td>
<td><a href="http://bit.ly/1Frk10E">http://bit.ly/1Frk10E</a></td>
</tr>
<tr>
<td>Walking 1.15 miles</td>
<td>1.15</td>
<td><a href="http://bit.ly/g35JNn">http://bit.ly/g35JNn</a></td>
</tr>
<tr>
<td>Walking 2.01 km</td>
<td>2.01</td>
<td><a href="http://bit.ly/gYoPDW">http://bit.ly/gYoPDW</a></td>
</tr>
<tr>
<td>Walking 3.06 km</td>
<td>3.06</td>
<td><a href="http://bit.ly/hacoxy">http://bit.ly/hacoxy</a></td>
</tr>
<tr>
<td>Walking 3.78 miles</td>
<td>3.78</td>
<td><a href="http://bit.ly/g0TvV7">http://bit.ly/g0TvV7</a></td>
</tr>
<tr>
<td>Walking 4.18 km</td>
<td>4.18</td>
<td><a href="http://bit.ly/g5Nhr">http://bit.ly/g5Nhr</a></td>
</tr>
<tr>
<td>Walking 5.81 km</td>
<td>5.81</td>
<td><a href="http://bit.ly/hqDzxe">http://bit.ly/hqDzxe</a></td>
</tr>
<tr>
<td>Run 11.20 km</td>
<td>11.20</td>
<td><a href="http://rnkpr.com/ah5w8z">http://rnkpr.com/ah5w8z</a> RKLive RunKeeper</td>
</tr>
<tr>
<td>Run 11.48 mi</td>
<td>11.48</td>
<td><a href="http://rnkpr.com/aifus6">http://rnkpr.com/aifus6</a> RKLive RunKeeper</td>
</tr>
<tr>
<td>Run 5.93 km</td>
<td>5.93</td>
<td><a href="http://rnkpr.com/aifus6">http://rnkpr.com/aifus6</a> RKLive RunKeeper</td>
</tr>
<tr>
<td>Run 5.04 km</td>
<td>5.04</td>
<td><a href="http://rnkpr.com/ahtoe6">http://rnkpr.com/ahtoe6</a> RKLive RunKeeper</td>
</tr>
<tr>
<td>Week 1 Runners Location</td>
<td><a href="http://dailyMile.com/e/0Op5S">http://dailyMile.com/e/0Op5S</a></td>
<td></td>
</tr>
</tbody>
</table>

9.2 Appendix B - Research Ethics Committee Approval

Sent on behalf of Dr Heike Felzmann, Acting Chair, Research Ethics Committee

Dear Mr Vickey

Ethics Application: SURVEY: A Web Framework for Social Networking and Exercise Adherence²

I write to you regarding the above proposal which was submitted for Ethical review. Having reviewed your response to my letter, I am pleased to inform you that your proposal has been granted APPROVAL.
All NUI Galway Research Ethic Committee approval is given subject to the Principal Investigator submitting an annual report to the Committee. The first report is due on or before 31st August 2012. Please see section 7 of the REC's Standard Operating Procedures for further details which also includes other instances where you are required to report to the REC.

Yours Sincerely

Dr Heike Felzmann
Research Ethics Committee
Do as I tweet, not as I do: comparing physical activity data between fitness tweets and Healthy People 2020

Ted Vickey¹, John G. Breslin²

¹Point Loma University, San Diego, CA, USA; ²National University of Ireland, Galway, Ireland

Contributions: (I) Conception and design: T Vickey; (II) Administrative support: T Vickey; (III) Provision of study materials or patients: NA; (IV) Collection and assembly of data: T Vickey; (V) Data analysis and interpretation: T Vickey; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Background: The goal of this research was to compare the self-reported estimates of daily physical-activity data provided to the Healthy People 2020 research team via a telephone survey to the mobile fitness app real-time reporting of physical activity using Twitter.

Methods: The fitness tweet classification data set was collected from mobile fitness app users who shared their physical activity over Twitter. Over 184 days, 2,856,534 tweets were collected in 23 different languages. However, for the purposes of this study, only the English-language tweets were analysed, resulting in a total of 1,982,653 tweets by 165,768 unique users. The information and data gleaned from this data set, which reflected 184 days of continuous data collection, were compared to the results from the Healthy People survey, which were compiled using telephone interviews of self-reported physical activity from the previous week.

Results: The data collected from fitness tweets using the five mobile fitness apps suggest lower percentages of people achieving both the 150 to 300 and 300+ min levels than is reflected in the Healthy People survey results. While employing Twitter and other social media as data-collection tools could help researchers obtain information that users might not remember or be willing to disclose face-to-face or over the telephone, further research is needed to determine the cause of the lower percentages found in this study.

Conclusions: Though some challenges remain in using social media like Twitter to glean physical-activity data from the public, this approach holds promise for yielding valuable information and improving outcomes.

Keywords: mHealth; physical activity; Twitter; mobile fitness apps

Received: 29 September 2015; Accepted: 27 November 2015; Published 30 November 2015.

doi: 10.3978/j.issn.2306-9740.2015.11.01

View this article at: http://dx.doi.org/10.3978/j.issn.2306-9740.2015.11.01

Introduction

The promotion and monitoring of physical activity have been a focus of public health efforts in recent years. However, objectively measuring population-level physical activity is challenging because it requires tracking a large number of people using expensive devices and imposing strict data-collection protocols (1). That said, emerging technology can provide reliable and valid alternative surveillance tools for self-reported measures of physical activity (1). Although there is an increase in the number of studies using integrated sensor technology to collect physical-activity data on a population level, there is little technical guidance for researchers who want to use this technology within their research (2).

According to Graham and Hipp, “Physical activity measurement research is achieving greater ease of use, precision and scope by incorporating emerging technologies. These emerging technologies are noteworthy because they can: greatly increase external validity of measures and findings through ease of use and transferability; significantly increase the ability to analyze patterns; improve the ongoing, systematic collection and analysis of public health surveillance due to real-time capabilities; and address the need for research about the cyber infrastructure required to cope with big data.” (3).
In 2010, the U.S. Department of Health and Human Services published the fifth instalment of the national report on health and wellness, reflecting the strong state of the science supporting the health benefits of regular physical activity based on the accomplishments of previous Healthy People initiatives (4). The report, entitled Healthy People 2020, introduced new 10-year objectives for health promotion and disease prevention. New to the objectives is “myHealthyPeople,” a challenge for technology application developers. The research discussed here reflects an attempt to meet that challenge. The use of the fitness tweet classification model which was developed for this study enables researchers to collect ongoing data in real time, which is a sharp contrast to phone interviews that rely on participant recall. The biggest challenge in using technology to track physical activity lies in accounting for the fact that many users are inconsistent in their use of the tracking devices.

One component of Healthy People 2020 involves physical activity, suggesting that Americans should engage in at least 150 minutes per week of moderate-intensity physical activity to obtain substantial health benefits and more than 300 minutes per week to obtain more extensive health benefits.

Current baseline and targets are presented in Table 1.

In 2008, when the goals and objectives for Healthy People 2020 were first developed, 43.5% of American adults met the goal of 150 min per week of moderate-intensity physical activity, with only 28.4% reaching 300 min per week (5).

### Methods

For this research project, a comparison between the collected physical-activity data provided in the Healthy People 2020 report and physical-activity data collected from five mobile fitness apps (Nike+, DailyMile, MyFitnessPal, Endomondo and RunKeeper) as publicly shared over Twitter was conducted.

Each mobile fitness app used in this research had a standard word phrase for the automatic sharing of physical activity using fitness tweets that include time and/or distance of the physical activity. In addition, some mobile fitness apps included in the standard word phrase a shortened URL that directed back to the mobile fitness app’s user page.

On that page, additional information not included in the fitness tweet could be collected (Figure 1). A data-scraping script was written to collect this information. Once the data were collected, a data cleaning removed low totals (less than 15 min of reported physical activity over 28 weeks) and high totals (more than 30,000 min of reported physical activity over 28 weeks) in order to account for one-time users or user error and invalid results stemming from technology-related issues (e.g., a fitness app being left open after a workout is completed, which would inflate the numbers and skew the data).

### Data

Data for this research was from two data sets:

(I) Healthy People 2020;

(II) Fitness Tweet Classification Data Set.

Results from the Healthy People survey were compiled using telephone interviews of self-reported physical activity from the previous week. There are considerable concerns about this methodology, as physical-activity questionnaires show limited reliability and validity (6). Even so, they have long been considered the only feasible means of collecting data in large populations, despite the fact that researchers know that responses can be influenced by cultural factors, language barriers and recall accuracy, particularly in older populations (6). One aim of this research study is to explore the use of Twitter as a more reliable and valid alternative.

The fitness tweet classification data set was collected from mobile fitness app users who shared their physical

### Table 1 Healthy people 2020 baseline and targets

<table>
<thead>
<tr>
<th>Measure</th>
<th>Baseline (%)</th>
<th>Target (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduce the proportion of adults who engage in no leisure-time physical activity</td>
<td>36.2</td>
<td>32.6</td>
</tr>
<tr>
<td>Increase the proportion of adults who engage in aerobic physical activity of at least moderate intensity for at least 150 min/week</td>
<td>43.5</td>
<td>47.9</td>
</tr>
<tr>
<td>Increase the proportion of adults who engage in aerobic physical activity of at least moderate intensity for more than 300 min/week</td>
<td>28.4</td>
<td>31.3</td>
</tr>
</tbody>
</table>
activity over Twitter. Over 184 days, 2,856,534 tweets were collected in 23 different languages. However, for the purposes of this study, only the English-language tweets were analysed, resulting in a total of 1,982,653 tweets by 165,768 unique users.

The fitness tweet classification model (7) was used to classify each tweet into main categories of activity, blarney and conversation and then into subcategories as shown in Figure 2.

**Results**

In total, 102,544 users mentioned workout duration in their tweets, accounting for 2.4 million min of physical activity. The addition of workout type, duration and distance allowed additional analysis to be conducted. Physical activity is a sporadic and complex behaviour to measure, but previous research suggests that three days of accelerometer data, four days of pedometer data or 4 days of physical-activity logs are needed to reliably measure physical-activity levels in older adults (8).

As demonstrated in Figure 3, the data collected from
fitness tweets using the five mobile fitness apps suggest lower percentages of people achieving both the 150 to 300 and 300+ min levels. The lower percentage for the 150 to 300 min range was expected, as it is difficult to know whether a person used their mobile fitness app during every workout session. What was a bit more surprising was the lower percentages for the 300+ min levels, as the more active population might be expected to be more dedicated users of their wearable devices and mobile fitness apps.

Consider, for example, users of Nike+, which could be considered the most physically active overall group, as Nike targets athletic shoe buyers through social media channels. Analysis of the fitness tweet conversations also indicated a higher number of mobile fitness app users using Nike+ to train for 5K, 10K, half-marathon and full-marathon events. That being the case, the hypothesis was that of the five mobile fitness apps, Nike+ would be one of the more used mobile fitness apps within the 300+ min category due to the daily training regimens of the participants. Figure 4 highlights the data analysis suggesting that weekly Nike+ users fitness tweet an average of 81 min per week of physical activity. In fact, RunKeeper, which is also geared toward runners, reported the second lowest average weekly minutes of physical activity, with just over 104 min per week.

The overall variance in the data derived from those who completed the Healthy People 2020 survey and mobile fitness app fitness tweeters could be due to users not sharing all of their physical activity via Twitter and/or an overestimation of weekly minutes of exercise collected during the phone surveys for Healthy People 2020. Table 2 shows how one aspect of physical-activity data collected from Twitter can be presented. To maintain confidentiality of the users, Twitter user names were replaced with generic ‘User xxx’ labels. It is important to determine the user’s first user date of the mobile fitness app within the data-collection period, as the data-collection timeframe is just a snapshot over time. A user could have already been using the mobile fitness app and sharing the data before the start of the data-collection period. Cells that contain the label “X” indicate that the first use date of the mobile fitness app by the user occurred after the week header. For example, the first use date for User 13 occurred sometime in week 4. It is also important to be able to determine gaps of weekly usage over time, showing that a user is not consistent in the sharing of physical-activity data from mobile fitness apps using Twitter, or that the user simply did not exercise for a time due to injury, illness, vacation, etc.

Discussion

This case study presents a comparison between weekly minutes of physical activity derived from Healthy People 2020 survey results and fitness activity tweets of mobile fitness app users, and provides physical-activity researchers an alternative method of data collection that could be more reliable than self-reported physical-activity survey data. The issue of why this research yielded lower percentages of physical activity than the Healthy People phone survey remains unaddressed. Is it possible that the Twitter data is more accurate and that people are over-reporting their activity levels over the phone? Can further research derive a means of accounting for any under-reporting that is taking place via Twitter? Recall bias is a considerable issue in phone surveys, as people tend to overestimate their physical activity and underestimate their sedentary time; thus, researchers have developed ways to account for this bias when analysing the resulting data (9). This needs to be done for Twitter-based data as well, but ongoing, real-time data analysis is an invaluable resource for researchers that should eventually prove to be more reliable then recall-based phone surveys.

This active data collection could provide numerous benefits when compared to passive data collection. For example, some evidence already suggests that the knowledge that their activities are being monitored could impact participants’ weekly minutes of physical activity (8). While this may be problematic in a research setting, as described above, it can lead to true lifestyle change in individuals who use social media to motivate themselves to stay on track.
Obtaining information from social media allows for crowdsourced participation, which can provide much more data diversity in terms of greater range of age, geography and ethnicity of users. Moylan, Derr and Lindhorst found that mobile technology was especially useful in reaching out to participants who were previously inaccessible due to geography or physical disability (10). Employing Twitter and other social media as data-collection tools could help researchers obtain information that users might not remember or be willing to disclose face-to-face or over the telephone. Ahlwardt found that patients are often willing to reveal information about their personal healthcare experiences on Twitter, allowing healthcare providers to glean insight on how to improve communication with...
patients and treat them more effectively (11). Because users are often relaying information in real time, some researchers posit that the personal details users share may be more accurate than data collected by traditional methods.

Social media data collection also provides the added benefit of allowing researchers to access more people in their target population over a shorter amount of time. Furthermore, Casler and College [2013] discovered that participants who signed up for studies online performed behavioural tasks just as well as people who participated face-to-face or over the phone (12). Information collected over social media has also provided additional useful healthcare data, from the presentation of menopause symptoms in women to the prevalence of children with ulcerative colitis. Healthcare practitioners can also access additional information through these methods, including demographics, current medication lists and potential diagnoses.

Limitations

A common limitation in this type of research lies in the fact that adolescents and adults do not always accurately report physical-activity levels (13), with many underreporting sedentary behaviours and over-reporting exercise (9).

As this research used all Twitter data (i.e., users were not assigned a specific mobile fitness app to share physical-activity data), one cannot assume that when a user reported zero minutes of weekly activity, this means that the user actually performed no physical activity during that period. There could have been any number of user, device, data collection or Twitter errors. Other than sending a tweet to each individual user, it would be difficult to determine the reason for the lack of data.

An additional limitation is the actual definition of physical activity. During the original data collection for the Healthy People project, depending on how the question was phrased, the respondent may have answered in one of two ways. First, he or she could have provided the number of minutes he or she performed traditional physical activity by going to the gym or going outside for a run, for example. Second, the respondent could have included all physical activity, including non-structured activity such as walking in a mall. The data set of the Fitness Tweets would suggest that this data is a collection of exercise-type activity rather than ongoing measurement, as such measurement would be difficult due to battery issues throughout the day.

While we are confident that during the data-collection process we had access to the Twitter firehose allowing for the collection of all publicly available tweets, there is no way to verify this without actually purchasing all of the tweets. There remains a challenge in the extraction of useful data within these repositories through data mining and knowledge discovery (14). Researchers could enhance our model by purchasing commercially available data sets for analysis in future studies.

While we created a very potential tool for large-scale research by collecting physical-activity data from Twitter, the demographics used in this research could suggest a bias in terms of the users of the mobile fitness apps and thus under-represent certain groups. If researchers wish to use Twitter and mobile fitness apps for physical-activity research, additional steps would need to be taken to ensure that all groups are represented in the data samples collected from Twitter.

These findings and interpretations should be regarded as exploratory and speculative, as they represent what can be potentially done in a short development time and with ease of use for non-computer programming health-promotion researchers.

Future work

Advancements in technology design for both smartphones and wearables allow for continuous monitoring of physical activity without a drain in battery life. Depending on the sharing ability, physical activity could be measured by hour or even by minute, thus providing an even greater detail of recorded physical activity.

One benefit of using the fitness tweet classification model was that the database included 184 days of continuous data collection, which stands in stark contrast to the one-week recall used in the Healthy People project. While not every subject had daily physical-activity measures, the same is true with the survey respondents in the Healthy People project. One future area of work could be the determination of how many days’ worth of fitness tweets would be needed to reliably measure physical activity.

Future research could also involve a study that uses fitness tweeting as a more effective data-collection tool, with participants understanding what is being measured and the need to share all physical-activity sessions—as opposed to passive data collection. Knowing they are being monitored could impact participants’ weekly minutes of physical activity, but perhaps not in longer-term studies. Because this type of research can be conducted on an ongoing basis,
the phenomenon of study participants outperforming their usual activity levels should dissipate over time as they return to their usual behavioural patterns.

**Conclusions**

Technologies currently used in other fields could be adopted for physical-activity measurement. This research used one such technology—Twitter—and created a method to collect physical-activity data from publicly available tweets. The precise measurement of physical activity, including type, amount, context and place is essential for increasing physical activity (2). While this approach shows promise in data collection, future research on how to account for user inconsistency in terms of reporting physical activity is needed before Twitter-based data can be considered truly reliable, but it is clear that Twitter, other forms of social media and smartphone apps are here to stay. Health and fitness professionals and researchers in this area would benefit from leveraging the ever-growing population of users in their work.

**Acknowledgements**

This research was conducted under grant number SFI/08/CEI1380 from the Science Foundation Ireland and IRCSET.

**Footnote**

**Conflicts of Interest:** The authors have no conflicts of interest to declare.

**References**

Online Influence and Sentiment of Fitness Tweets: Analysis of Two Million Fitness Tweets

Theodore Vickey1,2*, MBS; John G Breslin1*, PhD

1College of Engineering & Informatics, National University of Ireland Galway, Galway, Ireland
2Department of Kinesiology, Point Loma University, San Diego, CA, United States
*all authors contributed equally

Corresponding Author:
Theodore Vickey, MBS
National University of Ireland Galway
College of Engineering & Informatics
NUI Galway
University Road
Galway, H91 TK33
Ireland
Phone: 353 858 859 7111
Email: ted.vickey@gmail.com

Abstract

Background: Publicly available fitness tweets may provide useful and in-depth insights into the real-time sentiment of a person’s physical activity and provide motivation to others through online influence.

Objective: The goal of this experimental approach using the fitness Twitter dataset is two-fold: (1) to determine if there is a correlation between the type of activity tweet (either workout or workout+, which contains the same information as a workout tweet but has additional user-generated information), gender, and one’s online influence as measured by Klout Score and (2) to examine the sentiment of the activity-coded fitness tweets by looking at real-time shared thoughts via Twitter regarding their experiences with physical activity and the associated mobile fitness app.

Methods: The fitness tweet dataset includes demographic and activity data points, including minutes of activity, Klout Score, classification of each fitness tweet, the first name of each fitness tweet user, and the tweet itself. Gender for each fitness tweet user was determined by a first name comparison with the US Social Security Administration database of first names and gender.

Results: Over 184 days, 2,856,534 tweets were collected in 23 different languages. However, for the purposes of this study, only the English-language tweets were analyzed from the activity tweets, resulting in a total of 583,252 tweets. After assigning gender to Twitter usernames based on the Social Security Administration database of first names, analysis of minutes of activity by both gender and Klout influence was determined. The mean Klout Score for those who shared their workout data from within four mobile apps was 20.50 (13.78 SD), less than the general Klout Score mean of 40, as was the Klout Score at the 95th percentile (40 vs 63). As Klout Score increased, there was a decrease in the number of overall workout+ tweets. With regards to sentiment, fitness-related tweets identified as workout+ reflected a positive sentiment toward physical activity by a ratio of 4 to 1.

Conclusions: The results of this research suggest that the users of mobile fitness apps who share their workouts via Twitter have a lower Klout Score than the general Twitter user and that users who chose to share additional insights into their workouts are more positive in sentiment than negative. We present a novel perspective into the physical activity messaging from within mobile fitness apps that are then shared over Twitter. By moving beyond the numbers and evaluating both the Twitter user and the emotions tied to physical activity, future research could analyze additional relationships between the user’s online influence, the enjoyment of the physical activity, and with additional analysis a long-term retention strategy for the use of a fitness app.

doi:10.2196/publichealth.8507

KEYWORDS
Twitter; physical activity; mobile fitness apps; fitness tweet classification; sentiment
Introduction

Physical activity can reduce the risk for many different types of chronic diseases and can help people maintain a healthy weight. Although this knowledge is widely known, adults and children in many countries do not get recommended amounts of physical activity [1]. Recent advances in physical activity monitoring now provide researchers with unparalleled opportunities to increase and improve our understanding of the health benefits of physical activity by assessing daily quantities of activity, patterns, and trends [2], as well as the real-time sentiment of physical activity. Research suggests that technology is one factor that has contributed to the increase in sedentary behavior and decrease in physical activity, but it has also led to a number of innovative physical activity interventions [1].

One such innovation is through the use of mobile fitness apps and the sharing of one’s workout through a social network. This paper will focus on the collection of self-reported fitness data through a mobile fitness app that is then shared with one’s social network via Twitter. The dataset of these tweets along with other connected datasets of demographic information allows for a number of analyses, including but not limited to the potential influence of such tweets and the sentiment of these tweets. By combining the digital traces as people interact through mobile phones and emerging technology may now provide novel methods to assess a range of factors objectively and with minimal expense and burden to participants [3]. This paper will review both the potential online influence and the sentiment of the shared fitness tweets.

Social media has changed the way society is exposed to information [4]. Social networking sites such as Twitter have developed into increasingly useful platforms for the general public to share thoughts, ideas, and opinions. Twitter is a free social networking platform that is widely used around the world by businesses and individuals and is considered one of the most widely used microblogging platforms with 328 million monthly active users with more than 1 billion unique monthly visits to sites with embedded tweets with a mission to “to give everyone the power to create and share ideas and information instantly, without barriers” [5]. Twitter users can rapidly and directly share with and respond to a massive audience, using messages of 140 characters or less. With the creation and introduction of newly developing technologies such as Twitter, new opportunities to obtain global health data that may circumvent the limitations of traditional data sources used in population health and physical activity research are now available [3].

At the same time, these publicly shared data are resulting in vast and growing user-contributed repositories of data [6]. Twitter provides user-generated data that can be collected and analyzed to examine opinions around health-related foci, including discussions about physical activity, alcohol and marijuana use, depression, and suicide [3]. From a health-promotion standpoint, these data can be useful to measure participants’ dependence on social support, given that exercisers today are just as, if not more, likely to seek motivation and validation from social media—in particular, Twitter—than their in-person friends and family members [7]. Because it is possible to glean precise information from tweets, including the time of the tweet and location of the user, this suggests that the 140-character messages could be predictive in other areas, such as the types of physical activity that users engage in and where and when they engage in these activities.

Using Twitter integration with mobile fitness apps can be a helpful tool for obtaining descriptive and predictive real-time shared health information in a noninvasive way. New and innovative cloud-based data collection and analysis tools may aid research efforts because they can yield a large collection of tweets in a short period of time. They may also be useful for longitudinal data collection [8]. The link between publicly available health and fitness data sources is made possible as more users publicly share their self-collected data from devices and apps through social media services such as Twitter [9]. An enhanced understanding of mobile fitness apps and the sharing of physical activity through one’s social network, the different types of measurement properties, and the subsequent generated data are critical to furthering our understanding of daily physical activity.

Sentiment analysis is a classification process, the primary focus of which is to predict the polarity of words and to then classify these words as positive, negative, or neutral with the aim of identifying attitude and opinions [10]. Specific to Twitter, sentiment analysis is the task of automatically identifying and extracting subjective information from tweets. This method of data analysis has received increasing attention from the Web-mining community [11]. Although Twitter provides extremely valuable insight into publicly shared opinions, it also provides new big data challenges, including the processing of massive volumes of data and the identification of human expressiveness within short text messages [11]. Much of the existing research on textual information processing has been focused on the mining and retrieval of factual information, with little research on the processing of opinions [12].

The mining of Twitter for data provides a rich database of information on people’s thoughts and sentiments about a myriad of health topics, including physical activity. Analysis of social networks data using Twitter has become a powerful tool that is currently being used to answer research questions across the health spectrum, including local and national flu surveillance [13], the sharing of information between cancer patients [14], marijuana usage among teens [15], and drug safety surveillance [16]. This paper represents, to the best of our knowledge, the first analysis of shared tweets from mobile fitness apps specific to physical activity. A significant proportion of tweets contained nonneutral sentiments regarding the shared physical activity of the four mobile apps featured in this research.

The ability to evaluate the sentiment of an individual immediately after a bout of physical activity has been completed can be powerful. A typical tweet might include the type of exercise performed, the duration and intensity of that exercise, and how the person felt during and after the activity. If the sentiment is negative (eg, “Just hiked to the top of Mt Pisgah. Took me 2 hours and I’m completely exhausted. Don’t think I’ll do that again! #myfitnesspal”), a coach or trainer can intervene and modify the activity accordingly. Finding exercise
that is enjoyable and of the appropriate intensity is an important precursor to long-term adherence. Behavioral researchers suggest that one’s emotions can profoundly affect individual behavior and decision making [17]. Simply stated, a tweet can be a window into real emotion provided in real time.

Other research reported that when fitness promoters initiated a #PlankADay challenge on Twitter—which was designed to encourage core-strengthening exercise—72% of users participated for at least 30 days straight and at the end of the challenge reported an increased enjoyment of the activity and expressed interest in continuing to do abdominal exercise [18]. This indicates that Twitter and other social networks can be useful in spreading exercise awareness and encouraging positive exercise behaviors. Together, this information can facilitate research on how technology can be used to monitor and motivate physical activity and how online social networks may play a role in physical activity promotion and adherence. Identifying the types of people who use mobile fitness apps and finding ways to track what they do and motivate them to continue to engage in physical activity is a form of data mining for this “customer base.”

**Methods**

**Collection of Tweets**

After a review of online tools that could collect and manage tweets, an open-source program called TwapperKeeper was deemed appropriate as the Twitter data-collection tool. TwapperKeeper is a Web app designed to collect social media data via Twitter for long-term archival and analysis. The app uses a Twitter-enabled application program interface (API) that acts as an interface between the Twitter search function and a cloud database for tweet storage [19].

For this research, we chose four mobile fitness apps based on their availability on iPhone, the ability of the mobile fitness app to share workout information through Twitter, and the fact that they targeted beginner versus experienced exercisers. The research team used these criteria to narrow possible choices and reviewed additional academic research for previously used apps, researched publicly available reviews on different mobile fitness apps, interviewed both developers and users of mobile fitness apps to obtain their input, and met as a group to finalize the selected mobile fitness apps to study [20].

The four apps chosen were Endomondo, Nike+, RunKeeper, and DailyMile. Tweets were then collected from the mobile fitness apps using the following hashtags: #endomondo, #nikeplus, #runkeeper, and #dailymile. These were used because these apps automatically attach these hashtags to a tweet to indicate it has come from that particular mobile fitness app. It is through these hashtags that common themes or information can be grouped within Twitter.

Data collection using TwapperKeeper continued for 184 days. During this period, 2,856,534 user-generated mobile fitness app tweets were collected in 23 different languages. The Twitter data in this study was public, and the research was deemed exempt from human subjects review. This research was approved by the institutional review board of the National University of Ireland Galway in Galway, Ireland.

Two analyses were completed on a dataset of collected tweets from four mobile fitness apps. The first was to measure the online influence of Twitter users through their Klout Score. The second was to measure the sentiment of physical activity-related tweets.

**Analysis 1: Measuring Online Influence**

One important factor to consider when analyzing tweets to report physical activity is the credibility and authority of the person sending the tweets. Previous data collectors have looked at a Twitter user’s number of followers, although researchers discovered that monitoring retweets and the messages themselves are a better predictive tool [21].

Websites such as Klout have developed the means to determine a user’s reach or influence on social media. The Klout Score is the measurement of a person’s overall online influence, with scores ranging from 1 to 100; higher scores represent a wider and stronger sphere of influence. Scores greater than 50 are rare [22]. A Klout Score places less emphasis on a user’s number of followers and number of tweets, but rather measures the extent to which the user’s content is retweeted [23]. One’s influence on Twitter can be difficult to measure accurately. Klout uses more than 3600 features that capture the online social network activity of the user to conduct the influence analysis and calculate the Klout Score [24]. The Klout Score allows for tailored statistical analysis of social media usage and is tangible proof of the effect of the Internet on a person’s lifestyle [25].

With regards to influence, Internet users perceived a mock Twitter page with a high Klout Score as more credible than the same page with a moderate or low Klout Score [26].

Online influence services such as Klout are in the process of scoring millions, eventually billions, of people on their level of influence. To proponents, the measurement of online influence is an inspiring tool that encourages the democratization of influence, where one no longer must be a celebrity, politician, or media personality to be considered influential.

**Recruitment**

For this experimental approach, the user’s Klout Score—a measure of their online influence—was used to compare shared physical activity levels from mobile fitness apps. In this experiment, we examined the sharing of fitness tweets from within mobile fitness apps (Nike+, RunKeeper, DailyMile, and Endomondo) and analyzed the data based on the participant’s gender and online influence, as measured by their Klout Score. We identified two types of activity tweets from dataset: workout tweets, which included what was generated by the mobile fitness app, and workout+ tweets, which included the same information as a workout tweet but also contained user-created communication. We hypothesized that those with a higher Klout Score would share fewer minutes of activity and more overall workout+ tweets. We also hypothesized that across both genders, the higher the Klout Score, the lower the minutes of shared physical activity.
The data for this research were drawn from an existing dataset of fitness tweets from mobile fitness app users who shared their physical activity and, in some cases, additional conversation over Twitter. Over 184 days, 2,856,534 tweets were collected in 23 different languages. However, for the purposes of this study, only the English-language tweets were analyzed from the activity tweets, resulting in a total of 583,252 tweets.

The Fitness Tweet Classification Model [20] was used to classify each tweet into main categories of activity, blarney, and conversation and then into subcategories as shown in Figure 1.

The different types of collected information from the mobile fitness apps and each corresponding Twitter account provided a number of different and unique data points to review. For this experiment, those data points included the activity tweets, the user’s gender, the minutes of physical activity, and the user’s Klout Score. The statistical analysis of physical activity on Twitter from the four selected mobile fitness apps was performed in SAS 9.3, a software suite developed by the SAS Institute for advanced analytics, business intelligence, and predictive analytics, using two key datasets: (1) the first dataset included all user information about Twitter users who sent tweets relating to workout and workout+ and (2) the second dataset contained all the actual tweets sent by each user.

**Analysis 2: Sentiment Analysis**

**Recruitment**

Of the activity tweets, there were a total of 408,574 workout+ tweets. From this total, a random sample of 23,391 was created. These tweets were user-generated, where the end user provided additional text to a workout tweet (i.e., the user provided supplementary information beyond that which was created by the app itself). Tweets were then grouped by mobile fitness app using the corresponding hashtags. There were no significant numbers of emojis available in the fitness tweets for use in the sentiment analysis.

**Sentiment Analysis of Tweets**

The AYLIEN Text Analysis for Google Sheet add-on was utilized for the analysis of the sentiment for each collected information-sharing conversation tweet as filtered by the Fitness Tweet Classification Model.

The AYLIEN Tweet Sentiment Analysis function is a three-step process:

1. **Preprocessing:** tweets are normalized and reformatted, and the parts that are considered irrelevant to the sentiment are stripped.
2. **Parsing:** tweets are parsed and their structure, tags, and negations are extracted.
3. **Classification:** tweets are classified as positive, negative, or neutral by a pretrained classifier, assisted by a lexicon-based approach as a second judge.

For this experiment, the sentiment analysis tool that analyzed each tweet and returned the value of positive, neutral, or negative was used for classification. These data were saved into an Excel spreadsheet for additional data processing by converting the text value to a numerical value (positive=1, neutral=0, and negative=-1).

**Results**

**Analysis 1: Measuring Online Influence**

**Gender Assignment of Twitter Users Within the Dataset**

Twitter does not collect the gender of users. To be able to compare across genders, a means of identifying the possible gender of the Twitter users was needed. To accomplish this, we used the US Social Security Administration’s name database to match English names with gender. The name database from the Social Security Administration website included popular names ranked by gender since 1880.

The first gender-match calculation between the first names in the collected Twitter demographic database (the Twitter user’s full name was one of the many demographic characteristics
collected from Twitter) and the Social Security Administration database eliminated names that were used fewer than 200 times because many such names were much more popular among one gender than another (e.g., girls were named Aaron <0.5% of the time). The assumption was that this adjustment eliminated a vast majority of gender confusion among names. Once this was completed, names were matched to genders using the VLOOKUP function in Excel.

A second gender-match calculation was performed for those Twitter users with names that appeared less than 200 times, in which we attempted to assign gender to the remaining names that did not match in the first round. Usernames that did not match either gender (<2%) were not included in the analysis.

After gender assignment, a descriptive statistical analysis was performed to compute the frequency of the following: (1) total minutes by gender, (2) total minutes by Klout Score, (3) total minutes by gender and Klout Score, (4) total number of tweets, (5) minutes exercised per tweet, and (6) total number of workout and workout+ tweets (separately).

### Determination of Klout Quartiles
To examine the distribution of tweets, minutes of exercise described by said tweets and the categories mentioned in each tweet (workout or workout+), it was necessary to separate the users’ Klout Scores into quartiles. We used the quartile method of data classification to create categories with a rank-ordered dataset split into four equal parts.

This was done through a two-step process in SAS. First, the distribution of Klout Scores was examined using the univariate procedure in SAS (PROC UNIVARIATE) and assigned quartiles based on that distribution. Second, using a data step, values of 1, 2, 3, and 4 were assigned to observations within the first, second, third, and fourth quartiles, respectively (Table 1). The maximum of any Klout Score is 100 and the minimum is 1. It was determined that the median Klout Score from the collected dataset was 20.50. As reported by Klout, the mean Klout Score is 40, with users with a score of 63 ranked in the 95th percentile [27].

<table>
<thead>
<tr>
<th>Quartile</th>
<th>Klout Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Maximum</td>
<td>100.00</td>
</tr>
<tr>
<td>99%</td>
<td>56.59</td>
</tr>
<tr>
<td>95%</td>
<td>49.03</td>
</tr>
<tr>
<td>90%</td>
<td>44.09</td>
</tr>
<tr>
<td>75% Q3</td>
<td>35.65</td>
</tr>
<tr>
<td>50% Median</td>
<td>20.50</td>
</tr>
<tr>
<td>25% Q1</td>
<td>11.92</td>
</tr>
<tr>
<td>10%</td>
<td>10.10</td>
</tr>
<tr>
<td>5%</td>
<td>10.00</td>
</tr>
<tr>
<td>1%</td>
<td>10.00</td>
</tr>
<tr>
<td>0% Minimum</td>
<td>1.00</td>
</tr>
</tbody>
</table>

### Number of Activity Tweets (Male Versus Female)
The descriptive statistical analysis found that males produced 57.9% (336,109/583,252) of the total of activity tweets, whereas females produced 42.1% (247,143/583,252). This difference was consistent across Klout quartiles (Table 2).

### Number of Tweets (Male/Female Among Workout Groups)
The descriptive analysis was expanded to compare males and females in the activity category. It was found that both genders tweeted far more among the workout group than the workout+ group (72.01%, 420,010/583,252 vs 27.99%, 163,242/583,252) in the lowest Klout quartile. This trend decreased slightly through the second and third Klout quartiles and then dramatically among the highest quartile of Klout Scores. In that quartile, the number of tweets varied much less (56.79%, 70,229/123,656 vs 43.21%, 53,427/123,656).

### Mean Minutes Per Tweet (Males Versus Females)
The ANOVA procedure (PROC ANOVA) within SAS was used to compare the mean number of minutes tweeted by each gender using gender in the class statement and setting the model as minutes=gender. It was found that, overall, the mean number of minutes tweeted did not vary significantly between males and females. However, the mean number of minutes tweeted was almost double among females of the lowest Klout Score quartile (Klout ≤11.92).

### Tests of Significance Between Groups: Minutes Tweeted Between Workout Categories
After assigning quartiles, we examined the frequency of observations within each stratum of Klout Scores using PROC FREQ in SAS for the following (Table 3): (1) minutes by Klout Score quartile and (2) exercise types by Klout Score quartile.

---

**Table 1.** Klout Score quartiles.

<table>
<thead>
<tr>
<th>Quartile</th>
<th>Klout Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Maximum</td>
<td>100.00</td>
</tr>
<tr>
<td>99%</td>
<td>56.59</td>
</tr>
<tr>
<td>95%</td>
<td>49.03</td>
</tr>
<tr>
<td>90%</td>
<td>44.09</td>
</tr>
<tr>
<td>75% Q3</td>
<td>35.65</td>
</tr>
<tr>
<td>50% Median</td>
<td>20.50</td>
</tr>
<tr>
<td>25% Q1</td>
<td>11.92</td>
</tr>
<tr>
<td>10%</td>
<td>10.10</td>
</tr>
<tr>
<td>5%</td>
<td>10.00</td>
</tr>
<tr>
<td>1%</td>
<td>10.00</td>
</tr>
<tr>
<td>0% Minimum</td>
<td>1.00</td>
</tr>
</tbody>
</table>

---

http://publichealth.jmir.org/2017/4/e82/
groups (workout vs workout+) and found a statistically significant difference ($P=.01$; Table 4).

**Analysis 2: Sentiment Analysis**

**Sentiment Analysis of Workout+ Tweets**

In total, there were 23,391 unique tweets within the original dataset that fit the filtering criteria from this random sample. Four of the mobile fitness apps were used in this analysis: DailyMile, Endomondo, Nike+, and RunKeeper. The overall sentiment of all mobile fitness apps suggests that half of these workout+ activity tweets were neutral in nature (Table 5). In addition, there were four times as many positive tweets than negative. The breakdown of sentiment analysis for negative, neutral, and positive sentiment by mobile fitness apps is also presented in Table 5.

Table 2. Klout Score by activity tweet (N=583,252) and gender.

<table>
<thead>
<tr>
<th>Quartile and Klout Score</th>
<th>Activity tweets, n (%)</th>
<th>Male (n=336,109)</th>
<th>Female (n=247,143)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: ≤11.92 (n=179,831)</td>
<td>102,007 (56.7)</td>
<td>77,824 (43.3)</td>
<td></td>
</tr>
<tr>
<td>2: &gt;11.93 and ≤20.50 (n=154,669)</td>
<td>89,822 (58.1)</td>
<td>64,847 (41.9)</td>
<td></td>
</tr>
<tr>
<td>3: &gt;20.51 and ≤35.65 (n=125,096)</td>
<td>73,394 (58.7)</td>
<td>51,702 (41.3)</td>
<td></td>
</tr>
<tr>
<td>4: &gt;35.65 (n=123,656)</td>
<td>70,886 (57.3)</td>
<td>52,770 (42.7)</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Workout and workout+ tweets by Klout quartile.

<table>
<thead>
<tr>
<th>Quartile and Klout Score</th>
<th>Workout tweets (n=420,010)</th>
<th>Minutes per tweet, mean (SD)</th>
<th>Total Minutes</th>
<th>Tweets (% total)</th>
<th>Workout+ tweets (n=163,242)</th>
<th>Minutes per tweet, mean (SD)</th>
<th>Total Minutes</th>
<th>Tweets (% total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: ≤11.92</td>
<td>143,552</td>
<td>44.05 (97.26)</td>
<td>6,320,924</td>
<td>178,756</td>
<td>1,745,722</td>
<td>48.12 (128.83)</td>
<td>36,279</td>
<td>241,254</td>
</tr>
<tr>
<td>2: &gt;11.93 and ≤20.50</td>
<td>118,047</td>
<td>43.42 (65.54)</td>
<td>5,125,345</td>
<td>154,087</td>
<td>1,666,997</td>
<td>45.53 (91.67)</td>
<td>36,622</td>
<td>77,824</td>
</tr>
<tr>
<td>3: &gt;20.51 and ≤35.65</td>
<td>88,182</td>
<td>49.32 (324.43)</td>
<td>4,348,112</td>
<td>125,096</td>
<td>1,694,811</td>
<td>45.91 (104.47)</td>
<td>36,914</td>
<td>51,702</td>
</tr>
<tr>
<td>4: &gt;35.65</td>
<td>70,229</td>
<td>41.26 (54.97)</td>
<td>2,897,436</td>
<td>123,656</td>
<td>2,550,963</td>
<td>47.75 (285.42)</td>
<td>53,427</td>
<td>70,229</td>
</tr>
</tbody>
</table>

Table 4. Minutes exercised by gender and Klout Score among workout group.

<table>
<thead>
<tr>
<th>Quartile and Klout Score</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tweets (% total males)</td>
<td>Minutes per tweet, mean (SD)</td>
</tr>
<tr>
<td><strong>Workout</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1: ≤11.92</td>
<td>241,254</td>
<td>10,935,339</td>
</tr>
<tr>
<td>2: &gt;11.93 and ≤20.50</td>
<td>81,503 (33.78)</td>
<td>3,528,992</td>
</tr>
<tr>
<td>3: &gt;20.51 and ≤35.65</td>
<td>67,666 (28.05)</td>
<td>2,942,049</td>
</tr>
<tr>
<td>4: &gt;35.65</td>
<td>51,863 (21.50)</td>
<td>2,811,512</td>
</tr>
<tr>
<td><strong>Workout+</strong></td>
<td>40,222 (16.67)</td>
<td>1,652,786</td>
</tr>
</tbody>
</table>

**Workout+**

1: ≤11.92                | 20,504 (21.62) | 952,567 | 46.46 (114.94) | 15,775 (23.07) | 793,154 | 50.28 (144.89) |
| 2: >11.93 and ≤20.50    | 22,156 (23.36) | 1,002,024 | 45.24 (85.01) | 14,466 (21.15) | 664,973 | 45.97 (101.02) |
| 3: >20.51 and ≤35.65    | 21,531 (22.70) | 983,395 | 45.67 (112.10) | 15,383 (22.49) | 711,416 | 46.25 (98.06) |
| 4: >35.65               | 30,664 (32.33) | 1,499,587 | 48.90 (362.80) | 22,763 (33.29) | 1,051,375 | 46.19 (117.88) |

**a** There was no significant difference between males and females in the number of tweets for workouts ($P=.64$).

**b** There was no significant difference between males and females in the number of tweets for workout+ ($P=.55$).
Table 5. Total number of tweets by sentiment and app.

<table>
<thead>
<tr>
<th>Tweets and sentiment</th>
<th>Total</th>
<th>DailyMile</th>
<th>Endomondo</th>
<th>Nike+</th>
<th>RunKeeper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of tweets, n</td>
<td>23,391</td>
<td>9298</td>
<td>820</td>
<td>3999</td>
<td>9284</td>
</tr>
<tr>
<td>Positive sentiment, n (%)</td>
<td>9389 (40.14)</td>
<td>7097 (76.41)</td>
<td>211 (25.73)</td>
<td>418 (10.45)</td>
<td>1663 (17.91)</td>
</tr>
<tr>
<td>Negative sentiment, n (%)</td>
<td>2342 (10.01)</td>
<td>1392 (14.99)</td>
<td>51 (6.22)</td>
<td>350 (8.75)</td>
<td>549 (5.91)</td>
</tr>
<tr>
<td>Neutral sentiment, n (%)</td>
<td>11,660 (49.85)</td>
<td>799 (8.60)</td>
<td>558 (68.05)</td>
<td>3231 (80.80)</td>
<td>7072 (76.17)</td>
</tr>
</tbody>
</table>

Figure 2. Word clouds by mobile fitness app.

Discussion

Analysis 1: Measuring Online Influence

This study further explored a novel approach to classify fitness tweets through Klout influence score. The study further stratified by gender through the use of a validated government database, which was probability matched to our data using exact matching procedures. This gender validation allowed for additional analysis of the gender breakdown of the existing dataset. The data were filtered through the matching criteria twice to improve precision, resulting in a 97% gender match. Although we gender matched twice, the process used to gender match could still be missing a few names that appear more often today than they did even a few years ago. Because popular names can change with high frequency, some gender matching in this study may not be valid within several years.

Based on the current database of collected fitness tweets from five mobile fitness apps, the highest Klout quartile included those individuals with a Klout Score of 35.65 or greater. Klout Scores can reach 100; therefore, our highest score tier may not capture an accurate representation of the most influential people on the Twitter platform. Additional insights from this research are described subsequently.

Men Share Their Physical Activity From Mobile Fitness Apps Via Twitter More Often Than Women

Based on this research, men share their workouts using Twitter and mobile fitness apps more often than women (54.35%, 336,109/618,458 vs 45.65%, 282,349/618,458). Although we believe this to be the first gender analysis of the sharing of physical activity from mobile fitness apps using Twitter, previous research on the overall gender use of Twitter suggests that more women than men use Twitter [28], with some...
nonacademic research suggesting that 40 million more women use Twitter on a monthly basis, that 62% of Twitter users are women [29], and those with a higher Klout Score tend to be women. Additional research regarding gender suggests that women are likely to be more active on Twitter as opposed to men, with women tweeting once every 20 hours versus men tweeting once every 26 hours [30].

However, additional research into our dataset using third-party software called Demographics Pro suggests that the average mobile fitness app user in the fitness tweet dataset is a male in his early thirties, typically married with children and having a high income. Additional insights into the users of mobile fitness apps who also tweet their physical activity includes that this group’s most common professions are programmers, photographers, church leaders, designers, and teachers. The group has a notably high concentration of Web developers (within the top 10% of overall Twitter distribution in this respect). In their spare time, they particularly enjoy beer, political news, wine, comedy/humor, and cooking. People in this group are charitably generous and particularly health conscious. Sports that rise most notably above the Twitter norm include cycling, skiing, and golf. As a consumer, this group is relatively affluent, with spending focused most strongly on technology, wine/dining, and health/fitness. Their strongest brand affiliations include Apple Store, Trader Joe’s, CrossFit, Trek Bicycle, and MyFitnessPal.

The Design of the Mobile Fitness App and the Sharing of Physical Activity Data to Social Networking Sites

The sharing of workout+ tweets is dramatically enhanced by the user interface of the mobile fitness app. When comparing the four mobile fitness apps for the total number of activity tweets (workout tweets plus workout+ tweets), the most popular mobile fitness app was Endomondo (211,240 tweets), followed by NikePlus (203,991 tweets), DailyMile (183,732 tweets), and MyFitnessPal (70,723 tweets). The same usage ranking order was seen with men and women (men: 123,482 for Endomondo, 116,388 for NikePlus, 106,846 for DailyMile, and 70,723 for MyFitnessPal; women: 87,758 for Endomondo, 87,603 for NikePlus, 76,886 for DailyMile, and 30,233 for MyFitnessPal). However, there was a large difference when reviewing the workout+ tweets with 97.67% (173,790/177,943) of all workout+ tweets from DailyMile, 1.89% (3358/177,943) from NikePlus, 0.44% (776/177,943) from Endomondo, and no workout+ tweets from MyFitnessPal. In reviewing the user interface for all four mobile fitness apps, it is evident that the design of DailyMile made it much easier to share not only the workout, but also additional information about the workout when compared to the other three mobile fitness apps. Also during the evaluation time period for the activity tweets, Endomondo used a third-party service called @addthis to share workout+ tweets. With no workout+ tweets from MyFitnessPal, we determined that the app made a design decision to not allow users to share additional information regarding their physical activity workouts.

There is Brand Loyalty Regarding Mobile Fitness App Usage and the Sharing of Physical Activity Data Using Twitter

Of the 113,340 overall users in the dataset, 97.21% (110,186 users) tweeted their physical activity from just one mobile fitness app, 3105 (2.74%) used two mobile fitness apps, with 101 (0.09%) users sharing from three mobile fitness apps and just one user (0.0009%) sharing from four mobile fitness apps. We base this on the analysis of tweets per users and cannot determine the actual usage of the app, only the sharing of physical activity data from the apps. We surmise one reason that more than 97% used just one app could be loyalty, but other reasons such as poor user interface and difficulty in connecting one’s Twitter account to the mobile fitness app may account for other reasons.

Analysis 2: Sentiment Analysis

A better understanding of the online influence of those who are sharing their fitness tweets may lead to new and innovative ways to encourage their followers to be more physically active through peer-to-peer influence, similar to programs created by marketing agencies to influence consumer behavior. Analogous to the other health-related research, physical activity researchers can monitor and attempt to influence physical activity Twitter chatter sent by influential Twitter users who are physically active and popular among various demographic groups and age ranges [15]. The findings can be used to inform online and offline efforts that work to target individuals who are most at risk for the harms associated with a lack of physical activity.

The relatively high number of neutral tweets was expected because each of the mobile fitness apps had a predetermined structure that limited additional information that could be included by the user. There also is the fact that a majority of the tweets simply did not contain words or phrases that could be classified as either positive or negative. Additional insights from this research are described subsequently.

The Real-Time Shared Sentiment of the Physical Activity Can Provide Additional Insights to Physical Activity

We believe that the sharing of one’s physical activity with additional commentary (for the purposes of this research called workout+ tweets) from mobile fitness apps can provide researchers with new insights that in the past may have been difficult to measure. The design of many of the mobile fitness apps allows for the user to share characteristics such as who they were with, the type of weather, the location of the physical activity, and their immediate thoughts regarding the physical activity. These and other insights will allow physical activity researchers to have a greater understanding into the real-time reasons, thoughts, and sentiment of how and perhaps why a person partakes in physical activity. These data will enable a greater understanding surrounding the complexities of physical activity, which can then be used for an enhanced design of mobile fitness apps as a potential tool in the decrease of physical inactivity.
Most Shared Mobile Fitness App Physical Activity Is of a Structured Exercise Type

It is through the analysis and interpretation that the context of fitness tweeting from within mobile fitness apps provides insights into what is being shared, by whom, and for what reasons. Based on the type of information collected, it can be expected that a majority of the activities shared using mobile fitness apps through Twitter were of a more structured exercise type, as opposed to continuous monitoring of daily physical activity. This is possibly due to the additional battery drain on the mobile phone of the user, which would preclude daylong usage of the app. In addition, the structure of the tweets would also suggest that these activities were measured in terms of duration, suggesting activities such as a run, walk, bike, or traditional workout. Because of the nature of some of the activity tweets, it was possible to extract additional information, including the actual type, distance, and the amount of time spent on an activity. It was possible for outliers to be present within the database. For example, the first use of a mobile fitness app could be the user testing the mobile fitness app that may have prompted an activity tweet with a very short-duration activity (seconds rather than minutes), whereas very long-duration activities were sometimes recorded for activities when the person did not properly end his or her mobile fitness app activity session. It was possible that some of the longer-duration activities were, in fact, long exercise sessions. For example, a person training for a marathon would track long runs.

A Significant Majority of Users From Each App Used the App More Than Once

Based on the research data, the number of one-time users of a mobile fitness app that shared their activity via Twitter (activity tweets) was calculated. Although the research cannot determine if a person continued to use a mobile fitness app and decided not to share via Twitter, it was determined that of all users, between 17% and 27% used the sharing to Twitter feature only once depending on the app. A number of reasons could exist for one-time use, including user error, experimentation of sharing functionality, or testing by a user choosing a mobile fitness app. From the 165,768 users that posted activity using a mobile fitness app that was then shared via Twitter, the database included 76,192,059 minutes of activity over the 6-month time period. These minutes are equivalent to 52,911 days, 1738 months, or more than 145 years of combined activity. We cannot determine if this physical activity was the only performed physical activity by each user during the time period because it is understood that users may have completed physical activity without using their mobile fitness app.

These findings and interpretations should be regarded as exploratory and speculative because they represent what can be potentially done in a short development time and with ease of use for non-computer programming health-promotion researchers.

Limitations

There are a number of limitations to this research study. Utilizing outside data, in this case the US government, to determine each user’s gender leaves room for error.

This research was conducted using the Twitter firehose, which allows for the collection of all publicly available tweets. Although we are confident in this data-collection process, there is no way to verify it without a financial expense to purchase all tweets. There also remains a challenge in the extraction of useful data from these repositories through data mining and knowledge discovery [6] due to a rapidly evolving explosion of data services and tools that can be used for analysis. This is due in large part to commercial pressures and the potential for using social networking data for computational research [31]. To minimize this limitation, we were able to link different datasets using the user’s Twitter name as the unique identifier through free publicly available data. Future work could enhance our model by purchasing commercially available datasets for analysis.

There has been a steady growth of social media usage, from 5% of the US population in 2005 to close to 70% in 2015. As more Americans have adopted social media, the user base has also grown more representative of the broader population; however, it is still most used by younger age groups [32].

Comparison With Prior Work

The use of social media and emerging technologies to study physical activity and the possible lack thereof continues to increase with the development of such technologies. Previous research has shown an interest in specific characteristics of the social environments adversely affecting health outcomes [3]. Other research has studied the use of wearables and other smart devices to quantify various different health conditions with the self-reported data being shared on social networks, such as Facebook and Twitter [9], and have suggested that the adoption of such emerging technology to monitor physical activity has created new research opportunities to observe, quantify, and define physical activity in the real-world setting [2]. Our research continues to build on these previous studies by providing researchers with other options for data collection and different objectives to consider.

Previous work regarding the role of technology on physical activity through social media includes a dearth of studies that have studied various aspects of the impact of social media on physical activity. Some research has focused on the behavior change challenges that include self-monitoring, goal setting, and problem-solving strategies [33]. Other research has suggested a change in how we think about physical activity and sedentary behavior measurement, a research topic that includes the use of mobile fitness apps and social networks that can collect large amounts of real-time data that previously would have been difficult to collect [34]. Research by Tsoh [35] explores contextual and psychological factors that may underlie the observed low physical activity levels among mobile fitness app users. Our research is more closely related to that of Grundy et al [36] on the network analysis of prominent health and fitness apps and work by Haddadi et al [37] on the integration of shared health and fitness data from mobile fitness apps that are shared over social networks. Although these works are highly relevant to the research presented in this paper, we expand the research by carrying out data analysis including gender and online influence.
Similar approaches to inferring gender include works using a gender-based dictionary [38], through profile picture and background inference [39], and a third-party Web service that can often reveal gender through proprietary algorithms [40]. Specific research on using social media networks and physical activity include work by Althoff et al [41] on the influence of Pokemon Go, the tweeting of physical activity as a possible method to increase physical activity by Tsoh [35], and work by Liu and Young [42] on using social media data analysis for physical activity surveillance.

Future Work

We created a very powerful tool for conducting large-scale research by collecting physical activity data from Twitter, but the demographics used in this research could suggest a bias regarding the breakdown of mobile fitness app users and thus underrepresent certain groups. If researchers wish to use Twitter and mobile fitness apps for physical activity research, additional steps would need to be taken to ensure that all groups are represented in the data samples collected. Apart from technical limitations, there could be ethical challenges that are equally as challenging. Although tweets are considered public, they may contain information that many would consider “private” due to the possible misconception of the perceived audience (a user’s Twitter followers) versus the actual audience (data researchers) [9]. To expand on this work, additional investigation could address possible trends specific to forms of physical activity per gender that could constitute a higher Klout Score. The popularity of consumer-facing health wearables (eg, Fitbit, Garmin) that also share physical activity data with online social networks would be a topic worthy of future research. By using these tracking devices, which monitor physical activity on an ongoing basis, a more inclusive picture of daylong physical activity can be achieved. This is in contrast to mobile fitness app data, which is typically collected and shared following a traditional “workout” (eg, a walk, run, bike). The same data collection and classification model presented in this paper can be used with minimal changes. With regards to online influence, other work could use an alternate measure of online influence rather than Klout.

Conclusion

This research analyzed publicly shared physical activity data collected via Twitter from five different mobile fitness apps. From this dataset, two analyses on the data were conducted to highlight the unique ability to use this type of data within the study of physical activity. The first analysis categorized the users into four quartiles that represented their online influence as calculated by Klout as well as a method to assign gender to each Twitter user. The analysis suggests that men share their workout tweets more than women, that there is more basic sharing of physical activity data (workout tweets) when compared to tweets that also contain commentary by the user (workout+ tweets), and that there is no significant difference in the tweeting of men and women. The second analysis was conducted with workout+ tweets and showed, across all apps, most of the shared tweets were neutral, but for those with a sentiment there were four times as many positive tweets as negative.

Acknowledgments

This project was supported by Science Foundation Ireland (under grant numbers SFI/08/CE/11380 and SFI/12/RC/2289) and the Irish Research Council.

Conflicts of Interest

None declared.

References


©Theodore Vickey, John G. Breslin. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 31.10.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.
Twitter classification model: the ABC of two million fitness tweets

Theodore A. Vickey, Kathleen Martin Ginis & Maciej Dabrowski

Translational Behavioral Medicine
Practice, Policy, Research

ISSN 1869-6716

DOI 10.1007/s13142-013-0209-0
Your article is protected by copyright and all rights are held exclusively by Society of Behavioral Medicine. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer’s website. The link must be accompanied by the following text: “The final publication is available at link.springer.com”.

Springer

236
Twitter classification model: the ABC of two million fitness tweets

Theodore A. Vickey,1 Kathleen Martin Ginis, PhD,2 Maciej Dabrowski, PhD1

ABSTRACT
The purpose of this project was to design and test data collection and management tools that can be used to study the use of mobile fitness applications and social networking within the context of physical activity. This project was conducted over a 6-month period and involved collecting publicly shared Twitter data from five mobile fitness apps (Nike+, RunKeeper, MyFitnessPal, Endomondo, and dailymile). During that time, over 2.8 million tweets were collected, processed, and categorized using an online tweet collection application and a customized JavaScript. Using the grounded theory, a classification model was developed to categorize and understand the types of information being shared by application users. Our data show that by tracking mobile fitness app hashtags, a wealth of information can be gathered to include but not limited to daily use patterns, exercise frequency, location-based workouts, and overall workout sentiment.

KEYWORDS
Physical activity, Twitter, Mobile fitness apps, Online social network

INTRODUCTION
Technology, health, and physical activity
As much as technology has enriched society and expanded global communication, it can be argued that it has also negatively affected overall global health by lowering opportunities for physical activity [1] and by contributing to an overall secular decline in physical activity participation rates [2]. At the same time, research also indicates that there is a potential for technologies to be used as a means for improving health and increasing physical activity [3].

According to a report issued by mobihealthNews, more than 13,000 health and fitness apps were available via iTunes by August 2012 [4]. The use of smartphones in supporting health behavior change via mobile fitness apps is encouraging. Aside from expanded opportunities for users to access health information, mobile devices are becoming more persuasive behavior change tools, allowing for the facilitation of ongoing collection of personal data and the opportune timing of feedback and education to elicit a change in behavior [5]. The most recent health applications have been smartphone applications for personal health areas such as diabetes care, nutrition tracking, smoking cessation, and fitness [4].

The recent advent of smartphones has greatly enhanced both the reach and realm of mobile apps for health purposes by providing a platform for developers to design third-party applications (apps), which expand the functionality and utility of mobile devices [6]. These applications allow users to track their fitness activities via GPS from their smartphones. They also allow the immediate sharing of details about a workout with friends and family that make up one’s online community through a website hosted by the app company or by third-party social networks such as Facebook or Twitter. Indeed, simple mobile devices can function as inexpensive, accessible, and powerful triggers for behavior change and may be a particularly powerful mechanism for delivering social support [1].

Online social networking sites are a relatively new and innovative way to deliver social support for physical activity. Online social networking services have eliminated the four walls of brick and mortar found in traditional networking and social interaction [7] and facilitate the development and maintenance of social contacts. One example of a social network is Twitter. The structure of Twitter is simple—users send messages (a.k.a., tweets) to a network of people (a.k.a., followers) from a variety of devices (desktops, laptops, mobile devices, etc.). Tweets are text-based messages of up to 140 characters in length. The default setting for the sharing of tweets is public, which permits other Twitter users to follow and read each other’s tweets. Each user has a personalized Twitter home page where all their tweets are aggregated into a single list [8].

Implications
Practice: The fitness tweet classification model can be used by researchers to better understand and classify fitness information collected via Twitter.

Policy: A system was developed whereby policy decisions can be made more effectively by the classification of real-time, on-body data collection rather than self-reported measures.

Research: This study provides a research opportunity between health and exercise science and social networking/social software disciplines.
Recent academic research has explored the role of the Twitter hashtag—a short keyword, prefixed with the hash symbol “#”—as a means of collecting a distributed discussion between groups of users, who do not need to be connected through existing “follower” networks [9]. One function of Twitter is the ability for information to be shared not only to those who are part of the follower network but also to the entire Twitter population by default. Twitter co-founder Evan Williams suggests “Twitter lets people know what’s going on about things they care about instantly, as it happens. In the best cases, Twitter makes people smarter and faster and more efficient” [15]. But with over 400 million tweets sent every day [14], individual tweets can be inane; but taken collectively, analysis of a stream of messages can turn Twitter into a useful tool for solving problems, performing research, and providing insights into the digital moods of its users. Twitter hashtags have been studied to garner information on topics such as terrorism informatics, user modeling and personalization, online security, spam detection, and information streaming [10].

In addition, research has focused on how Twitter is used as a communication platform and understanding why and how people use online social networks. By understanding the reasons, improvements to the overall structure of the network can occur [11]. From this work, researchers have derived standard metrics for measuring a user’s Twitter behavior, such as the number of tweets, retweets, and followers, [12] as well as text classification models to help understand the content of each tweet [13]. Retweets are the forwarding of tweets received by one user to their own personal social network, thus allowing for tremendous “virtual” sharing of information. Twitter followers are fellow Twitter users where one user “follows” the tweets of another.

Research on text classification within Twitter has shown that people use Twitter for different reasons. Java et al. [11] identified four main user intentions on Twitter: (1) Daily Chatter—most posts on Twitter talk about daily routine or what people are currently doing and this is the largest and most common user of Twitter; (2) Conversations—about one-eighth of all posts contain a conversation and this form of communication was used by almost 21% of users; (3) Sharing information—about 13% of posts contained a URL (i.e., website address), directing readers to another information source; and (4) Reporting news—many Twitter users report latest news or comment about current events on Twitter. Some automated users or agents post updates like weather reports and new stories from RSS feeds.

Text classification is one of the most important research fields in information retrieval and data mining, and its solutions are at the core of several technology applications ranging from the automatic cataloging of newspaper pages and web pages to the management of incoming e-mails and from the annotation of DNA genome sequences to sentiment analysis of tweets [16]. By tapping into the world’s collective brain, researchers have found that efforts to dig through the millions of individual tweet can provide a glimpse into public sentiment and activity and perhaps can even help shape it [15]. To the best of our knowledge, no research to date has conducted text classification within the context of Twitter and physical activity. An understanding of how Twitter is used in physical activity contexts could lead to improvements in the development of mobile fitness apps that promote and support physical activity behavior change.

Thus, the overarching purpose of this study was to develop an understanding of the types of information being shared from mobile fitness apps via Twitter. Our specific objectives were (1) to develop and implement a method for collecting fitness tweets sent from mobile fitness apps, (2) to develop a conceptual model to classify tweets, and (3) to analyze and interpret a sample of tweets. Given the preliminary nature of this research, no hypotheses were put forth.

METHODS AND RESULTS

Development and implementation of a fitness tweet collection method

After an online review of online tools that could collect and manage tweets, an open source program called TwapperKeeper was chosen. TwapperKeeper is a web application designed to archive social media data via Twitter to allow for long-term archival and analysis. The application uses a Twitter-enabled API that acts as an interface between the Twitter search function and a cloud database for tweet storage. The application allows users to monitor and archive specific hashtags and to provide additional metadata to describe an archive that can later be viewed in multiple.

Once the hashtags were defined, the application began two archiving processes (TwapperKeeper, 2011):

- “The Crawl”—For the keyword defined (by a hashtag), the crawling processes began to poll the Twitter Search API to find all tweets in the search cache that match the desired hashtag. This allowed for TwapperKeeper to fill in older tweets (limited by the Twitter API) as well as continually monitor tweets that might be missed by “The Stream” archive process. A disruption of service is possible during disconnects/reconnects with the Twitter Streaming API, rate limits imposed by Twitter, and possible service interruptions on the Twitter service itself.

- “The Stream”—A persistent connection was also created with the Twitter Streaming API for the desired hashtags. The archiving process inserted all inbound tweets into a database table for later processing. A second process ran to analyze each
TBM page 3 of 8

To finalize the selected mobile fitness apps to study. The five apps chosen were Endomondo, MyFitnessPal, Nike+, RunKeeper, and dailymile.

We then collected tweets from the five mobile fitness apps by gathering tweets that used the following hashtags: #endomondo, #myfitnesspal, #nikeplus, #runkeeper, and #dailymile. These are the hashtags that the apps automatically attach to a tweet to indicate it has come from that particular application. It is through these hashtags that common themes or information can be grouped within Twitter. Tweet collection was done by TwapperKeeper which began the archiving process by searching publicly available tweets, identifying tweets that contained the desired hashtags, and inserting identified tweets into a database for later processing. The type of information collected from each tweet is shown in Table 1.

In addition, supplementary information was collected regarding demographics and Twitter usage. To collect and process such information, a JavaScript was created. The JavaScript extracted Twitter information collected from the TwapperKeeper database and requested specific information about the publicly available Twitter user account. After limiting the process to unique users, the script has the ability to send Twitter user information to websites such as Twitter, Klout, and other information websites to collect general information about the Twitter users such as their start date on Twitter, number of total sent tweets, their Klout score, and frequency of tweets.

The information collected about the Twitter user is shown in Table 2. All information collected was publicly available with each user of Twitter agreeing to this public sharing of information by their agreement to the terms and conditions of their Twitter account.

### Development of a conceptual model to classify fitness tweets

Our second research objective was to develop and validate a strategy to classify the collected tweets. This strategy was based on two fields of exploration within the Twitter research tracks that seek to better understand the context of tweets: data mining and text classification. Data mining is a relatively young and interdisciplinary field of computer science with the process that results in the discovery of new patterns in large datasets by using methods at the intersection of artificial intelligence, machine learning, statistics, and database systems, with the overall goal being to extract knowledge from an existing dataset and transform it into a human-understandable structure for further use [17]. Text classification is the labeling of natural language texts (in this case, a tweet) into one or more categories drawn from a predefined set. This may be done manually or algorithmically. For the purposes of this research, the complete dataset was sorted and evaluated manually to determine any apparent similarities. This evaluation allowed the algorithms to be established that were then used to classify the entire database of fitness tweets. Data verification tests

<table>
<thead>
<tr>
<th>Data point</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archive source</td>
<td>Twitter Search or Twitter Stream</td>
</tr>
<tr>
<td>Text</td>
<td>The actual tweet</td>
</tr>
<tr>
<td>To_User</td>
<td>Name of recipient user if the tweet was sent to a specific Twitter user</td>
</tr>
<tr>
<td>From_User</td>
<td>Name of the Twitter user that sent the tweet</td>
</tr>
<tr>
<td>ID</td>
<td>Specific Twitter identification number for the associated tweet</td>
</tr>
<tr>
<td>From_User_ID</td>
<td>Specific Twitter identification number for the associated Twitter user name that sent tweet</td>
</tr>
<tr>
<td>Iso_Language_Code</td>
<td>Identified language of the tweet</td>
</tr>
<tr>
<td>Source</td>
<td>Twitter platform used to send tweet</td>
</tr>
<tr>
<td>Profile_Img_URL</td>
<td>URL to the picture of the tweeter</td>
</tr>
<tr>
<td>Geo_Type</td>
<td>Either “point” if geolocation was used with tweet or blank if not</td>
</tr>
<tr>
<td>Geo_Coordinates_0</td>
<td>Latitude of the location where the tweet was sent</td>
</tr>
<tr>
<td>Geo_Coordinates_1</td>
<td>Longitude of the location where the tweet was sent</td>
</tr>
<tr>
<td>Created_At</td>
<td>Day, date, and time the tweet was sent</td>
</tr>
<tr>
<td>Time</td>
<td>UNIX time the tweet was sent</td>
</tr>
</tbody>
</table>
The development of the fitness tweet classification model was based on available macro topic classification models where tweets were categories into broad categories of content, based on prior literature [18], and sourced from other works [8,11,13,19]. These prior literatures indicated four major categories for sharing on Twitter—Conversational, Pass-Along, News, and Status—categories that are consistent with Java’s [11] research on the primary purposes of tweets. These four categories provided the starting point for the development of our framework. Once the theoretical foundation was established, a custom computer program was created that incorporated data mining and text classification of the collected tweets.

Consistent with previous research [13], we used a grounded theory approach to develop the framework. The grounded theory research approach is opposite to other traditional social science research where the researcher chooses a theoretical framework and then applies the model to the phenomenon to be studied. Rather than beginning with a hypothesis, the grounded theory starts with data collection. From the collected data, key points are marked with a series of codes, which are extracted from the text. These codes are then grouped into similar concepts, making them more workable. From these concepts, categories are formed, which are the basis for the creation of a theory or a reverse-engineered hypothesis [20].

Development of the framework began by taking a random sample of 500 public tweets (100 from each of the five apps), tweeted over a 2-week period. The researchers sorted the tweets into groups with two general themes emerging: tweets about a recent workout (i.e., “Activity”) and tweets about other non-exercise-related conversational topics (i.e., “Conversation”). Tweets that share a person’s workout, specific to the tweet structure as defined by the five different mobile fitness apps, were classified as Activity. Each mobile fitness app used a different data structure that was able to be defined. Table 2 provides examples within the Activity category. These 500 tweets were then submitted to a computerized text classification procedure programmed to identify Activity and Conversation tweets. This procedure revealed subcategories within the Activity and Conversation groupings as well as a third category, subsequently labeled “Blarney.”

Specifically, further analysis of the Activity tweets showed that some users added additional messages along with the information about their actual workout (e.g., I just ran 4 mi using #RunKeeper in the sunshine of San Diego, felt great); thus, the “Workout Plus” subcategory was added. A Workout Plus tweet has the same foundation of a Workout tweet but adds the additional variable of information.

Further analysis of the Conversation category indicated tweets pertaining to four areas: requests for technical support (requests to the app company or the broader community), marketing (e.g., press releases and updates that came from the app company itself or the community), statements of support (where people within the app community congratulated others on reaching milestones, personal bests, etc.), or information sharing (e.g., those within the app community that wanted to run together in an upcoming 10-km race would post messages using the hashtag per the app). Thus, the following subcategories were added to the Conversation category: Technical Support, Marketing, Statements of Support, and Information Sharing. In addition, a new third category was added (Blarney) that tagged spam tweets (tweets with only a URL) or tweets that had little relevance to exercise (e.g., Test FB http://t.co/IKIQjTi #myfitnesspal). Blarney is defined as skillful flattery, nonsense, or blandishment [22].

Table 2 | Additional Twitter data point descriptions from the Fitness Tweet Crawler

<table>
<thead>
<tr>
<th>Data point</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>User_Name</td>
<td>Twitter user name of the person that sent the tweet (same user name as the From_User listed in TwapperKeeper)</td>
</tr>
<tr>
<td>Location</td>
<td>Location of the tweeter as recorded in their Twitter user profile</td>
</tr>
<tr>
<td>Tweets</td>
<td>Number of total tweets sent by user at the time of the query</td>
</tr>
<tr>
<td>Following</td>
<td>Number of people the user is following at the time of the query</td>
</tr>
<tr>
<td>Followers</td>
<td>Number of people that follow the user</td>
</tr>
<tr>
<td>Klout</td>
<td>Klout score of the user</td>
</tr>
<tr>
<td>Style</td>
<td>Klout style classification of the user</td>
</tr>
<tr>
<td>Access_Date</td>
<td>Date of the query</td>
</tr>
<tr>
<td>Access_Time</td>
<td>Time of the query</td>
</tr>
<tr>
<td>Twitter_Startdate</td>
<td>Date the user started using Twitter</td>
</tr>
<tr>
<td>Times_Per_Day</td>
<td>Number of times per day the user sends any tweet</td>
</tr>
</tbody>
</table>
To examine the strength of the framework, six coders were given the same 500 tweets previously used and were asked to classify them according to the framework. Examples of messages placed into each category are shown in Table 3. Agreement among raters was high. An intraclass correlation coefficient, using the two-way mixed-effects model, yielded an ICC = .925 (95% confidence interval = .914–.935). With the general recommendation for reliability being .7 [21], reliability among the raters was high, providing strong evidence that the framework could be reliably used to classify fitness tweets. This framework was then used to generate computer code for computerized classification of the tweets, as reported in the next section.

Analysis and interpretation of a sample of tweets

The third objective was to provide an analysis of tweets using our classification framework. Data collection using TwapperKeeper began on Thursday April 21, 2011 at 00:00 Greenwich mean time and continued until September 21, 2011 at 23:59 for a total collection of Twitter data of 184 days. During this period, 2,856,534 tweets were collected in 23 different languages. However, for the purposes of this study, only the English tweets were analyzed. After reviewing the human classification of the tweets, minor adjustments to the code enhanced the overall reliability of the computer classification of the tweets.

The total number of processed tweets in English was 1,982,653, which were tweeted by 165,768 unique users. Figure 2 displays the breakdown of the fitness tweets categorization using the fitness tweet classification model. Of the English language tweets, 1,446,462 (73%) were classified as Activity, 104,360 (5%) as Blarney, and 420,603 (21%) as Conversation. Of additional interest is the subclassification breakdown. Figure 3 displays the breakdown of fitness tweets in the subclassification. Of the Activity tweets, 53% of the total tweets were Workout, with 21% as Workout Plus. There was a small sample of Blarney tweets with 1% Pointless Babble and 5.2% Spam tweets from the total dataset of tweets. Of Conversation tweets, 4% was of Technical Support, 5% was of Corporate Marketing, 1.3% was of Statements of Support, and 19% was of Information Sharing relative to the total number of tweets in the dataset. Of the Activity fitness tweets, over 76,192,059 min of exercise was shared via the five mobile apps via Twitter equaling over 145 years of physical activity.

DISCUSSION

Using the data collection and data processing tools described in this paper, we have been able to create a growing dataset of information that people publically share from their smartphones and other devices, via Twitter, about their workout activities. This information includes data collected by the app itself—such as exercise type, length, day of the week, mood, geographical location, and time—as well as data on how people use fitness apps to share information and engage in social networking regarding their fitness activities. Together, this information can facilitate research on how technology can be used to monitor and motivate physical activity and how online social networks may play a role in physical activity promotion and adherence.

We have created a Twitter classification model that allows for analysis of mobile fitness app tweets through data collection, data processing, and data
<table>
<thead>
<tr>
<th>Activity</th>
<th>dailymile</th>
<th>Endomondo</th>
<th>RunKeeper</th>
<th>Nike+</th>
<th>MyFitnessPal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workout</td>
<td>Ran <a href="http://dmile.com/e/Ok7N">http://dmile.com/e/Ok7N</a></td>
<td>Was out cycling 0.32 km with #Endomondo. See it here: <a href="http://bit.ly/hNKf3y">http://bit.ly/hNKf3y</a></td>
<td>Just completed a 5.69 km walk with @runkeeper. Check it out <a href="http://mkpr.com/ahr0yw">http://mkpr.com/ahr0yw</a> #RunKeeper</td>
<td>I just finished a 2.00 mi run with a time of 20:05 with Nike+GPS. #nikeplus</td>
<td>burned 157 calories doing 60 minute of “Yoga” #myfitnesspal</td>
</tr>
<tr>
<td>Workout Plus</td>
<td>Did a cross training workout for 45 mins and felt good. Morning upper: 2x15 L-pullups +25 bicep curl <a href="http://dmile.com/e/Ols1">http://dmile.com/e/Ols1</a></td>
<td>Was out running 4.4 miles with #Endomondo. Had PB 2nd mile. See it here: <a href="http://bit.ly/dRLrgn">http://bit.ly/dRLrgn</a></td>
<td>Just posted a 2.25 mi run - just some miles to stretch the legs before the race tomorrow. <a href="http://mkpr.com/ahm649">http://mkpr.com/ahm649</a> #RunKeeper</td>
<td>9.69 miles this morning. Longest run since the move to Colorado. #nikeplus #fb</td>
<td></td>
</tr>
<tr>
<td>Blamey Pointless Babble</td>
<td>Test #dailymile</td>
<td>Rated on LUUUX <a href="http://bit.ly/ohjo71">http://bit.ly/ohjo71</a></td>
<td>Other Activity 0.00 km</td>
<td>#nikeplus software is crap. Runs far too slow?</td>
<td>Test FB <a href="http://t.co/kIJOjT">http://t.co/kIJOjT</a> #myfitnesspal</td>
</tr>
<tr>
<td>Spam</td>
<td>#dailymile <a href="http://bit.ly/jht2Wn">http://bit.ly/jht2Wn</a></td>
<td>#endomondo</td>
<td>#US91K Congrats! Your RunKeeper Rank has been updated! <a href="http://t.co/ymRjs0b">http://t.co/ymRjs0b</a> #WorldRankin</td>
<td>I just finished not caring w/ a time of 2.5 seconds using #nikeplus</td>
<td>MUST SEE! <a href="http://bit.ly/bhtvDB">http://bit.ly/bhtvDB</a> #myfitnesspal</td>
</tr>
<tr>
<td>Conversation Technical Support</td>
<td>@akeyo we’ll look into that! We do have a way to export all your data.</td>
<td>#Endomondo now supports #Facebook integration for your Facebook apps: <a href="http://bit.ly/k29sYK">http://bit.ly/k29sYK</a></td>
<td>@bsheamcincyny it’s buggy on runkeeper. Like I said, it’s perfect on my iphone4.</td>
<td>@nonthone A good place to start when you experience weird iPod behavior is with a reset</td>
<td>@MyFitnessPal can the BB app scan barcodes?</td>
</tr>
<tr>
<td>Statements of Support</td>
<td>@dailymile I can do it. Only so many miles to go!</td>
<td>:) Yay to moving my lazy ass! RT @ishsal: Was out running 4.04 km with #Endomondo. See it here: <a href="http://bit.ly/og24">http://bit.ly/og24</a></td>
<td>Achieved a new personal record with @RunKeeper: Farthest distance... <a href="http://bit.ly/hy8xUZ">http://bit.ly/hy8xUZ</a> #FitnessAlerts</td>
<td>Good job! RT @Stom21T just finished a 3.51 mi run with a time of 36:29 with Nike+GPS. #nikeplus</td>
<td>#myfitnesspal New Member Struggle with losing weight and keeping it off <a href="http://bit.ly/hx2hb2">http://bit.ly/hx2hb2</a></td>
</tr>
<tr>
<td>Information Sharing</td>
<td>#dailymission Too many of us focus on the training, what about core? I work on core everyday. It’s what I... <a href="http://dailymile.com/e/QhSH">http://dailymile.com/e/QhSH</a></td>
<td>My last Endomondo was walking, not cycling. I forgot to change.</td>
<td>Watch my bike ride right now with @RunKeeper Live <a href="http://mkpr.com/ahstp9">http://mkpr.com/ahstp9</a> #RKLive #RunKeeper</td>
<td>RT @sneakernoize: Reading Material For the Sneaker Fiend Kicksclusive Magazine &amp; <a href="http://bit.ly/gsc42L">http://bit.ly/gsc42L</a> #nikeplus</td>
<td>lost 4 pounds since his last weigh-in! He’s lost 28.8 pounds so far. #myfitnesspal</td>
</tr>
</tbody>
</table>
analysis. We have shown that a simple 140-character Twitter message can lead to a wealth of pertinent and valuable demographic- and action-oriented information that when processed through the fitness tweet classification model can show patterns of associations between user, online fitness communities, and Twitter as a whole. Our research extends the Twitter classification models established by Naaman et al. [13] who created models within the context of how to establish categorization of tweets. Our classification model was created using a grounded theory approach and its reliability was confirmed across six raters who successfully coded 500 sample tweets. Given the breadth of mobile fitness apps included in our analysis, we are confident that the classification model can be applied to categorize data obtained from other mobile fitness apps that have the ability to share information via Twitter. Indeed, an important contribution of this project is the identification of data structures from within mobile fitness applications when sharing via Twitter. These structures can now be analyzed using text classification processes. Given the tremendous amount of data generated by Twitter (e.g., we have collected over 12 million tweets from just five mobile apps over the past 15 months), researchers need tools to manage and analyze these data in order to address research questions regarding the use of technology and social networks to promote health behavior change. Our work has yielded such tools.

We have also provided preliminary data on how people are engaging with their online social communities to share information on their fitness activities. While there is a substantial amount of information being shared via Twitter regarding actual workouts (i.e., Activity tweets), there appears to be only a small amount of conversation about the workout themselves (Conversation tweets). For those users who track their workouts for internal quantified self reasons, our data would indicate that the mobile fitness apps can provide such a tool. However, it is unclear regarding the reason...
why these people would decide to share their workouts on a social networking service such as Twitter. The lack of meaningful and engaged conversation between mobile fitness app users is an important question to be addressed for future research. If it is the intent of mobile fitness app developers who have the objective of using these apps to increase physical activity behavior by way of strengthening social support via one’s social network, then having a true understanding of why conversation is not occurring is critical. However, our findings would indicate similar usage patterns for general Twitter usage. General Twitter research suggests that Twitter is used more as a one-way, one-to-many publishing service than a two-way, peer-to-peer communication network [23].

In summary, this study has provided tools for advancing research on mobile fitness app use, social networking, and physical activity. With these tools available, researchers can now examine a wide range of questions such as how the use of mobile fitness apps and the sharing of workout information using Twitter is related to possible exercise motivation of why conversation is not occurring is critical. Researchers can now examine a wide range of questions such as how the use of mobile fitness apps and the sharing of workout information using Twitter is related to possible exercise motivation.

Fitness—There's an App for That

Review of Mobile Fitness Apps

TED VICKEY, JOHN BRESLIN, AND ANTONIO WILLIAMS
Fitness—There's an App for That: Review of Mobile Fitness Apps

Ted Vickey, National University of Ireland, Ireland
John Breslin, National University of Ireland, Ireland
Antonio Williams, Indiana University, USA

Abstract: The purpose of this study was to research the new emerging technology of mobile health, the use of mobile fitness apps to share one’s workout with their Twitter social network, the workout tweets and the individuality of the Tweeters. A comparison of five popular mobile fitness apps include Nike+, RunKeeper, DailyMile, MyFitnessPal and Endomondo.

Keywords: Physical Activity, Twitter, Endomondo, RunKeeper, Nike+, DailyMile, MyFitnessPal, Online Social Network

Introduction

As much as technology has enriched society and expanded global communication, it can be argued that it has also negatively affected overall global health by lowering opportunities for physical activity and contributing to an overall decline in physical activity participation rates (Foster, Linehan, & Kirman, 2010). Perhaps it is time to find new and innovative ways to not only promote physical activity, but also to motivate people in a more persuasive manner. One such way is through mobile technology.

There is a substantial body of research regarding social networking and increased physical activity, but little has been studied regarding the effective use of advanced internet technologies, such as mobile technologies, that can be used as a means for generating positive change by functioning as inexpensive, accessible and powerful “just in time” triggers toward behavioral change. The prevalence of mobile smartphones being always on, always near and always connected means they can become social-support tools without the need for overly complex and expensive applications or devices (Foster et al., 2010). Mobile technology can be used as a tool in the fight against physical inactivity.

In September 2011, 17% of mobile phone users reported using their mobile phones to search for health information. That number increased to 31% September 2012 (Matthews, 2013). During that same year, the number of global users of mobile health apps, of which mobile fitness apps is a subset, doubled from 124M to 247M (Matthews, 2013). With annual increases in both the ownership of smartphones and the growing use of these smartphones to access health information from the web, mobile fitness apps are poised to help transform the healthcare system in the foreseeable future. It is the convergence of the social network and mobile technology that is needed to combat the lack of physical activity.

The most recent health applications have been smartphone applications for a number of personal health areas including but not limited to diabetes care, nutrition tracking, smoking cessation and fitness (Dolan, 2012). The recent advent of smartphones has greatly enlarged both the reach and realm of mobile apps for health purposes by providing a platform for developers to design third-party applications (apps), which expand the functionality and utility of these mobile devices (West et al., 2012).

One example of emerging technology within the mobile health space that could help increase physical activity is the Mobile Fitness App (MFA). These applications allow a user to track their fitness activities from their mobile phone; and to then share the workout with their social network either via the application’s website, Facebook or Twitter. According to a report issued by MobiHealthNews, more than 13,000 health and fitness apps were available via iTunes in August 2012 (Dolan, 2012). The use of smartphones in supporting health behavior change via mobile
fitness apps is encouraging. Aside from expanded opportunities for users to access health information, mobile devices are becoming more persuasive behavior change tools, allowing for the facilitation of on-going collection of personal data and the opportune timing of feedback and education to elicit a change in behavior (Patrick, Intille, & Zabinski, 2005). In addition to allowing a user to track their fitness activities via a global positioning satellite (GPS) from their smartphone, these applications also allow the immediate sharing of a workout with friends and family that make up one’s online community through a website hosted by the app company or by third party social networks such as Facebook or Twitter.

Review of Mobile Fitness Apps

This research reviewed five mobile fitness apps to establish a baseline of functionality for this type of mobile application. Once the baseline was established, a criterion was created to allow for multi-application comparison with regards to the sharing of a user’s workout and additional fitness related information. Of the thousands of mobile fitness apps available, five were chosen for analysis based on their availability via an iPhone, the ability of the mobile fitness app to share workout information through Twitter, and apps that represented larger versus smaller corporations, based in U.S. versus abroad, and targeting beginner versus experienced exercisers. Criteria for the selection of the types of mobile fitness applications to be included in this research were established as follows:

1. Operating System – iPhone. At the time of evaluation, the iTunes store and associated health applications had significant market share over other platforms (Blackberry, Android, Windows).

2. Social Sharing – Twitter. Each mobile fitness app that was to be included in the model needed to have the ability to share workouts and associated information via Twitter. While sharing of information occurred via other methods (i.e. on the application’s website or Facebook), researchers were unable to query these postings. Since tweets were publicly available, had character limitations and standard sharing options, Twitter was chosen.

3. Diversity – In order to classify as many fitness tweets as possible within a limited timeframe, diversity of design was included as the last criteria for selection. To encompass a wide range of mobile fitness apps, the following categories were established (with chosen MFA identified):
   - One large well known / public company app (Nike+),
   - One non-USA based app (Endomondo),
   - One running/walking app (RunKeeper),
   - One community-based app (DailyMile),
   - One general health app that included both fitness and nutrition (MyFitnessPal).

Endomondo

Described as a personal athletics tracker, Endomondo is a free mobile/GPS-powered Sports Tracker app which runs on multiple platforms, including iPhone, Android and Garmin watches (Endomondo, 2012). Endomondo can be used for distance-based activities and utilises both GPS and Microsoft Bing interactive maps to track routes, distance, duration, split times and calorie consumption while providing audio feedback on performance. By incorporating aspects found in leading social networks, Endomondo allows for user interaction in order to facilitate motivation to both become active and stay active. This is specifically accomplished in that users can send real-time pep talks to friends while they are exercising, compete against friends, challenge co-workers and share it all on Facebook or Twitter (StreetInsider.com, 2012).
Christian Birk, co-founder of Endomondo stated "Our vision is to make fitness fun. To succeed in this goal, Endomondo’s service takes full advantage of Microsoft’s Health Vault platform. We’ve worked hard on this app so recreational athletes across the world will like it." (StreetInsider.com, 2012). As of October 2010, the mobile fitness app had more than one million downloads with 500,000 registered users. The application has had growth from 40,000 registered users in January 2010 to 100,000 in April 2010, a doubling of its user base every 10 weeks over the past year (O’Hear, 2010).

The alpha version of the application was released in September 2008 in connection with the world’s largest running race, the DHL race with 100,000 participants in Copenhagen, Denmark. Enhancements to the application from the feedback provided by the initial users allowed for a beta version of Endomondo.com launch in July 2009 (Endomondo, 2012). As one of the more popular mobile fitness apps, Endomondo - through their online social fitness network - centres on sports that unify active individuals and promotes exercise for both recreational and serious athletes. In 2012, Endomondo reported that the service had over 7 million worldwide users and was tracking up to 200,000 workouts per day with a 28% increase in exercise activities for those engaged in the community throughout Europe (StreetInsider.com, 2012).
In addition to the basic tracking of a workout route, split times, calorie consumption and challenges, Endomondo provides the user with an audio coach. For each mile or kilometre, a voice will inform the user about distance and speed. In addition, the app enables friends to follow the user’s run in real-time from their PC, from which they can send messages of encouragement that are converted to audio and played during the workout (Endomondo, 2012). In the first quarter of 2012, Endomondo was named a winner of the Microsoft Health Users Group 2012 Innovation Award (StreetInsider.com, 2012).
RunKeeper

RunKeeper suggests that their mobile fitness app “makes tracking your workouts fun, social, and easy to understand so that you can improve the quality of your fitness” (RunKeeper, 2012). As explained in a PC World magazine article, “the RunKeeper app employs your smartphone’s GPS radio to track the distance, time, pace, route, and elevation of your jogs. You can then sync your data to the RunKeeper Website and later view a history of your activity.” (Chiodo, Hopkins, & Mies, 2010).

![RunKeeper Screenshots](web)

RunKeeper is an app developed by Boston based company FitnessKeeper and Founder/CEO Jason Jacobs. Originally designed by Jacobs to help train for his own marathon, the app has expanded to allow for additional fitness tracking including walking, biking, hiking, skiing, rowing, and user-defined fitness activities. As of June 2011, RunKeeper has an online community of 6 million fitness enthusiasts and in addition to the iPhone app, RunKeeper is also available on Android and Windows Phone 7 platforms (Jacobs, 2011).
Jacobs and RunKeeper continue to market the app through high profile partnerships with other technology companies such as location-based Foursquare and promotions such as Jacobs running the Boston Marathon dressed as an iPhone, and had raised $1.51 million in funding as of August of 2010 (McDermid, 2010). In December 2010, RunKeeper changed their business model to allow for the app to be available for free. This type of free promotion strategy is a common, with the goal to increase the total number of downloads which increases the likelihood that the app appears in the Apple top-selling lists of the App Store, where the increased visibility will allow for long term additional downloads (Ha, 2010). As of December 1, 2010, the RunKeeper app was downloaded from the Apple App store over 171,000 times, a ten-fold increase from normal distribution up to that point (Ha, 2010).
Hoping to become the ‘Facebook of Fitness’, in June 2011, FitnessKeeper announced the launch of the Health Graph, allowing developers, fitness sensors and websites to connect into the wealth of health and fitness data collected by the RunKeeper community. Jacobs describes the Health Graph as an extension of the social graph.

“The social graph has evolved into the Open Graph - a system of connections that includes not just personal relationships, but also your personal ‘likes’ and interests. Any website, individual or group that you ‘like’ is eligible for inclusion in your open graph.” (Jacobs, 2012)

The Health Graph was designed to be a system of health connections; an ever-changing digital map of one’s personal health recording body measurement statistics or health related actions that impact personal health. The data sharing established by the Health Graph should provide a snapshot of health, giving users a picture of their health and how it has changed over time.
Table 1: Health Graph Connects Health Activities to Social Interactions over Time (Jacobs, 2012)

<table>
<thead>
<tr>
<th>Data</th>
<th>Week of May 17th</th>
<th>Week of May 24th</th>
<th>Week of June 1st</th>
<th>Week of June 7th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>180lbs</td>
<td>192lbs</td>
<td>176lbs</td>
<td>169lbs</td>
</tr>
<tr>
<td>Avg. daily Calories</td>
<td>3,400</td>
<td>3,500</td>
<td>2,400</td>
<td>2,300</td>
</tr>
<tr>
<td>Avg. daily sleep hrs.</td>
<td>6</td>
<td>5</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Avg. daily activity duration</td>
<td>15 minutes</td>
<td>18 minutes</td>
<td>30 minutes</td>
<td>35 minutes</td>
</tr>
<tr>
<td>Social Network Motivation</td>
<td>Received 2 comments</td>
<td>Received 2 comments</td>
<td>Received 6 comments</td>
<td>Received 10 comments</td>
</tr>
</tbody>
</table>

According to Jacobs (2012), the Health Graph will also provide an understanding of how a person’s activity and behavior correlate with a change in health and how a person’s social interactions influence changes in wellbeing. Since it opened up its Health Graph API, RunKeeper has integrated with over 40 third party services and while data sharing is relatively small, it is doubling every month (Schonfeld, 2011).

In November 2011, RunKeeper raised $10 million in Series B financing led by Spark Capital and AOL founder Steve Case’s Revolution Ventures (Schonfeld, 2011).

Nike+

Nike+ iPod Sports Kit is a set of tools designed by global shoe maker Nike. As suggested by the company, these tools “…allow a user to measure distance, pace, map runs, track progress and get the motivation needed to go even further” (Nike, 2011).

Figure 7: Nike+ Screenshots

The original design of the Nike+ iPod Sports Kit was first introduced in May 2006 and included a small accelerometer that was attached or embedded into a Nike shoe. The accelerometer is designed to communicate with either a Nike+ Sportband, a receiver plugged into
an iPod Nano, or directly with a 2nd, 3rd or 4th Generation iPod Touch, iPhone 3GS or iPhone 4 (Wikipedia, 2011). As of August 2011, the Nike+ iPod sport kit sells for a retail price of $29 and can be purchased worldwide from existing distribution channels of Nike. In September 2010, Nike released the Nike+ GPS App on the Apple platform, which used a tracking engine powered by mobile sensor company MotionX that does not require the separate shoe sensor. This application works using the existing iPhone accelerometer and GPS and the accelerometer of the iPod Touch and sells in the iTunes store for $1.99.

The relationship between Apple and Nike is strong, having first been established in 2001 soon after the release of the first Apple iPod. Nike President and CEO Mark Parker stated that year "Most runners were running with music, we thought the real opportunity would come if we could combine music and data" (McClusky, 2009). Parker had a personal friendship with Apple CEO Steve Jobs and both companies saw profit potential if they could develop the system together, with Apple working on advancement of the Nike sensor prototype by making it smaller and more durable and Nike focused on the shoes, the concept of fitness goal setting and the interface for the Web and the iPod (McClusky, 2009).

Nike+ has introduced additional technology called “Cheer Me On” to allow for motivation during runs. A cheer is heard by the runner, which is activated whenever a friend likes or comments on the user’s run status from the Nike+ website or Facebook (Van Grove, 2010). Friends can also track real time progress by monitoring a user’s workout via the web. Best runs are celebrated by the Nike+ community with motivational messages from Nike’s top athletes including Tiger Woods. These personal milestones and other accomplishments can also be broadcasted to a user’s social network via Facebook and Twitter integration.
By creating a simple way to collect data with tools to use and share it, Nike has created a community of more than 1.2 million runners. Data analysis of these collected runs would suggest that the group has tracked more than 130 million miles and burned more than 13 billion calories (Nike, 2011). Personal habits have also been surmised such as workout patterns during the winter months (people in the US run more often than those in Europe and Africa, but for shorter distances), the average duration of a run worldwide is 35 minutes, and the most popular Nike+ Powersong, which runners can set to give them extra motivation, is "Pump It" by the Black Eyed Peas (McClusky, 2009).

DailyMile

DailyMile is a San Francisco based company described by University of Wisconsin alumni founders Kelly Korevec and Ben Weiner as “a social experience for active people, a community of people just off the couch to ultramarathoners alike, who encourage and inspire one another as we achieve our goals” (DailyMile, 2012).

Founded in 2008, DailyMile was originally designed to cater to active types such as runners and cyclists that often trained alone while incorporating the sharing of workouts via social media, which allows people to train together virtually. The service is a combination fitness log, motivational tool and social-networking hub aimed at using social media to help people achieve their health and wellness goals such as training for a big race or losing weight, all the while connecting with others that are trying to bring fitness and health into their offline lifestyles (Henning, 2010).
As of November 2011, DailyMile reported that over 10.1 million workouts by members were completed, with over 8.9 million member interactions via comments posted, and total member activity accounting for over 72 million doughnuts being burned. The site has reached over 200,000 members and adds over 3,000 new members weekly (DailyMile, 2012). DailyMile currently interfaces with devices such as Nike+, Garmin, Apple mobile platforms and Android clients. Members can download and embed personalised widgets of code that can be added to their own blog or website that tracks exercise mileage.

The DailyMile website has three areas of focus including profile, training and community sections. Within the community section, a member can interact with other members, participate in challenges and forums, view shared exercise routes and enter local fitness events. Unlike the other mobile fitness applications discussed in this research, DailyMile does not have their own app, but rather uses an API to enable third party developers to build applications on the DailyMile Social Workout Platform. These third party apps include but not limited to Electric Miles, Runmeter, LogYourRun, Kinetic and Jog Log and allow functions such as data entry and deletion, comments, likes, friends, routes and GPS location (DailyMile, 2012). DailyMile also uses members to provide crowdsourcing ideas to developers. Members have suggested concepts such as mobile apps, blog integration, Google Health data transfer, nutritional information exchange and workout logging via SMS.

MyFitnessPal

MyFitnessPal is an online health and fitness community that offers useful tools, advice and support to help a person meet their weight loss and fitness goals. The site allows a person to track what they eat and how much exercise they perform in order to give a clear picture of daily/weekly/monthly caloric intake. In addition, the online community offers tips and support to help with motivation along the way. While similar with the other mobile fitness apps discussed in this paper, MyFitnessPal also includes a robust daily food tracking option with a database of over 1.2 million searchable items via a database maintained by the USDA. One feature of the set up process is personalised goal setting with respect to body weight. A user can decide to
gain, lose or maintain weight, however the app restricts the user from losing less than 2 pounds (1 kilogram) (Duffy, 2011).

Based on the user’s fitness profile, MyFitnessPal recommends a daily net calorie target. The tracking of exercise (calories out) and food consumption (calories in) throughout the day adjusts the daily net calorie target. Re-occurring exercises and/or food can be saved as “favourite”, thus allowing for quick logging.

In 2011, a Walden University study proposed that positive social change by tracking calories via smart phones using MyFitnessPal could encourage users to make healthy choices and thus reduce the overall prevalence and incidence of obesity and related health conditions (hypertension, diabetes type 2, and cardiovascular diseases) within their communities (Hijazi, 2012). The MyFitnessPal website suggests:

“Study after study has confirmed the benefits of keeping track of the food you eat and the activity you do. It’s simple - the more consistently you track your food intake, the more likely you are to lose weight. That’s why every successful weight management program suggests that you keep a food diary and/or an activity log. But recording everything you eat without the right tools can be tedious at best, or simply impossible at worst.” (MyFitnessPal, 2012)

The MyFitnessPal service is a free online service with supplemental apps on iPhone and Android platforms for additional methods of data collection. MyFitnessPal integrates with Facebook and Twitter allowing for customised sharing of activities.
The five mobile fitness apps, (Endomondo, RunKeeper, Nike+, DailyMile and MyFitnessPal) that were reviewed, each provided similar functionality with additional capabilities. Below are complete breakdowns of the functions.

Figure 12: MyFitnessPal Screenshots (web)
Table 2: Mobile Fitness App Comparisons (as of Dec 2012)

<table>
<thead>
<tr>
<th></th>
<th>Nike+</th>
<th>RunKeeper</th>
<th>MyFitnessPal</th>
<th>DailyMile</th>
<th>Endomondo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (iOS version)</td>
<td>18.6 MB</td>
<td>11.1 MB</td>
<td>15.9 MB</td>
<td>n/a</td>
<td>5.3 MB</td>
</tr>
<tr>
<td>Native App</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No(^4)</td>
<td>Yes</td>
</tr>
<tr>
<td>Platform</td>
<td>iOS</td>
<td>iOS, Android</td>
<td>iOS</td>
<td>n/a</td>
<td>iOS, Android, Blackberry, Windows</td>
</tr>
<tr>
<td>Cost (iOS version)</td>
<td>$1.99</td>
<td>Free</td>
<td>Free</td>
<td>n/a</td>
<td>Free (Pro $3.99)</td>
</tr>
<tr>
<td>Live Cheering</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>n/a</td>
<td>Yes</td>
</tr>
<tr>
<td># Activities</td>
<td>2(^1)</td>
<td>14(^2)</td>
<td>350(^3)</td>
<td>n/a</td>
<td>51(^5)</td>
</tr>
<tr>
<td>Sharing with Twitter and Facebook</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>On Web</td>
<td>Yes</td>
</tr>
<tr>
<td>Real-time Coaching</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>n/a</td>
<td>Yes</td>
</tr>
<tr>
<td>Real-time Tracking</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>n/a</td>
<td>Yes</td>
</tr>
<tr>
<td>Challenge Friends</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>On Web</td>
<td>Yes</td>
</tr>
<tr>
<td>HR integration</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>n/a</td>
<td>Yes</td>
</tr>
<tr>
<td>GPS</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>n/a</td>
<td>Yes</td>
</tr>
<tr>
<td>Auto Pause</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>n/a</td>
<td>Yes</td>
</tr>
<tr>
<td>Languages</td>
<td>English, Chinese, French, German, Italian, Japanese, Portuguese, Spanish</td>
<td>English</td>
<td>English</td>
<td>n/a</td>
<td>English, Danish, French, German, Spanish</td>
</tr>
</tbody>
</table>

---

1. Walk, Run
2. Running, Cycling, Mountain Biking, Walking, Hiking, Downhill Ski, Cross Country Ski, Snowboarding, Skating, Swimming, Wheelchair, Rowing, Elliptical, Other
3. Activities are manually entered to the app or website.
4. DailyMile is a website portal that uses a number of different mobile apps to enter data
5. Walking, Cricket, Running, Cycling Transport, Cycling Sport, Mountain Biking, Skating, Roller Skiing, Skiing Downhill, Skiing Cross Country, Snowboarding, Kayaking, Kite Stair Climbing, Cross Training, Dancing, Fencing, Football, Rugby, Soccer, Handball, Hockey, Pilates, Polo, Scuba, Squash, Tennis, Table Tennis, Beach Volleyball, Volleyball, Weight Training, Yoga, Martial Arts, Gymnastics
Fitness Tweet Data Collection

To effectively gather and analyse the vast amount of Twitter-related information being generated by the five MFAs in this study, collection and analytical tools were needed including a tweet collection tool and a Twitter interface to build a database of publicly-available data. The cloud based application called TwapperKeeper was used for tweet data collection specific to mobile fitness app (MFA) hashtags, while the Wang/Vickey Tweet Crawler API was used to gather information that linked MFA tweeters and their publicly available demographic data (Vickey, Ginis, & Dabrowski, 2013).

Data collection using TwapperKeeper began on Thursday April 21, 2011, at 00:00 Greenwich Mean Time and continued until September 21, 2011, at 23:59, for a total collection of Twitter data of 184 days. After review of the collected Twitter data specific to the five mobile fitness app hashtags used for this research (#Nike+, #runkeeper, #dailymile, #myfitnesspal and #endomondo), it was determined that 2,856,534 tweets were collected in 23 different languages.

![Tweets by Language (Non-English)](chart)

Figure 13: Collected Fitness Tweets by Language (excluding English)

Figure 13 provides insights into the emerging popularity of fitness tweeting from mobile fitness apps from around the world. Of the 23 identified languages from Twitter, English was by far the most represented with more than 69% of all Tweets. Of particular interest is the popularity of such fitness tweeting activities from Asian speaking users (Japanese and Indonesian) with the exception of China. The lack of Chinese fitness tweets may be related to the national Chinese regulations against the use of Twitter, with the reported 1,402 Chinese tweets most likely being posted from users from outside China. The relative high ranking of Danish speaking users versus total population is due to the fact that Endomondo was developed in Denmark where the primary spoken language is Danish.
However, for the purposes of this research only the English-language tweets were used. Thus, the total number of processed tweets analysed was 1,971,425. Of these tweets, 814,872 were from RunKeeper, 405,708 from Nike+, 296,182 from Endomondo, 255,250 from DailyMile and 199,413 from MyFitnessPal.

![English vs. Non-English Tweets Using Mobile Fitness Apps](image)

**Figure 143: English vs. Non-English Tweets**

Analysis performed on the database was to determine the daily usage of each mobile fitness app. Since there was an increase in the popularity of mobile fitness apps during this time period, as expected, there was a steady increase of the overall daily usage of each mobile fitness app over time. Unexpected were the dramatic decreases in usage during each individual week which affected all mobile fitness apps. Additional analysis suggested that this weekly decrease occurred at the end of the work week (Thursday) and continued throughout the weekend and then increased at the start of the following week. The dates listed on the horizontal axis in Figure 15 indicate Thursdays during the selected time period. It is the opinion of this researcher that this weekly trend is due to the forthcoming weekend with a majority of mobile fitness app users partaking in physical activity, and thus fitness tweeting on work-days rather than weekends. This insight could be very beneficial in the overall planning of a physical activity program by personal trainers or by the user themselves.
Conclusion

The lack of physical activity has been formally recognised as a serious public health burden associated with increased risk for cardiovascular disease, diabetes, obesity, and some cancers (Almeida, 2008). As highlighted in this research, the use of mobile phones and mobile fitness apps can be used as one of the many tools to encourage people to become more physically active. Five mobile fitness apps (Endomondo, RunKeeper, Nike+, DailyMile and MyFitnessPal) were reviewed in this research. As identified, each provided similar functionality with additional capabilities.

Research indicates that there is a potential for technologies to be used as a means for generating positive behavior change. Simple mobile devices can function as an inexpensive, accessible and powerful trigger towards behavior change without the need of overly complex and expensive applications or devices (Foster et al., 2010). This research has shown that there is a tremendous amount of data available to monitor physical activity. Future work will include a more robust mobile fitness app tweet classification system, an in-depth analysis of a mobile fitness app user’s social network, comparison studies of multiple mobile fitness apps and the integration of such research into the behavioral changes needed to impact the overall health and fitness of these users.
REFERENCES


ABOUT THE AUTHORS

Ted Vickey: Ted Vickey is a PhD researcher from the Digital Enterprise Research Institute at the National University of Ireland at Galway. He holds a Bachelor’s degree in exercise and sport science from Penn State and a Masters in international business from the University of Limerick (Ireland). In addition to his academic work, Ted is president of FitWell LLC, a global fitness and wellness consulting company. He served as the executive director of the White House Athletic Center, serving the staff of the executive office of the president for 11 years under the Bush, Clinton, and Bush administrations; is the Nominations Chair for the Board of Directors for the American Council on Exercise; and has worked at the past two British Open Golf Tournaments for ESPN / ABC.

John Breslin: National University of Ireland, Ireland

Dr. Antonio Williams: Antonio S. Williams, Ph.D., is an Assistant Professor at Indiana University. His research interests include fitness and sport brand management as well as new media marketing. In addition to his refereed presentations at professional conferences such as NASSM and SMA, his work has been featured in the ACSM Health and Fitness Journal, The New York Times, and Fox News.
The International Journal of Sport and Society provides a forum for wide-ranging and interdisciplinary examination of sport, including: the history, sociology and psychology of sport; sports medicine and health; physical and health education; and sports administration and management. The discussions in the journal range from broad conceptualizations of the fundamental logics of sport, to highly localized readings of sporting practices in particular times and places.

As well as papers of a traditional scholarly type, this journal invites case studies that take the form of presentations of practice—including documentation of sports practices and exegeses analyzing the social effects of those practices.

The International Journal of Sport and Society is a peer-reviewed scholarly journal.