Text mining of social media feeds to perform sentiment analysis for technology release

Peter Carew*, Ali Akbar Jilani**

**Lecturer in Computing,*
Waterford Institute of Technology, Waterford, Co. Waterford
Ireland (Tel: 353-051-302628; e-mail: pcarew@wit.ie).

***Student, Department of Science and Computing,*
Waterford Institute of Technology, Waterford, Co. Waterford
Ireland (Tel: 353-085-212-6940; e-mail: 20078735@mail.wit.ie).

Abstract: Textual Domain offers a unique proposition for sentiment analysis. After the popularization of online social networks, social media were naturally picked by the industry to evaluate the sentiments that related to their market segments. Twitter represents one of the most popular social media platform where people share their opinions and sentiments using within the 140 characters space made available to them. Keeping in view the volume of content growing on Twitter on daily basis, it was imperative to device some automated tool that performs this sentiment score calculation for a define-able volume of content. A number of techniques have been developed for different aspects of sentiment analysis i.e. document level, sentence level etc. This study is targeted at development of another sentiment analysis tool that uses these available techniques to find their mutual correlation. Something that it calls "axis of honesty" and connects these axes to develop the overall success score of a specific technology product from within the fed Twitter corpus.

Keywords: Social Media, Web 2.0, Text Mining, Sentiment Analysis, technology release.

Introduction:

After the popularization of online social networks (OSNs), Sentiment Analysis has become an extremely popular tool for its application in several analytical domains especially the web and social media. Social media includes a variety of specific tools or applications, such as blogs, microblogs and social networks. Twitter is one of the most popular medium among social media that has contributed to reshaping the web from a mere static repository to a dynamic forum (microblogging service) where users can publish their thoughts and opinions along with other types of "user-generated content (UGC)" on any topic of interest. This content carries valuable information particularly for applications that require analysis of public opinion on a certain topic. This study focuses on one such application where different large organizations tap into this resource as they try to understand public opinion (sentiment analysis) about their products that are either launched or about to be launched. Below semblance (Figure 1: Source Google Trends) depicts the growing trend on the topic of "Sentiment Analysis".

Figure 1: Searches on Google for the query: 'Sentiment Analysis'. This figure shows the steady growth on the number of searches on the topic, according to Google Trends, mainly after the popularization of online social networks (OSNs).

According to Figure 2, most of the searches on Sentiment Analysis are found to be made from India while Singapore and Sri Lanka stand at second and third rank in terms of web searches on the subject. Below is the top 5 ranking of search volumes from different countries across the globe.

Figure 2: Top 5 ranking countries w.r.t. volume of search queries on 'Sentiment Analysis'.

As Figure 3 shows, On the topic of 'Sentiment Analysis', 'Twitter sentiment' is the most common search term as per Google (Trends). The most common application of sentiment analysis is in the area of reviews of consumer products and services.

(Feldman, 2013) Defines 'Sentiment Analysis' or 'Opinion Mining' as the task of finding opinions of authors about specific entities. He explains how there is a huge explosion of 'sentiments' available from social media including Twitter, Facebook, message boards, blogs, and user forums. This opinionated information is a gold mine for companies and individuals that want to monitor their reputation or get timely feedback about their products and actions, may they be about product release. Sentiment analysis offers these organizations the ability to monitor the different social media sites in real time and act accordingly. Marketing managers, campaign managers, politicians, equity investors or even online shoppers can directly benefit from this sentiment analysis technology. (Feldman, 2013).

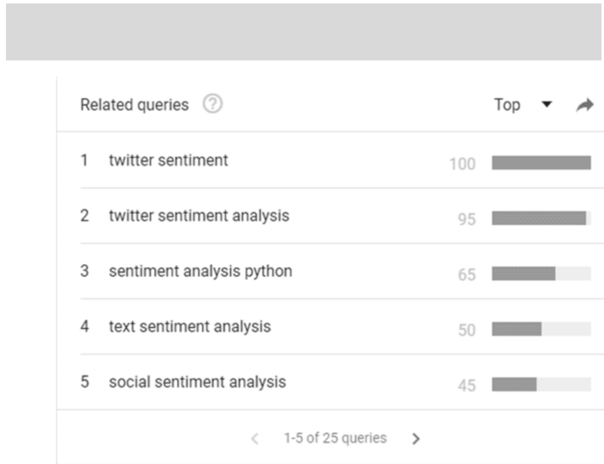


Figure 3: 'Twitter sentiment' is the most common query on the topic of 'Sentiment Analysis' as per Google Trends.

Twitter:

Twitter is one of the most popular microblogging platform that was launched in 2006. As a rough estimate taken in 2016 by (Giachanou and Crestani, 2016), Twitter had 284 million users who posted 500 million messages per day. Characterized by the ease of access and download of posts published through the system, it was considered one of the largest datasets of user-generated content. Twitter is considered an informal mode of social media content.

Tweet:

A single message posted on Twitter is called a "Tweet". Its content may at maximum stretch over 140 characters that can vary from personal information or opinion about products or events to other content types such as photos, videos, news or even links.

Scope:

Why Twitter: Sentiment Analysis is possible across the broad range of social media platforms available today. Below are some of the unique characteristics of twitter that distinguish it from other microblogging platforms such as Tumblr, FourSquare, Google+, and LinkedIn for sentiment analysis.

1. *Standard length:* Tweets have a standard length limitation of 140 characters which gives enough room to the Twitterati to explain his/ her opinion while remaining relevant to the topic.

2. *Informal type of medium:* Twitter seems to be the most suitable out of all other social media platforms as it offers an informal medium of expression (more suitable for subjective content) to its registered users while limiting them to 140 characters which helps control content relevance. Other microblogging platforms are either formal (LinkedIn) or are less popular than Twitter (Tumblr, Google+).

3. *Volume of content:* Over the years, Twitter's interface has remained simple, which is why a lot of tweets take place through third-party sites and applications that make the experience more useful. There could be other sources considered but volume and content relevance become important questions when you consider analyzing sentiments in products that are yet to be announced. In the context of technology release, there is a better chance of finding pre-release product centered content on twitter than any other social media platform also because of its popularity.

It is relevant to mention here that this research is likely to have a universal application. The main outcome of this research is supposed to be a binary indicator representing prediction about the success of technology product release. This could either be 'Success' or 'Failure'. The overall sentiment score across the Twitter corpus can finally be presented to the user using some visualization tool.

According to (Feldman, 2013), More than 7,000 articles have been written about sentiment analysis and various startups are developing tools and strategies to extract sentiments from text. The scope of this study is therefore being drifted more towards contribution to quality (as compared to variety) of results and to make it more manageable, it is also being confined only to *subjective sentences* (that contain opinions, beliefs and views) as opposed to *objective sentences* (that contain factual information). Subjective sentences carry the essence of sentimental information (opinions, beliefs and views) while objective sentences contain factual information that is more suitable to areas like stock picking. A Tweet may vary in the number of sentences it contains. These sentences may carry different opinions about the same entity. In order to develop an accurate and fine-grained view of different opinions, the proposed tool is required to attach sentiment annotations to individual sentences within a tweet. However, In order to limit the scope of research, following assumptions are being made.

- That tweets are written in English language. Re-tweets are excluded from the analysis.
- Since a tweet comprises up of more than a sentence, it may be assumed that the entire tweet contains an opinion on one main object expressed by the "Twitterati" (more reasonable in the context of document-level sentiment analysis)
- That we know the identity of the entity discussed in the sentence.
- It is assumed that each phrase in the sentence also contains just one opinion.
- To further relax the situation, it is assumed that there is a single opinion in each sentence.

Ethical considerations:

Privacy in Twitter is not an issue since Twitter allows users to post messages on its platform after a registration phase during which the user is asked to select a unique pseudonym (username) that further serves as the user's identity. Users may choose to use their original identity instead. All "Mentions" in a tweet indicate the username the tweet is directed at and in order to refer to other users, it uses '@' followed by the username to which it is directed (@username). Across all interactions (replies, follows, retweets), user keeps control over the choice to disclose his/ her original identity or to use a pseudonym. Twitter even gives a user the option to decide if his/ her tweets will be visible to everyone or only to his/ her approved followers.

The study is designed around sentiment analysis of a particular subject that will limit the scope at group level, not an individual user. The topic of interest is also related to "Technology Release" that lies in the public domain and does not pose any privacy challenges. Furthermore, the scope of this study at every level will be defined after a detailed consideration of all possible privacy aspects. The possibility of a misuse or breach of privacy will be minimized.

Research Questions:

The purpose of this paper is to develop a tool/ technique that can be used to answer the following research questions.

RQ1: Is it possible to develop/ a tool or technology to predict success of technology release based upon sentiment analysis performed over pre-release historical social media data?

RQ2: What would such a tool/ technology look like?

Methodology:

Research Approach:

There are multiple methods for measuring sentiments, including lexical-based and supervised machine learning methods. This study is designed to be conducted through development of a software tool that will gather relevant data through a supervised process. The supervised approach assumes that there is a finite set of classes into which data should be classified and training data is available for each class.

The scope for data gathering will be guided through filters applied to isolate only subjective data that represents a more suitable type of data for the adopted techniques.

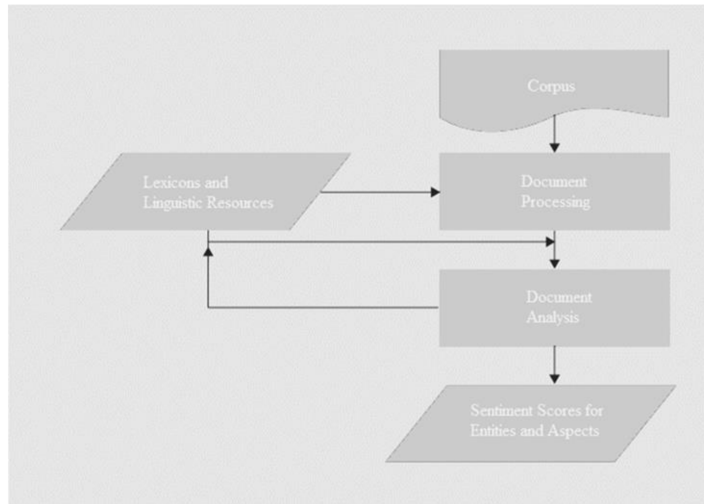


Figure 4: General architecture of a lexical Sentiment Analysis System.

A generic sentiment analysis tool may employ a variety of linguistic techniques such as stemming, tokenization, part of speech tagging, entity extraction, and relation extraction used for pre-processing textual artefacts or the target subject (e.g. a Tweet). The tool may also utilize lexicons or other linguistic resources. The primary component in the tool is Document Analysis module that attach sentiment annotations to the subject (document, sentence or aspect of entity) by utilizing linguistic resources. These sentiment annotations being the primary output of the system may be presented to the user using a variety of visualization tools (Feldman, 2013).

Sentiment Analysis Tool:

Twitter Sentiment Analysis has been performed by several researchers using a variety of tools developed using different languages such as Java. As part of this study, the tool however is planned to be developed using the programing language “R”. The language R is considered a gold standard for development in the numerical analysis and machine learning space. R’s biggest advantage worth a mention here is its package ecosystem (Krill, 2015).

On the other hand “R” is known to have issues in security and memory management. Since the Sentiment Analysis Tool to be developed in this study, is a stand-alone application with no other modules, security should not be an issue. Memory management however is something that requires serious attention especially when talking about large datasets. This risk in application performance should not be serious as a tweet is typically a “bag” of 140 characters in its anatomy. If however the issues arises, it can be addressed by some trade off in terms of simplicity of sentiment analysis algorithm.

Data Collection:

As per (Giachanou and Crestani, 2016), Twitter provides an easy way for developers and researchers to access and gather twitter data using either of two APIs namely, “REST” and “Streaming”. Twitter APIs provide an easy access to large amounts of tweets that have specific characteristics. This allow the creation of a data filtering system that can define scope for data collection e.g. having certain terms or emoticons. Usually, lists of entities, hashtags or emoticons are used to crawl Tweets that are returned in JSON format. JSON is widely used for storing and exchanging data, and can be easily parsed by many programming languages. The metadata returned by the Twitter APIs include information such as publication date, author’s username, location, hashtags, retweets, followers, and many other data.

According to (Giachanou and Crestani, 2016), most researchers prefer the Streaming API because it provides unlimited and real time access to tweets that meet a specific requirement.

Available Online Resources:

A number of sentiment lexicon are available to be used for sentiment analysis. Given below are a few references.

1. General Inquirer Lexicon: http://www.wjh.harvard.edu/~inquirer/ spread-sheet_guide.htm.
2. Emotion Lexicon: http://www.purl.org/net/emolex
3. Financial Sentiment Lexicons: http://nd.edu/~mcdonald/Word_Lists.html.
4. MPQA Subjectivity Lexicon: http://www.cs.pitt.edu/mpqa/subj_lexicon.html
5. SentiWordNet: http://sentiword-net.isti.cnr.it/
6. Sentiment Lexicon: http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

Preliminary Literature Review:

1. Text Mining:

According to (Giachanou and Crestani, 2016), Twitter APIs provide an easy access to large amounts of tweets that have specific characteristics. This allow the creation of a data filtering system that can define scope for data collection e.g. having certain terms or emoticons.

2. Social Media:

(Carol et al., 2013) defines social media or Web 2.0 technologies as "innovative online tools designed to enhance communication and collaboration".
Sentiment Analysis is possible across the broad range of social media microblogging platforms such as Tumblr, FourSquare, Google+, and LinkedIn etc.

3. Sentiment Analysis Methods:

(Peng , 2014) Invented a method for Social Media Sentiment Analysis. Accordingly, embodiments generally relate to systems and methods for generating a sentiment dictionary and calculating sentiment scores of adjectives within the sentiment dictionary. A set of seed words can be identified and expanded using synonyms and antonyms of the set of seed words. Social media data can be parsed to identify adjectives that link to the set of seed words with the words "and" or "but." Matrices representing the attraction and repulsion among the linked adjectives can be generated. A factorization algorithm can be minimized to determine an output matrix that comprises positive and negative sentiment scores for each of the adjectives. In embodiments, a sentiment score for part of all of the social media data can be calculated using the output matrix, and one or more parts of the social media data can be classified as a positive or negative sentiment.

Authors	Data	Approach	Outcome
Sommar and Wielondek (2015)	Movie reviews	Use the outcome of lexicon-based classification to feed machine learning for improved performance and convenience in sentiment classification	Combined approach outperforms the lexicon-based approach, in turn being outperformed by the learning based approach
Mudinas *et al.* (2012)	Software and movie reviews	Lexicon-based output is used to train a learning-based classifier	Hybrid approach improves the accuracy of sentiment classification compared to lexicon only approach, but is less accurate than learning based methods only
Liu *et al.* (2011)	Tweets	A classifier is trained using data given by the lexicon-based approach, instead of being labeled manually	Combined approach improves recall compared to lexicon-based approach only
Prabowo and Thelwall (2009)	Movie reviews, Product reviews, MySpace comments	Multiple sentiment classifiers are used in sequence so that if one classifier fails to classify a document, the classifier will pass the document onto the next classifier, until the document is classified or no other classifier exists	The use of multiple classifiers in a sequential manner can result in better effectiveness than any individual classifier. However, documents were assigned to one sentiment only (binary classification), so that a document containing both conveying both positive and negative sentiment, was necessarily classified as either positive or negative
Tan *et al.* (2008)	Movie Reviews, Computer Reviews, Education Reviews, and House Reviews	Use a lexicon-based technique to label data; then learn a new supervised classifier based on the labeled data	The experimental results indicate that proposed scheme could dramatically outperform "learn based" and "lexicon-based" techniques

Table 1: Previous studies combining lexicon-based and machine learning approaches to sentiment analysis.

(Jeong et al.) Explains different approaches to identification of author's sentiment along with the degree of sentiment

1. Lexicon-based:

This approach uses predefined dictionaries that define sentiment words and their corresponding sentiment values. E.g. SentiwordNet. A number of sentiment lexicon are available to be used for sentiment analysis. Given below are a few references.

- General Inquirer Lexicon: http://www.wjh.harvard.edu/~inquirer/ spread-sheet_guide.htm.
- Emotion Lexicon: http://www.purl.org/net/emolex
- Financial Sentiment Lexicons: http://nd.edu/~mcdonald/Word_Lists.html.
- MPQA Subjectivity Lexicon: http://www.cs.pitt.edu/mpqa/subj_lexicon.html
- SentiWordNet: http://sentiword-net.isti.cnr.it/
- Sentiment Lexicon: http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

(Ribeiro et al., 2016) have conducted a benchmark comparison (apple-to-apple) of twenty-four popular sentiment analysis methods (called the state-of-the-practice methods) across eighteen labeled datasets, originating from messages posted on social networks, movie and product reviews, as well as opinions and comments in news articles was conducted as they are used in practice, across multiple datasets.

2. Sentiment Analysis Algorithms:

There are multiple methods for measuring sentiments, including Classification method (supervised machine learning) and lexical-based method.

Classification Method Algorithms:

(Feldman) Classification algorithms used Given the training data, the system learns a classification model by using one of the common classification algorithms such as Support Vector Machine (SVM), WSVM, C4.5 tree, AdaBoost, Linear kernel SVM, Naïve Bayes (NB), MaxEnt, Multi Naïve Bayes (MNB), CRF, Perceptron with Best Learning Rate, Voted Perceptron, Ensemble Method, Logistic Regression, or kNN.

Lexicon Based Algorithms:

SentiStrength, SentiCircles, Clustering-based word sense disambiguation (WSD), Lexicon-based classifier, Rule-based and ESSA.

This classification is then used to tag new documents into their various sentiment classes.

As per (Giachanou and Crestani) these characteristics themselves pose challenge to sentiment analyst.
(Bermingham and Smeaton) have concluded that classifying tweets is easier than classifying longer documents such as blogs.

3. Challenges in sentiment analysis:

(Giachanou and Crestani) have explained below characteristics of twitter as the main challenges faced by sentiment analysts.

1. *Text Length:* have explained tweet length limitation (140 characters) and informality of medium as challenges.
2. *Topic Relevance:* many researchers of twitter sentiment analysis have been considering presence of a word in a tweet as an evidence of topic relevance while others studies consider the hashtag symbol as a strong indicator or topic relevance. To a certain degree, these approaches may be correct as commonly the sentiment does target the topic.
3. *Incorrect English:* Length Limitation and informality of communication make the language used in tweets is very different than the one used in other geners (web, blog, news etc.)
4. *Data Sparsity:* Owing the large volume of incorrect English and misspelled words, tweets contain an extensive amount of noise called "Data Sparsity" that negatively impacts sentiment analysis. Another reason for this noise is the use of non-standard textual artefacts such as emoticons and informal language. (Jeong et al., 2017) have also mentioned emoticons ('^^', ':-D') and onomatopoeic words ('haha', 'blah') as a type of noice.
5. *Compositional Sentiments:* (Feldman) has expressed the need for better modeling of Compositional Sentiments. At sentence level, this means more accuracy is required in overall sentence sentiment calculation from sentiment-bearing words, the sentiment shifters and the sentence structure.
6. *Anaphora and Auto-Entity Resolution:* Typically in an informal mode of communication, a product may be referred to by multiple names within a context. Anaphora resolution refers to aspect extraction e.g. "battery life" and "power usage" both mean the same thing(Feldman).

1. Technology Release

(Jeong et al.) See social media as an emerging source of customer voice since it assumed the form of a channel for exchanging and storing consumer-generated, large-scale, and unregulated voices about products. The authors have proposed a 4 step opportunity mining (identification of product opportunities) approach based upon topic modeling and sentiment analysis of large-scale customer generated social media data using open APIs. Below are the different steps discussed in the approach.

1. Use topic modeling to identify latent product topics used by product customers in social media
2. Quantify the importance of each product topic.
3. Use sentiment analysis to evaluate satisfaction level of each product using sentiment analysis.
4. Use the opportunity algorithm that uses product topic importance and satisfaction to determine opportunity value and improvement direction of each product topic from a customer centered view.

As a case study, opportunity mining of Samsung Galaxy Note 5 has been described as performed through the use of AIChemyAPI included in IBM's Watson platform.

Description of the Experimental Design/ Verification Methodology:

As per (Feldman, 2013), there are two main approaches to sentiment analysis, "Supervised" and "Unsupervised". A single tweet may contain more than one sentence. It is important to note here that each sentence can have its own **Sentiment Orientation** (SO). Which simply means that, it is fair to treat it as a document to apply document sentiment analysis techniques to it or to treat it as a "bag of words". A tweet level sentiment score can therefore be calculated (using linguistic resources) and associated with each tweet. The same tweet in another way can also be broken-down into sentences, which implies that sentence level sentiment scores can also be associated to each sentence. There are separate techniques/algorithms used for document sentiment calculation and sentence sentiment calculation. It should be possible to compare the two outcomes (document sentiment & sentence sentiment) with each other.

As part of this study, it is hypothesized that

H1: The document sentiment of a tweet and its aggregate sentence level sentiment are equivalents/ correlated.

H2: There is a strong correlation between document (tweet) sentiment score calculated using one technique and the aggregate sentence-sentiment (score) calculated for the same tweet using at least one of the sentence level techniques. This correlation determines the "axis of honesty".

H3: These two techniques (tweet level technique and sentence level technique) are not always the same for every set of tweets taken from the Twitter corpus.

H4: The overall success of a technology release will be likely if the aggregate of these document level sentiment scores is positive and unlikely if their aggregate is negative (falls below zero).

H5: A dominant pattern is visible when all these sentiments are presented using a visualization tool.

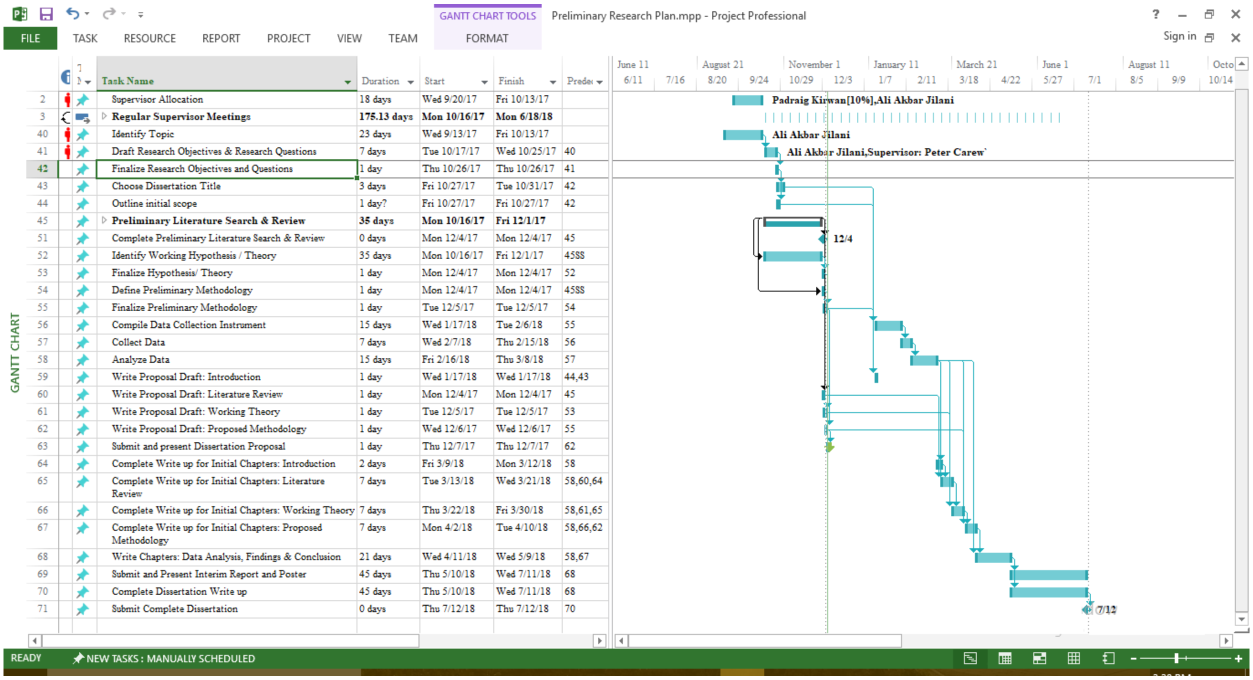
Special Resources Required:

Since the Software development tool is to be developed using **"R"** as a programming language, An IDE would be required. A number of freely downloadable IDEs are available online. The most suitable in terms of features seems to be "**RStudio**" which has a Desktop and Server edition. Open Source versions of both, the RStudio Desktop and RStudio Server are freely available for download (https://www.rstudio.com/products/rstudio/), A number of R Packages for Text mining and Sentiment Analysis are available out of which **TM** (Text Mining) and **RSentiment** initially seem to be interesting. For visualization and graphics, **Shiny** (https://shiny.rstudio.com/) will be used.

Main Milestones Anticipated:

Task #	Task Name	Duration	Start	Finish	Predecessors
2	Supervisor Allocation	18 days	Wed 9/20/17	Fri 10/13/17	
3	**Regular Supervisor Meetings**	**175.13days**	**Mon 10/16/17**	**Mon 6/18/18**	
40	Identify Topic	23 days	Wed 9/13/17	Fri 10/13/17	
41	Draft Research Objectives & Research Questions	7 days	Tue 10/17/17	Wed 10/25/17	40
42	Finalize Research Objectives and Questions	1 day	Thu 10/26/17	Thu 10/26/17	41
43	Choose Dissertation Title	3 days	Fri 10/27/17	Tue 10/31/17	42
44	Outline initial scope	1 day	Fri 10/27/17	Fri 10/27/17	42
45	**Preliminary Literature Search & Review**	**35 days**	**Mon 10/16/17**	**Fri 12/1/17**	
46	Literature Search & Review : Social Media	35 days	Mon 10/16/17	Fri 12/1/17	
47	Literature Search & Review : Twitter	35 days	Mon 10/16/17	Fri 12/1/17	
48	Literature Search & Review : Sentiment Analysis	35 days	Mon 10/16/17	Fri 12/1/17	
49	Literature Search & Review : Technology Release	35 days	Mon 10/16/17	Fri 12/1/17	
50	Literature Search & Review : Advertising Campaign	35 days	Mon 10/16/17	Fri 12/1/17	
51	Complete Preliminary Literature Search & Review	0 days	Mon 12/4/17	Mon 12/4/17	45
52	Identify Working Hypothesis / Theory	35 days	Mon 10/16/17	Fri 12/1/17	45SS
53	Finalize Hypothesis/ Theory	1 day	Mon 12/4/17	Mon 12/4/17	52
54	Define Preliminary Methodology	1 day	Mon 12/4/17	Mon 12/4/17	45SS
55	Finalize Preliminary Methodology	1 day	Tue 12/5/17	Tue 12/5/17	54
56	Compile Data Collection Instrument	15 days	Wed 1/17/18	Tue 2/6/18	55
57	Collect Data	7 days	Wed 2/7/18	Thu 2/15/18	56
58	Analyze Data	15 days	Fri 2/16/18	Thu 3/8/18	57
59	Write Proposal Draft: Introduction	1 day	Wed 1/17/18	Wed 1/17/18	44,43
60	Write Proposal Draft: Literature Review	1 day	Mon 12/4/17	Mon 12/4/17	45
61	Write Proposal Draft: Working Theory	1 day	Tue 12/5/17	Tue 12/5/17	53
62	Write Proposal Draft: Proposed Methodology	1 day	Wed 12/6/17	Wed 12/6/17	55
63	Submit and present Dissertation Proposal	1 day	Thu 12/7/17	Thu 12/7/17	62
64	Complete Write up for Initial Chapters: Introduction	2 days	Fri 3/9/18	Mon 3/12/18	58
65	Complete Write up for Initial Chapters: Literature Review	7 days	Tue 3/13/18	Wed 3/21/18	58,60,64
66	Complete Write up for Initial Chapters: Working Theory	7 days	Thu 3/22/18	Fri 3/30/18	58,61,65
67	Complete Write up for Initial Chapters: Proposed Methodology	7 days	Mon 4/2/18	Tue 4/10/18	58,66,62
68	Write Chapters: Data Analysis, Findings & Conclusion	21 days	Wed 4/11/18	Wed 5/9/18	58,67
69	Submit and Present Interim Report and Poster	45 days	Thu 5/10/18	Wed 7/11/18	68
70	Complete Dissertation Write up	45 days	Thu 5/10/18	Wed 7/11/18	68
71	Submit Complete Dissertation	0 days	Thu 7/12/18	Thu 7/12/18	70

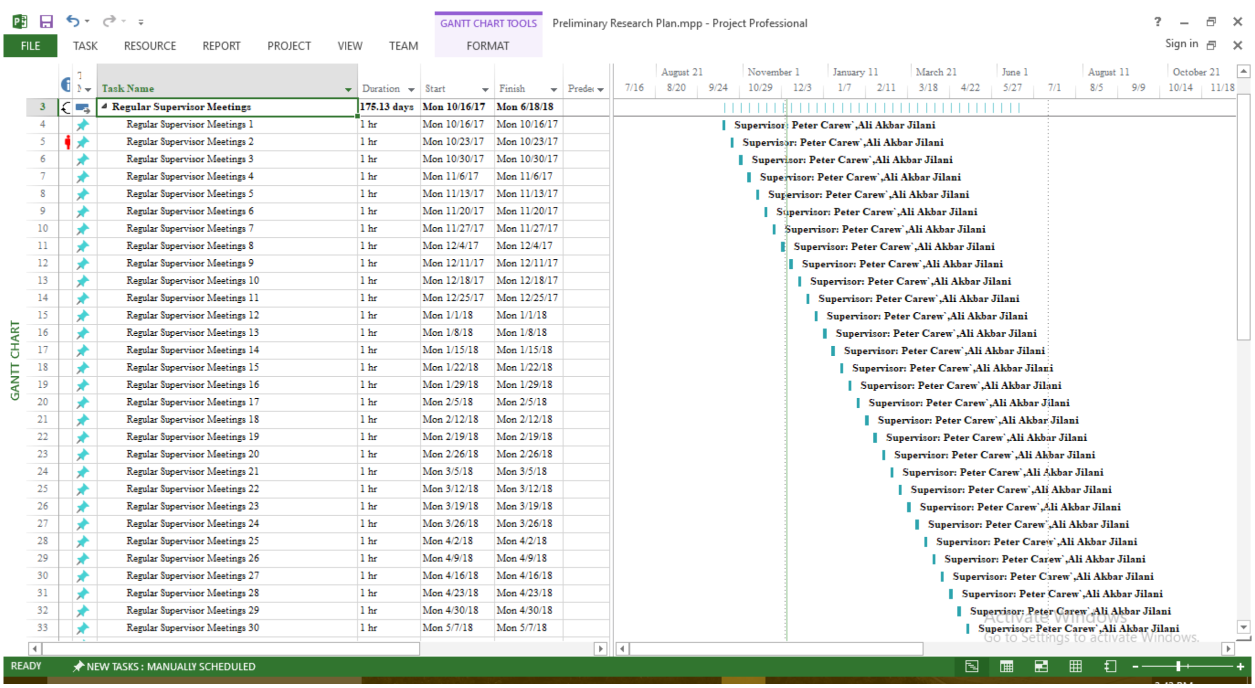
Gantt chart:



	Task Name	Duration	Start	Finish	Predecessors
2	Supervisor Allocation	18 days	Wed 9/20/17	Fri 10/13/17	
3	Regular Supervisor Meetings	175.13 days	Mon 10/16/17	Mon 6/18/18	
40	Identify Topic	23 days	Wed 9/13/17	Fri 10/13/17	
41	Draft Research Objectives & Research Questions	7 days	Tue 10/17/17	Wed 10/25/17	40
42	Finalize Research Objectives and Questions	1 day	Thu 10/26/17	Thu 10/26/17	41
43	Choose Dissertation Title	3 days	Fri 10/27/17	Tue 10/31/17	42
44	Outline initial scope	1 day?	Fri 10/27/17	Fri 10/27/17	42
45	Preliminary Literature Search & Review	35 days	Mon 10/16/17	Fri 12/1/17	
51	Complete Preliminary Literature Search & Review	0 days	Mon 12/4/17	Mon 12/4/17	45
52	Identify Working Hypothesis / Theory	35 days	Mon 10/16/17	Fri 12/1/17	45SS
53	Finalize Hypothesis/ Theory	1 day	Mon 12/4/17	Mon 12/4/17	52
54	Define Preliminary Methodology	1 day	Mon 12/4/17	Mon 12/4/17	45SS
55	Finalize Preliminary Methodology	1 day	Tue 12/5/17	Tue 12/5/17	54
56	Compile Data Collection Instrument	15 days	Wed 1/17/18	Tue 2/6/18	55
57	Collect Data	7 days	Wed 2/7/18	Thu 2/15/18	56
58	Analyze Data	15 days	Fri 2/16/18	Thu 3/8/18	57
59	Write Proposal Draft: Introduction	1 day	Wed 1/17/18	Wed 1/17/18	44,43
60	Write Proposal Draft: Literature Review	1 day	Mon 12/4/17	Mon 12/4/17	45
61	Write Proposal Draft: Working Theory	1 day	Tue 12/5/17	Tue 12/5/17	53
62	Write Proposal Draft: Proposed Methodology	1 day	Wed 12/6/17	Wed 12/6/17	55
63	Submit and present Dissertation Proposal	1 day	Thu 12/7/17	Thu 12/7/17	62
64	Complete Write up for Initial Chapters: Introduction	2 days	Fri 3/9/18	Mon 3/12/18	58
65	Complete Write up for Initial Chapters: Literature Review	7 days	Tue 3/13/18	Wed 3/21/18	58,60,64
66	Complete Write up for Initial Chapters: Working Theory	7 days	Thu 3/22/18	Fri 3/30/18	58,61,65
67	Complete Write up for Initial Chapters: Proposed Methodology	7 days	Mon 4/2/18	Tue 4/10/18	58,66,62
68	Write Chapters: Data Analysis, Findings & Conclusion	21 days	Wed 4/11/18	Wed 5/9/18	58,67
69	Submit and Present Interim Report and Poster	45 days	Thu 5/10/18	Wed 7/11/18	68
70	Complete Dissertation Write up	45 days	Thu 5/10/18	Wed 7/11/18	68
71	Submit Complete Dissertation	0 days	Thu 7/12/18	Thu 7/12/18	70

Weekly Supervisor Meetings:

The task is arranged to occur recursively for 36 instances as scheduled below.



	Task Name	Duration	Start	Finish	Predecessors
3	Regular Supervisor Meetings	175.13 days	Mon 10/16/17	Mon 6/18/18	
4	Regular Supervisor Meetings 1	1 hr	Mon 10/16/17	Mon 10/16/17	
5	Regular Supervisor Meetings 2	1 hr	Mon 10/23/17	Mon 10/23/17	
6	Regular Supervisor Meetings 3	1 hr	Mon 10/30/17	Mon 10/30/17	
7	Regular Supervisor Meetings 4	1 hr	Mon 11/6/17	Mon 11/6/17	
8	Regular Supervisor Meetings 5	1 hr	Mon 11/13/17	Mon 11/13/17	
9	Regular Supervisor Meetings 6	1 hr	Mon 11/20/17	Mon 11/20/17	
10	Regular Supervisor Meetings 7	1 hr	Mon 11/27/17	Mon 11/27/17	
11	Regular Supervisor Meetings 8	1 hr	Mon 12/4/17	Mon 12/4/17	
12	Regular Supervisor Meetings 9	1 hr	Mon 12/11/17	Mon 12/11/17	
13	Regular Supervisor Meetings 10	1 hr	Mon 12/18/17	Mon 12/18/17	
14	Regular Supervisor Meetings 11	1 hr	Mon 12/25/17	Mon 12/25/17	
15	Regular Supervisor Meetings 12	1 hr	Mon 1/1/18	Mon 1/1/18	
16	Regular Supervisor Meetings 13	1 hr	Mon 1/8/18	Mon 1/8/18	
17	Regular Supervisor Meetings 14	1 hr	Mon 1/15/18	Mon 1/15/18	
18	Regular Supervisor Meetings 15	1 hr	Mon 1/22/18	Mon 1/22/18	
19	Regular Supervisor Meetings 16	1 hr	Mon 1/29/18	Mon 1/29/18	
20	Regular Supervisor Meetings 17	1 hr	Mon 2/5/18	Mon 2/5/18	
21	Regular Supervisor Meetings 18	1 hr	Mon 2/12/18	Mon 2/12/18	
22	Regular Supervisor Meetings 19	1 hr	Mon 2/19/18	Mon 2/19/18	
23	Regular Supervisor Meetings 20	1 hr	Mon 2/26/18	Mon 2/26/18	
24	Regular Supervisor Meetings 21	1 hr	Mon 3/5/18	Mon 3/5/18	
25	Regular Supervisor Meetings 22	1 hr	Mon 3/12/18	Mon 3/12/18	
26	Regular Supervisor Meetings 23	1 hr	Mon 3/19/18	Mon 3/19/18	
27	Regular Supervisor Meetings 24	1 hr	Mon 3/26/18	Mon 3/26/18	
28	Regular Supervisor Meetings 25	1 hr	Mon 4/2/18	Mon 4/2/18	
29	Regular Supervisor Meetings 26	1 hr	Mon 4/9/18	Mon 4/9/18	
30	Regular Supervisor Meetings 27	1 hr	Mon 4/16/18	Mon 4/16/18	
31	Regular Supervisor Meetings 28	1 hr	Mon 4/23/18	Mon 4/23/18	
32	Regular Supervisor Meetings 29	1 hr	Mon 4/30/18	Mon 4/30/18	
33	Regular Supervisor Meetings 30	1 hr	Mon 5/7/18	Mon 5/7/18	

References:

BERMINGHAM, A. & SMEATON, A. F. 2010. Classifying sentiment in microblogs: is brevity an advantage? Ireland, Europe: Association for Computing Machinery.

CAROL, T., RACHEL, V. & DONALD, W. K. 2013. Social media and scholarly reading. *Online Information Review,* 37**,** 193-216.

FELDMAN, R. 2013. Techniques and applications for sentiment analysis. *Commun. ACM,* 56**,** 82-89.

GIACHANOU, A. & CRESTANI, F. 2016. Like It or Not: A Survey of Twitter Sentiment Analysis Methods. *ACM Computing Surveys,* 49**,** 28-28:41.

JEONG, B., YOON, J. & LEE, J.-M. 2017. Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International Journal of Information Management*.

KRILL, P. 2015. *Why R? The pros and cons of the R language* [Online]. Available: https://www.infoworld.com/article/2940864/application-development/r-programming-language-statistical-data-analysis.html [Accessed 08/12/2017 2017].

PENG , E. A. May 13, 2014. *Systems, methods and devices for generating an adjective sentiment dictionary for social media*

sentiment analysis: [US Patent & Trademark Office, Patent Full Text and Image Database]. United States, CA, Sunnyvale. patent application 13/082,963.

RIBEIRO, F. N., ARAÚJO, M., GONÇALVES, P., ANDRÉ GONÇALVES, M. & BENEVENUTO, F. 2016. SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science,* 5**,** 23.

TRENDS, G. *Google Trends - Sentiment Analysis* [Online]. Available: https://trends.google.com/trends/explore?date=all&q=sentiment%20analysis [Accessed 19/11/2017].