**Using binary classifier for predicting success of technology release based upon sentiment analysis of social media data**

Ali Akbar Jilani

Student No. 20078735

Department of Science and Computing,

Waterford Institute of Technology, 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. The abstract should summarize the contents of the paper in short terms, i.e. 150-250 words.

**Keywords:** Sentiment Analysis, Binary Classification, Social Media, Twitter, Machine Learning, Text Mining, Technology Release, Product Success, Online Review.

1. Introduction
   1. Background

As online social networks (OSNs) have enabled worldwide consumers to openly communicate their experiences, it has created opportunities for technical communicators, marketing and public relations writers and pretty much any company or individual that want to monitor their reputation or get timely feedback about their products and actions. Social media platforms including Twitter, Facebook, Message Boards, Blogs and user forums offer an explosion of user-generated content (UGC) that can be tapped to ad-hoc corpus building processes thus creating word lists relevant to specific organizational interests. This way, technical communicators as well as marketers can listen to their external users and accurately identify area of need. 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.

While sentiment analysis technology doesn’t stop us from employing artificial intelligence in a program to measure opinion scores on a specific subject, this study focuses on the current need of a much simpler approach of understanding public opinion patterns (sentiment analysis) about a certain technology while maintaining the socio-technical context from which these patterns emerge. It utilizes Amazon online public reviews about technology to train a binary classifier model on success of technology release and the same model is later used to predict success for upcoming technology products.

* 1. Research Objective and Questions

Objective of this study is to investigate,

Objective should cover the question and vice versa.

Scope of study.

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 or mere as WordCloud. It is relevant to mention here that this research is likely to have a universal application.

According to research1, 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 only English tweets will be considered in this study. 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 registration, 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.

* 1. Research Questions:

The purpose of this research is to investigate public sentiments from their experiences of technology and to tally them as they occur on social media as well as in technology reviews found in sites such as Amazon. This research also involves the correlation between these sentiments as recorded through both the above sources. The underlying research question can therefore be defined as the following.

**RQ1:** Does aggregate social media sentiment agree with online user reviews of technology?

Social Media Data.

Binary Classifier.

1. Literature Review
   1. Sentiment Analysis

Feldman1 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.

* + 1. Challenges in Sentiment Analysis

Giachanou and Crestani2 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 noise.
5. *Compositional Sentiments:*  Feldman1 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:* Feldman1 specifies twitter as 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.
   * 1. Sentiment Analysis Algorithms

*Classification Based Algorithm:*

In the territory of document level sentiment analysis as suggested by Feldman1, Support Vector Machine (SVM), Linear kernel SVM, Naïve Bayes (NB), WSVM, C4.5 tree, AdaBoost, MaxEnt, Multi Naïve Bayes (MNB), CRF, Perceptron with Best Learning Rate, Voted Perceptron, Ensemble Method, Logistic Regression, or kNN are a few mentioned algorithms that can be used to perform sentiment analysis.

*Lexicon Based Algorithms:*

Giachanou and Crestani2 have mentioned that Lexicon Based Algorithms have been extensively applied on conventional text such as blogs, forums and product reviews but have less been explored for Twitter Sentiment Analysis. SentiStrength, SentiCircles, Clustering-based Word Sense Disambiguation (WSD), and Lexicon-based classifiers are a few mentioned lexicon based algorithms. In addition, a three step technique for TSA (proposed by Ortega et al) comprises preprocessing as step one, polarity detection as step two, and rule based classification as step three. Last two steps were based on WordNet and SentiWordNet.

SentiStrength being one of the most well-known lexicon-based algorithms developed for social media uses a list of emoticons, negations, and boosting words to effectively identify sentiment strength of informal text including tweets using a human-coded lexicon that contains words and phrases frequently confronted in social media.

* + 1. Sentiment Analysis Approaches

Giachanou and Crestani2 mention document, sentence and entity levels as the three different levels at which Sentiment Analysis (SA) have been applied in literature. Document level sentiment analysis aims to identify sentiment polarity that is expressed in the whole document. Sentence level sentiment analysis on the other hand classifies each sentence as positive or negative and entity level SA detects sentiment polarity of specific entity/ target of a specific object.

Mainly due to size limitations imposed by twitter, most of tweets contain a single sentence. Therefore, there is no fundamental difference document and sentence level when it comes to TSA. In case of tweets, SA can be applied on message/ sentence and entity level.

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

Document level..

Comparable with twitter..

**Supervised Learning:**

As per research1, supervised approach to document level sentiment analysis expects a finite set of classes with access to training data for each one of them. A common example would be to classify (tag) a document into either positive or a negative class. The case can further be extended to also include a third neutral class. A relatively advanced case could contain a discrete numeric scale to classify the document into (like the five-star system employed by Amazon).

Using the training data, the system learns a classification model by using one of the common classification algorithms. The model is then used to tag new documents into various sentiment classes. Regression can further be used to assign a numeric value (in a finite range) to a document. Feldman also refers to research that shows that good accuracy is achieved even when each document is represented as simple **bag of words**. More advanced representations utilize TFIDF, POS (Part of Speech) information, sentiment lexicons, and parse structures.

**Unsupervised Learning:**

Unsupervised learning approach to document level sentiment is based upon the determination of semantic orientation (SO) of specific phrases within the document. These phrases can be selected using either of POS pattern or a lexicon of sentiment words and phrases. Thus, if the average SO of these phrases falls above a specific threshold, it is classified as positive whereas, the average SO below the predefined threshold value classifies the document into negative class.

Another classic way of calculating SO of a given word or phrase is to calculate the statistical difference between the Pointwise Mutual Information: PMI(P,W) of the phrase ‘P’ with two sentiment words ‘W’ based on their co-occurrence in a given corpus or over the web. As an example quoted by Feldman, the SO measures whether the phrase P is closer in meaning to ‘excellent’ being positive word or ‘poor’ (negative word).

* 1. Data Cleaning Activities

A step-by-step3 approach to harvesting data via the Twitter application program interface (API), Cleaning of the data and the basic sentiment analysis code in R language provides value as a single and quick reference that still holds valid. Other closest references from Packt4, de Vries approach5, The Airline Consumer Sentiment Analysis6 and Breen’s approach7 are all outdated.

* 1. Binary Classifier

Sample Heading (Third Level). Content:

* 1. 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.

* 1. Social Media Data
     1. **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.
     2. **Structure of a Tweet:** 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 (R Script) that will gather relevant data through a supervised process using “twitter” package. The supervised approach to binary classification assumes that there is a finite set of classes into which data should be classified and training data is available for each class.

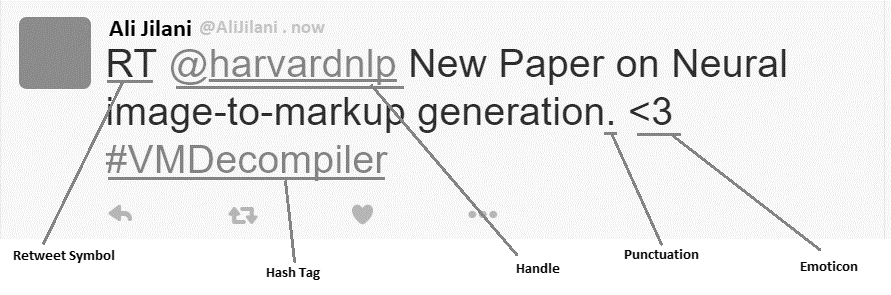


Figure 1: Components of a typical (re)-tweet.

* 1. Machine Learning
  2. 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.

* + 1. Supervised Learning
    2. Unsupervised Learning
  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.

* 1. Success of technology release
  2. Online Review

Sample Heading (Third Level). Content:

1. Theory

As part of this study, it is hypothesized that

* 1. Hypothesis 1: Aggregate sentiment analysis score on social media will correlate to aggregate online user reviews of technology.

Sentiment Analysis Algorithm employed as part of this study will make use of a set of sentiment lexicons (positive and negative lexicons) of different sentiment words to gauge the aggregate sentiment of a corpus related to a specific technology product release. Average star rating as taken from online reviews site (e.g. Amazon) is as such a measure of public sentiment about the same technology product released but on a different scale. It is hypothesized that there is a correlation between both the sentiment scores. (Relate it back to literature..)

* 1. Hypothesis 2: Binary Classification could be used to classify social media sentiment as positive or negative in a way that corresponds with online user reviews of technology.

Instead of aggregate sentiment score, Our binary classification model such as SVM or Perceptron will be fed with twitter data. All the words in twitter corpus related to a specific technology product release will serve as variables while their frequencies of occurrence in the corpus will be treated as values. Our model will be trained using supervised learning to be told if the overall set of words and values corresponds to success or to failure (calculated through the average online review on Amazon as a side process). Twitter corpora relating to a minimum of 10 technology products released will be used to develop a learning for our model. Upon successful learning, the same model will be utilized in predicting the outcome against in terms of either of success or failure. It is hypothesized that this binary outcome of our model actually corresponds with online user reviews. (Relate it back to literature..)

1. Methodology

Explain Methodology at a much more specific level but relate it to research questions e.g. Twitter data SVM / Perceptron, tools used for analysis, data sources, and data parameters (1000 tweets). Justify your choice. Refer back to sections in Literature Review.

* 1. Introduction.. purpose of this section: outlines the methodology adopted and r.design used in order to test the hypo stated/ presented in section 3…

1. Can you get to correltate SVM with reviews.
2. Sentiment scores correspond to stars on reviews. Are they the same as social media buzz.
3. does the star review go up when sentiment score goes up.. and vise versa..

Success Criteria

1. how many items were sold until the next model was launched.. could be a criteria of success.

Success of a technology product is a subjective matter that can be explained in lots of different ways. Market share of a technology product or number of items sold since product release could be considered different examples of success criteria. Online product reviews representing public sentiments about a product can also be taken as success criteria.

Online product reviews are discrete representations of public sentiments and to some extent a representation of product success. To be able to relate it to public sentiments about the same product as found on twitter, we device a script using R programming language to procedurally convert these public words into their equivalents of sentiments over a scale comparable with online user reviews. This study employs a binary classifier to investigate the possibility of correlating social media buzz with public reviews on Amazon. This section outlines the methodology adopted and the research design used in order to test the hypotheses presented in section 3.

For the scope of this study, Tweets are considered ‘a bag of words’ as mentioned in section 2.1.3. Similarly, either of document or sentence level supervised Machine Learning technique can confidently be applied on tweets on message/ sentence and entity level.

Sentiment Analysis Algorithm in the study is applied to a corpus of 1000 tweets pulled live based upon a criteria matching our specified keywords (name of technology product released E.g. iPhone 4) and defined language filter using twitter APIs called into our R Script through twitteR package using step-by-step3 approach. The same approach is followed further in cleaning the data. These vanilla tweets are converted into data frame being more suitable for the purpose of text analysis. The harvested tweets are extracted from this collection to form a corpus comprising Text vectors. A series of data cleaning steps is performed using the text mining package TM. Thus the corpus is processed to remove stopwords, Retweet RT and @Handle labels, #Hashtags, different types of URLs (tiny and normal), all special characters other than English letters and spaces, Numbers and Punctuations. These vectors are then homogenized into lower case before stripping off whitespaces. Other techniques like removal of stem-words etc. are also applied to prepare the corpus for analysis.

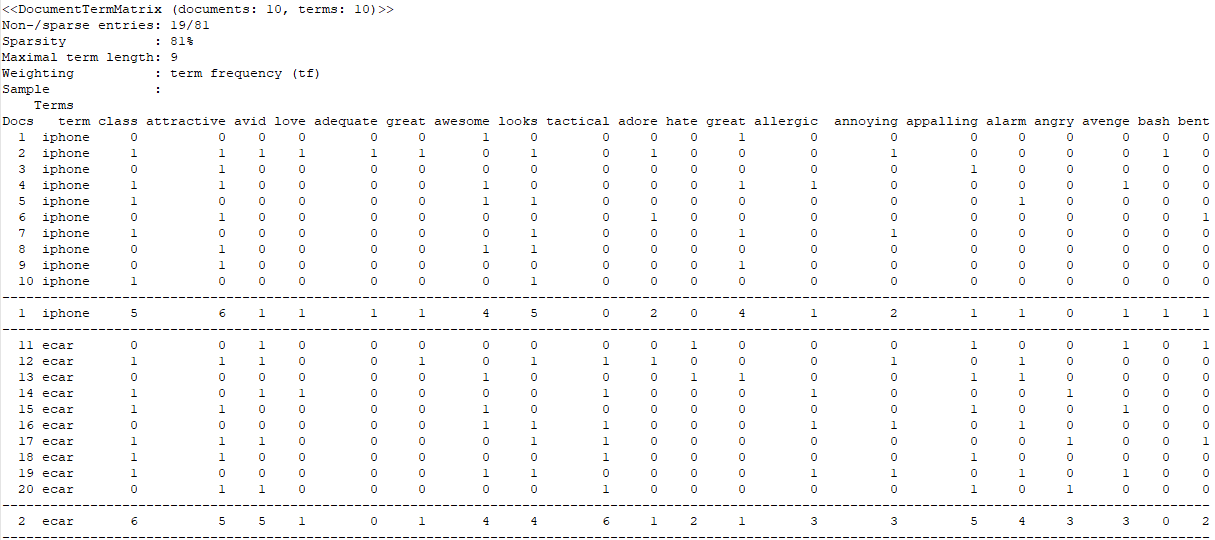


Figure 3: Term frequencies summed up to form a single data row.

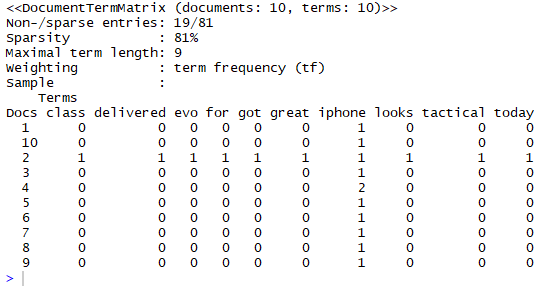


Figure 2: A sample Document Term Matrix (DTM).

Frequencies of different words are then calculated per tweet in the form of Document Term Matrix (DTM). A DTM is a matrix that arranges documents along rows while individual terms/ words in the tweets are arranged into columns. These terms comprising both positive as well as negative terms are compared with a lexicon of positive and negative opinion/ sentiment words.

The negative sentiment words lexicon comprises 4782 frequently occurring words while the positive sentiment words lexicon comprises 2006 frequently occurring words.

The entire DTM of 1000 rows is combined and their term frequencies summed up to form a single data row.

The same process is repeated per technology keywords to form a matrix having rows equal to the number of technologies being investigated in the experiment and columns equal to the union of columns from all the respective DTMs. This would certainly mean that a term not found in one DTM will assume all zeroes as the respective frequencies against columns of words not found in that particular tweet.

A Sentiment Analysis Algorithm will further be employed to formulate the aggregate sentiment score per technology product as a function of sentiment scores of happy words and the sad words (say on the scale of 1 to 10).

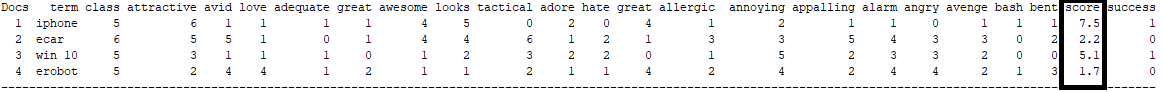
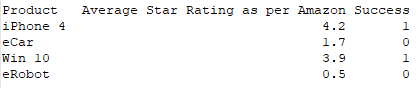


Figure 4: DTM with aggregate sentiment scores calculated.

On the other hand, product reviews are inspected over reputable websites to gauge an average public liking for the same technology product over a unified scale. Amazon is one such reputable website that covers the largest number of technology products.

The same set of technology products are inspected to identify their average review over Amazon. A typical Amazon review is marked over a scale of 1 to 5 where 1 represents the least rated product while 5 marks a highly rated product. For the sake of clarity, a threshold value of 3 is decided to mark the boundary value between a successful and non-successful product. i.e. a product given rating <=3 will be considered non-successful while any product rated above 3 (three) will be considered successful and a representation of positive sentiments towards that technology. Below table represents a technology product, its average star rating taken from Amazon and a column representing success/ failure (+1/-1).



Pearson’s R will be employed to find a correlation between the overall Unified Sentiment Scores and the Amazon average star rating. As per Pearson’s correlation, a P value (measure of error) less than 0.05 will be considered to be significant enough to confirm the validity of the first hypothesis. Similarly the value of Pearson’s R (between 0 and 1) will represent the measure of correlation or correlation coefficient.

To test our second hypothesis, a machine learning (binary classification) model will treat our tweet corpus (in data frame format and relating to a separate technology product) as a bag of words. After performing a set of data cleaning and the pre-processing activities defined under section (2.n), the data frame will be converted to document term matrix (DTM) where terms will represent columns and their respective frequencies of occurrence as rows. The entire DTM rows will be combined into a single row (summary row) representing sum of term frequencies per column. The above mentioned steps will be performed with all data belonging to all the technology products being considered in the study and their resulting rows of DTM terms will be combined into a single DTM.

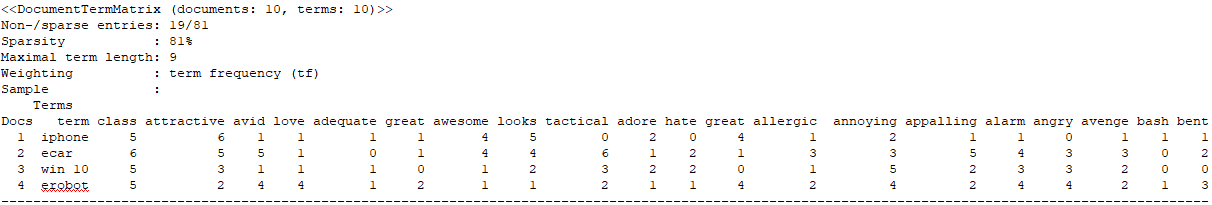


Figure 6: DTM combining the summary rows representing each of technology products.

Figure 7: DTM with each technology product represented as a single row.

Figure 8: A matrix of all technologies summarized per data row.

.

A binary classification model will treat these terms as features and their respective frequencies as values to be trained on the process of calculating success as an outcome. As such, the model will be classifying twitter sentiment as positive or negative. How many technology products and their categorization remains arguable but in respect for time and resource limitations, this study is planned to be limited to 10 technology products.

On the other hand, the Amazon product reviews will be used as a standard to determine success of the same technology product. This study assumes a rating above 3 as a measure of product success while anything equal to or below 3 as failure and is considered to be a representation of negative sentiment.

The binary classifier will learn by continuously accommodating the hyperplane to redefine a decision boundary by also including the newly fed data. Each row of data adds to the decision making capabilities of the model until it is able to generalize the overall behavior.

A new data frame with another similar technology product will then be fed to the learned binary classier and asked to predict success of technology. It is hypothesized that the trained binary classifier will be able to classify sentiments as positive or negative in a way that corresponds with online user reviews of technology.

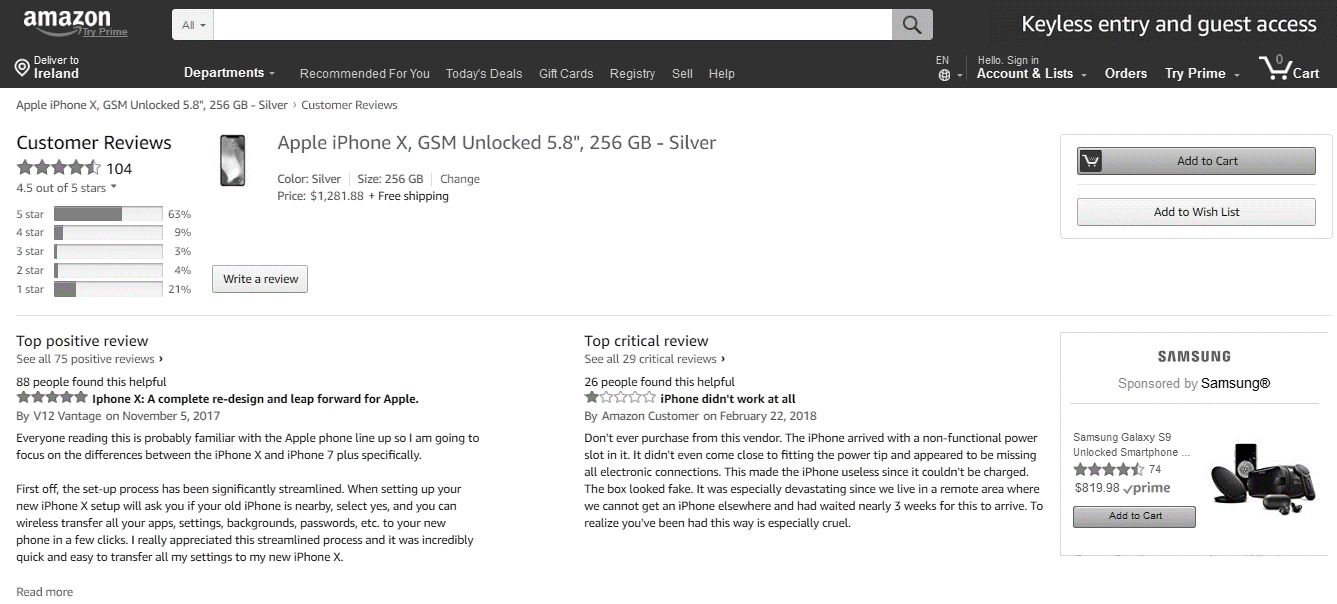


Figure 5: Amazon reviews on iPhone X.

.

A positive correlation will mean that Amazon star reviews will go up when then aggregate twitter sentiment goes up and the vice versa but the reverse will be true if the correlation is found to be negative.

As per Section (3.n.x), the resulting prediction will be compared with the Amazon average star rating to calculate error which will be taken as the discrete difference between what the classifier is expected to predict and what it actually predicts.

* 1. Visual Representation of Results:

The step-by-step approach to harvesting, and cleaning twitter data goes further into different visual representations like WordCloud and Histogram given below.

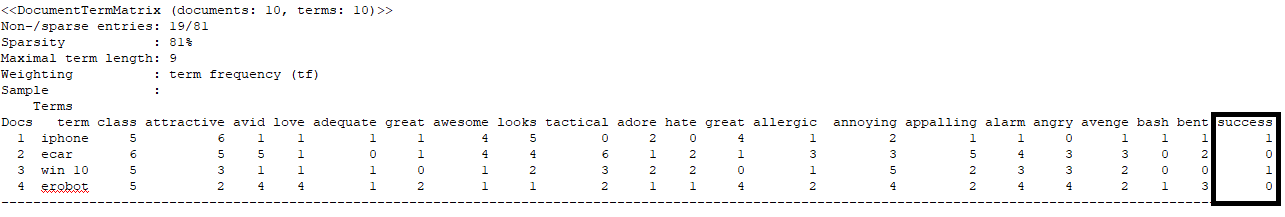
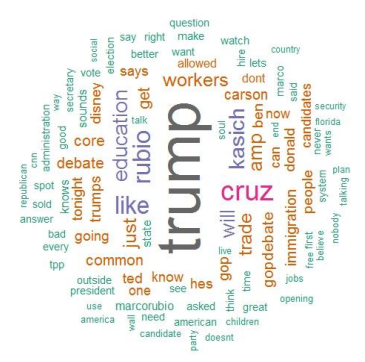


Figure 9: Success of a technology product is internally related to the DTM data row within the sentiment analysis algorithm.

.



Goes to machine learning for sentiment analysis in Literature review..

Co-relational measure will be easier to implement in R

Algorithms...

Instead of combining features, we can add their totals at the end to come up with one grand total sentiment.

train your data with test data.. success again.. again.. split into two.. by identifying the patterns of success clear enouph pattern .. able to converge into ..

words or phrasess or sentiments on success sides that can lead us to judge success based upon them..

### Sample Heading (Third Level). Only two levels of headings should be numbered. Lower level headings remain unnumbered; they are formatted as run-in headings.

#### Sample Heading (Forth Level). The contribution should contain no more than four levels of headings. The following Table 1 gives a summary of all heading levels.

**Table 1.** Table captions should be placed above the tables.

|  |  |  |
| --- | --- | --- |
| Heading level | Example | Font size and style |
| Title (centered) | **Lecture Notes** | 14 point, bold |
| 1st-level heading | **1 Introduction** | 12 point, bold |
| 2nd-level heading | **2.1 Printing Area** | 10 point, bold |
| 3rd-level heading | **Run-in Heading in Bold.** Text follows | 10 point, bold |
| 4th-level heading | *Lowest Level Heading.* Text follows | 10 point, italic |

Displayed equations are centered and set on a separate line.

*x* + *y* = *z* ()

Please try to avoid rasterized images for line-art diagrams and schemas. Whenever possible, use vector graphics instead (see Fig. 1).

**Fig. 10.** A figure caption is always placed below the illustration. Short captions are centered, while long ones are justified. The macro button chooses the correct format automatically.

For citations of references, we prefer the use of square brackets and consecutive numbers. Citations using labels or the author/year convention are also acceptable. The following bibliography provides a sample reference list with entries for journal articles [1], an LNCS chapter [2], a book [3], proceedings without editors [4], as well as a URL [5].

References

1. FELDMAN, R. 2013. Techniques and applications for sentiment analysis. *Commun. ACM,* 56**,** 82-89.
2. GIACHANOU, A. & CRESTANI, F. 2016. Like It or Not: A Survey of Twitter Sentiment Analysis Methods. *ACM Computing Surveys,* 49**,** 28-28:41.
3. STEPHEN, H. & REBECCA SCOTT (2017) ‘Developing an Approach to Harvesting, Cleaning, and Analyzing Data from Twitter Using R’, Information Systems Education Journal (ISEDJ), v15 n3 (May 2017), P42-54.
4. DANNEMAN, N., & HEIMANN, R. (2014). Social media mining with R. Packt Publishing Ltd.
5. DE VRIES, A. (2016) Text Analysis 101: Sentiment Analysis in Tableau & R. The Information Lab. Retrieved 30 May 2016, from <http://www.theinformationlab.co.uk/2016/03/02/text-analysis-101-sentiment-analysis-in-tableau-r/>.
6. BREEN, J. (2011). Mining Twitter for Airline Consumer Sentiment. Inside-r.org. Retrieved 30 May 2016, from <http://www.inside-r.org/howto/mining-twitter-airline-consumer-sentiment>.
7. Bryl, S. (2014). Twitter sentiment analysis with R. AnalyzeCore.com. Retrieved 30 May 2016, from http://analyzecore.com/2014/04/28/twitter-sentiment-analysis/.
8. Author, F., Author, S.: Title of a proceedings paper. In: Editor, F., Editor, S. (eds.) CONFERENCE 2016, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2016).
9. Author, F., Author, S., Author, T.: Book title. 2nd edn. Publisher, Location (1999).
10. Author, F.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010).
11. LNCS Homepage, <http://www.springer.com/lncs>, last accessed 2016/11/21.