e-ISSN: 2349-4689

Research Result

Using an Open Source Tool Analyzing the Sentiment of Events on Twitter

Dr. Meghna Utmal

Prof. & Head MCA Department, GGITS, Jabalpur,(M.P.), INDIA

ABSTRACT

The popularity of the Internet has exploded in recent years. The field of study that analyses people's opinions, sentiments, assessments, attitudes, and emotions from written language is known as sentiment analysis and opinion mining. Users contribute a large amount of user-generated material. The rise of social media platforms such as reviews, forum discussions, blogs, microblogs, Twitter, and social networks has increased the importance of sentiment analysis. Analyzing and summarizing user-generated content is tough. The majority of users express their opinions and thoughts on blogs, social media sites, and e-commerce sites, among other places. As a result, these contents are critical for individuals, businesses, and research projects to make decisions. Sentiment analysis and opinion mining research, which falls under the Natural Language Processing umbrella, is a hot research field in this regard. R is a free and open-source programming language. In this paper, we examine the many ways to sentiment analysis and opinion mining for various datasets in order to determine which strategy is optimal for which dataset, hence assisting researchers in selecting an approach and dataset. In the proposed work, we used the R tool to gather tweets from different events on Twitter, pre-processed them, and calculated sentiment scores from them. We created a Wordcloud for each event, highlighting the most frequently used terms from tweets, as well as calculating the amount of positive, negative, and neutral tweets for each event.

KEYWORDS

Sentiment Analysis and Opinion mining, Natural language processing, SentiWordNet, General Inquirer

1. INTRODUCTION

Sentiment analysis (also known as opinion mining) is the technique of identifying and extracting subjective information from source materials using natural language processing, text analysis, and computational linguistics. Sentiment analysis is commonly used in reviews and social media for a variety of purposes, including marketing and customer service. The goal of sentiment analysis, also known as opinion mining, is to discern a speaker's or writer's attitude toward a topic or the overall contextual polarity of a document. Sentiment Analysis' main goal is to classify the polarity of a given word, phrase, sentence, document, or another object. We can also find other emotional states such as "Happy," "Sad," "Angry," "Fear," "Surprise," and so on. When it comes to business intelligence, sentiment analysis is applied in a variety of ways. In marketing, for example, it can be used to assess the performance of an ad campaign or a new product launch, establish which versions of a product or service are popular, and even detect which demographics prefer or dislike certain features.

Sentiment analysis presents a variety of issues. One is a subjective term; what is considered positive in one context may be regarded negative in another. A second issue is that people do not always communicate their opinions in the same way. The majority of classical text processing is based on the difference between two words. However, in Sentiment Analysis, "the picture was nice" is not the same as "the picture was not nice."

People's viewpoints can be contradictory. People make both positive and negative comments, which can be managed by looking at statements one at a time. People communicate their opinions through many informal mediums such as twitter, blogs, Facebook, Amazon, and others, which are human readable yet tough for machines to interpret. Because of the lack of context, even other individuals have difficulties understanding what someone believed based on a brief bit of text. "That movie was as good as its last movie," for example, is entirely contingent on how the individual delivering the judgement felt about the previous model. English and Chinese are the two most researched languages in sentiment analysis. There are now very few scholars who conduct research in languages other than English, such as Arabic, Italian, and Thai. This study covers the years 2004 and onwards, and includes several approaches on various datasets such as movie reviews and product reviews.

Sentiment Analysis, according to Bing Lui, comprises three stages. The three levels of Sentiment Analysis are listed below.

1.1 Levels of Sentiment Analysis

a. *Document level Sentiment Analysis*

This level of Sentiment Analysis examines the entire document and classifies whether it displays a positive or negative sentiment [1], [2]. Only one product has been reviewed in a single document. And your job is to find out what people think about that product. As a result, this process is sometimes referred to as document-level sentiment classification. The expressed opinion is about a single entity at this level. When there is a document with several product reviews, this does not apply.

b. Sentence Level Sentiment Analysis

At this level, the aim is to examine each statement and evaluate whether it conveys a good, negative, or neutral viewpoint. Subjectivity Classification [3], which divides objective and subjective statements, was closely related to this level. Subjective sentences express subjective information about sentences, whereas objective sentences represent factual information. Opinions can be found in many objective sentences. Sentence Level Sentiment Analysis is the term for this endeavour.

c. Aspect level Sentiment Analysis

Feature level (feature-based opinion mining and summarization) Sentiment Analysis was previously known as Aspect Level Sentiment Analysis [4]. Sentiment Analysis at the document and sentence level does not reveal what individuals liked or disliked. It allows for more fine-grained analysis. Instead of looking at papers, paragraphs, sentences, clauses, or phrases, this level looks at the opinion itself. This level considers the entity, its aspect, the aspect's viewpoint, the opinion holder, and the passage of time. Because of these criteria, this level can determine what people genuinely like, i.e., which features of the product are most popular among customers and at what times. This assignment is both more intriguing and challenging.

2. DATA SOURCE

The user's opinion is crucial in improving service quality. Individuals or businesses can use these opinions to learn about popular events or items in the online realm. On many social media and e-commerce websites, users express their opinions. Twitter, Facebook, Amazon, and Flipkart are just a few examples. Many of these blogs include reviews of various products, events, and topics.

a. Blogs

A blog is a website where someone expresses their own thoughts, actions, and experiences. Blogging is one of the most useful methods for businesses to engage with customers and, in turn, make their lives easier. With the rise of user-generated content on the internet, the number of blogging pages is fast expanding. Blog sites are the most common way to communicate one's personal viewpoint. Bloggers use blogs to document their daily lives and convey their thoughts, feelings, and emotions [5]. Many studies [6], [7], [8] have employed blogs as a source of user opinion in sentiment analysis and opinion mining research.

b. Data Set

In the subject of movie review categorization, the movie review dataset (http://www.cs.cornell.edu/People/pabo/movie-review-data) is commonly utilized. Researchers use the website (http://www.cs.jhu.edu/mdredze/datasets/sentiment) for the multi-domain sentiment (MDS) dataset. The Amazon.com multi-domain sentiment analysis dataset contains four different categories of product reviews: books, DVDs, electronics, and kitchen appliances, each containing 1000 positive and 1000 negative ratings.

Another review dataset is available at (http://www.cs.uic.edu/liub/FBS/CustomerReviewData.zip). [9],[10],[11], [12],[13],[14],[15], [16], [17], [18],

There are 330 review texts in the Micropinion Generation Dataset (CNET). The reviews include a wide range of devices, including televisions, cell phones, and GPS. This dataset was used to summarize opinions in text form. This dataset can be found on Kavita Ganesan's website (http://kavita-ganesan.com/micropinion-generation). The MovieLens Dataset, created by the GroupLens Research Project at the University of California, contains 100,000 ratings (1-5) from 943 individuals on 1682 films.

Minnesota. At least 20 movies have been rated by each user. The dataset can be found at http://grouplens.org/datasets/movielens/. There are 52077 reviews in the Restaurant Review Dataset. The fields on the website (http://www.cs.cmu.edu/mehrbod/RR/) comprise rating information, review counts, percent, and cuisine type. The SNAP Review Dataset includes 34,686,770 Amazon customer reviews from a total of 6,643,669 people. This dataset was originally utilized for recommendation algorithms and is now available on the Stanford Snapshot website (http://snap.stanford.edu/data/web-Amazon.html).

c. Review sites

Any user's opinion can play a significant role in making a purchasing decision. On the Internet, there is a growing quantity of unstructured data in the form of user-generated evaluations. Reviews of products or films are based on unstructured opinions expressed in a variety of ways. The majority of reviewer data for research is gathered from various social media or e-commerce websites, such as www.amazon.com (product reviews), www.yelp.com (restaurant reviews), www.CNET download.com (product reviews), and www.reviewcentre.com, which offers millions of consumer product reviews. Aside from this, there are professional review sites like www.dpreview.com and www.zdnet.com, as well as consumer opinion sites like www.consumerreview.com, www.epinions.com, and www.bizrate.com that cover a wide range of topics and items. [19],[9],[20],[21].

d. Micro-blogging

Users create status messages called "tweets" on Twitter, a popular microblogging site. These tweets contain the user's ideas and opinions on various issues. Tweets are sometimes used to categorize people's feelings on a particular issue or a product review. On Twitter, there are several mobile reviews.

3. RELATED WORK

The study in [1] is about sentiment analysis at the document level on a movie review dataset. They used a variety of machine learning techniques, including (Naive

Bayes, maximum entropy classification, and support vector machines). The IMDB website provided the data for this study. They only included reviews in which the author's rating was stated in stars or a numerical number. Unigrams, unigrams+bigrams, bigrams, unigrams+POS, adjectives, top 2633 unigrams, and unigrams+position are among the features they have chosen. In comparison to the human-generated baselines, the results were quite good. In comparison to other classifiers, SVM is the most effective.

[2] shows how to classify a review as recommended or not recommended using a basic unsupervised learning algorithm. Average The reviews are classified based on the semantic orientation of phrases in the review that contain adjectives or adverbs. Sentiment Analysis at the document level is the focus of this article. The semantic orientation of a phrase and a word is calculated using point wise Mutual Information (PMI). They used Epinions reviews for many domains (Automobiles, Banks, Movies, and Travel Destinations). They found that accuracy ranged from 84% for automotive reviews to 66% for movie reviews.

They created an algorithm for determining semantic orientation in [22]. Rather than phrases containing adjectives or adverbs, this algorithm is built for isolated adjectives. To infer the semantic orientation of adjectives from conjunction restrictions, they employed a four-step supervised learning technique. They found that depending on the amount of training data, accuracy for classification of adjectives ranged from 78 percent to 92 percent.

They created a system that creates sentiment timelines in [23]. It monitors online movie comments and generates a plot with the quantity of good and negative sentiment messages over time. They employed movie-specific domain lexicons. It is used instead of a lexicon created by hand. This work is used in automatic review grading, advertising campaign tracking, public opinion tracking for politicians, financial opinion tracking for stock traders, and trend analysis for entertainment and technological trends.

It is related with subjectivity tagging in [24]. They assessed factual material in an objective manner. The results of a method for grouping words according to distributional similarity are used in this paper to identify strong signals of subjectivity. Characteristics based on both similarity clusters and lexical semantic features offer higher precision than features based on either alone, according to 10-fold cross validation results.

[25] used a multi-way measure to categorise a document's polarity and enlarged the task of classifying a movie review as positive or negative to predicting star ratings on a 3 or 4 star scale. They looked at how well humans performed at the job. Meta algorithm is used, which is based on metric labelling. When we use a novel similarity measure relevant to the task, this Meta method outperforms both multi-class and regression versions of SVMs. They used the Dataset of movie reviews.

They found out what people thought about several qualities or attributes of entities, such as a cell phone, a digital camera, or a bank, in [4]. They examine various aspects of various entities. This article completed three tasks: (1) mining product attributes that customers have mentioned, (2) recognizing opinion statements in each review and determining if each opinion sentence is good or negative, and (3) summarizing the findings. They used POS tagging, Frequent Features Identification, Opinion Words Extraction, and Opinion Word Orientation Identification, Infrequent Feature Identification, Predicting the Orientations of Opinion Sentences, and Summary Generation in their suggested technique. They obtained the following results for opinion sentence extraction and sentence orientation prediction: for five goods, the average precision and recall are 0.64 percent and 0.69 percent, respectively. Sentence Orientation accuracy was 0.84 percent for them.

3.1 Sentiment Analysis Process

Supervised Learning Approach and Unsupervised Learning Approach are the two basic approaches to sentiment analysis. The following steps are required in the sentiment analysis process.

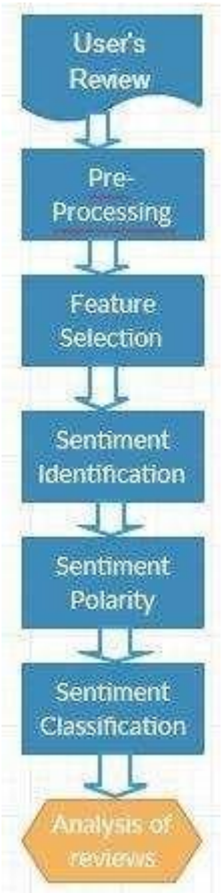

Figure 1. Sentiment Analysis Process

a. *Collection of User's Reviews*

The Sentiment Analysis Task necessitates the use of reviews. This survey employs a variety of ways for collecting reviews. Product reviews are gathered from various E-commerce sites such as Amazon.com, Epinion.com, and Cnet, among others, while movie reviews are gathered from IMDB and other sites. Structured, semi-structured, and unstructured reviews are all possible. There are open-source frameworks for sentiment analysis research where researchers can acquire their data for research purposes. The R Foundation for Statistical Computing supports R [27], a programming language and software environment for statistical computing and visualization. Crawling the reviews from a social website is a simple procedure after installing the necessary packages and going through the authentication process. We can use the text data we have for pre-processing purposes once we obtain it.

b. Pre-Processing

The incomplete, noisy, and inconsistent data are removed using data pre-processing. Before using data in a feature selection task, it must be pre-processed. The following are some things to complete during pre-processing:

- Removing URLs, Special characters, Numbers, Punctuations etc.
- Removing Stopwords
- Removal of Retweets (in case of twitter dataset)
- Stemming
- Tokenization

c. Feature Selection

The challenging task in sentiment analysis is feature selection from pre-processed text. The basic purpose of feature selection is to reduce the feature space's dimensionality and consequently the computing cost. The overfitting of the learning scheme to the training data will be reduced by feature selection. Different machine learning algorithms were analysed on a movie review dataset with different feature selection strategies in [1]. Typically, features are unigrams, biagrams, and ngrams. In feature selection strategies, POS tagging is used.

d. Sentiment Word Identification

Many applications of sentiment analysis and opinion mining, such as review mining, opinion holder identifying, and review categorization, rely on sentiment word identification. Positive, negative, and neutral words can all be categorized as sentiment words.

e. Sentiment Polarity Identification

The primary goal of SA is to classify the polarity of a text at the document, phrase, or feature level. Positive, Negative, and Neutral polarity are the three types of polarity. Different lexicons, such as Bing Lui sentiment lexicon, SentiWordNet, and others, are used to determine sentiment score, sentiment strength, and other metrics.

f. Sentiment Classification

Sentiment categorization of movie and product review datasets is performed using supervised machine learning techniques such as naive Bayes, SVM, and Maximum Entropy, among others. The accuracy of classification systems is determined on the dataset used. The training dataset is used to train the classification model, which then helps to classify the test data in supervised machine learning approaches.

g. Analysis of Reviews

Finally, it is critical to analyse the results in order to make decisions for both individuals and businesses. In the instance of movie reviews, if the majority of the responses are positive, the user may opt to see the film. In business intelligence, analysis is used.

3.1 Sentiment Classification Approaches

There are two main approaches in sentiment analysis

- Supervised learning
- Unsupervised LearningApproach

3.2.1 Unsupervised Learning Approach

When we have training data, we can use this strategy to categorise sentiment, and it can alleviate the problem of domain dependency and the necessity to decrease the training data. Turney [2] calculated the semantic orientation of the statement using two seed words (bad and great). The relationship of seed words with their phrases was discovered using point-wise mutual information. The average semantic orientation of all such sentences is used to calculate the document's sentiment. At the document level, this method achieved a 66 percent accuracy for the movie review dataset.

A) Lexicon based Methods

Lexicon-based sentiment classification techniques are founded on the idea that the polarity of a piece of text may be determined by the polarity of the words that make it up. [28] Lexical resources are used in this method to map words to a category (Positive, Negative, Neutral) or numeric score produced by the algorithm to determine the overall sentiment of the text. Some lexical resources are included below:

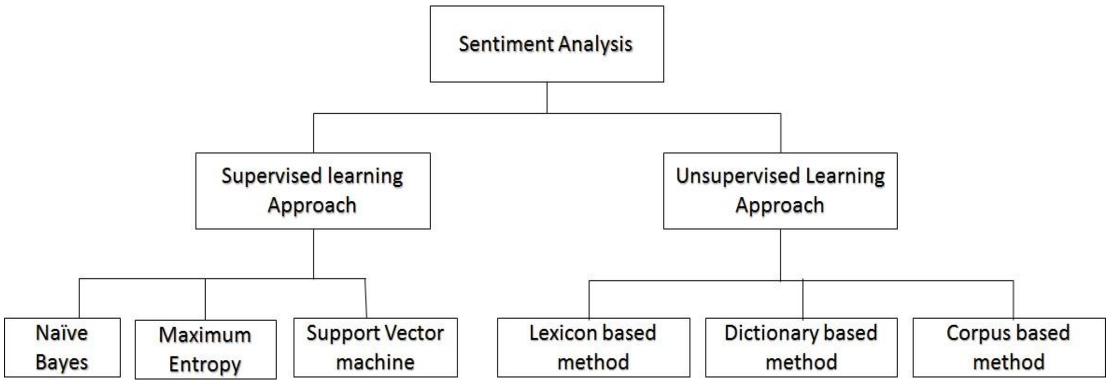

Figure 2: Sentiment Classification Approaches

a. *SentiWordNet*

The lexical resource SentiWordNet [29] is utilised in sentiment analysis applications. For each WordNet synset, it provides an annotation based on three numerical emotion scores (positivity, negativity, and neutrality) [30]. It gives a synset-based sentiment representation, with distinct sentiment ratings for different senses of the same Term. The most promising meaning is determined using the Word Sense Disambiguation (WSD) algorithm.

b. *WordNet-Affect*

The linguistic resource WordNet-Affect [31] is a lexical representation of emotional knowledge. It's a WordNet extension that labels affective-related synsets with A-Labels (affective concepts) (e.g. the term euphoria is labelled with the concept positive-emotion, the noun illness is labelled with physical state, and so on). The mapping is done using a domain-independent hierarchy as a foundation.

c. MPQA

It [32] is connected to subjectivity lexicons and contains lexicons of 8,222 phrases labelled as subjective expressions culled from various sources. This includes a list of words and associated POS-tagging, as well as polarity (positive, negative, and neutral) and intensity labels (strong, weak).

d. SenticNet

This lexical resource [33] is utilised for sentiment analysis at the concept level. It is based on Sentic Computing [34], a multi-disciplinary Sentiment Analysis paradigm. SenticNet can link polarity and affective information to complicated notions such as achieving a goal, celebrating a memorable occasion, and so on. For 14,000 common sense notions, it assigns a sentiment score ranging from -1 to 1.

e. Hu and Liu's lexicon

This English opinion lexicon includes a list of positive and negative words, as well as sentiment words. This list, which contains 2006 positive and 4783 negative sentiment words, was created for the [4] paper.

B) *Dictionary Based Methods*

f. *WordNet*

WordNet [35] is a huge English lexical database. Cognitive synonyms (synsets) are groups of nouns, verbs, adjectives, and adverbs that each communicate a distinct concept. Conceptual-semantic and lexical relationships link synsets together. In the way that it organises words together depending on their meanings, WordNet resembles a thesaurus. The shortest distance between "good" and "bad" is used to determine a word's polarity in [36]. In our experiment, we extract the words from our lexicon that are contained in WordNet for comparison.

g. General Inquirer

The General Inquirer (GI) is a text analysis programme that uses one of the oldest manually generated lexicons available. The GI has been in the works since 1966 as a tool for content analysis, which is a technique used by social scientists, political scientists, and psychologists to objectively identify specific qualities of messages [37]. There are almost 11K terms in the lexicon, which are divided into 183 categories. There are 1,915 words in GI that are labelled Positive and 2,291 words that are labelled Negative. It's useful for determining sentiment qualities of textual data in a variety of studies.

h. **LIWC**

LIWC is text analysis software [38] that can be used to investigate the emotional, cognitive, structural, and process components of text samples. LIWC employs a proprietary vocabulary with about 4,500 words classified into one (or more) of 76 categories, including 905 words in two sentiment-related categories. Positive Emotions (e.g., love, lovely, good, great) account for 406 of the total, whereas Negative Emotions (e.g., hurt, ugly, sad, awful, worse) account for 499.

i. **AFINN**

AFINN [49] is a list of English words with valence ratings ranging from minus five (negative) to plus five (positive) (positive). Finn rup Nielsen manually labelled the words between 2009 and 2011. The file is split by tabs. There are 1468 unique words and phrases on 1480 lines in the first edition AFINN-96, and 2477 unique words and phrases on 1480 lines in the second revised version AFINN-111.

C) Corpus Based Methods

j. Darmstadt Service Review Corpus

It consists of consumer reviews that have been annotated at the sentence and expression levels with opinion-related information. They gave Niek Sanders in [39]: "He has built a Twitter Sentiment Corpus that "consists of 5513 hand-classified tweets."

Table 1. Summary of Some Research Article

Author & Year	Dataset	Features & Techniques	Results			
			Accuracy	Precision	Recall	F1 Score
Kaiquan Xu (2011)[40]	Amazon Reviews	linguistic features, Multiclass,SVM,CRF	61.38	61.96	93.49	74.26
Long Sheng (2011) [41]	Movie Reviews	PMI,SO-A,SO-LSA, Back-Propagation neural network	64	60	98	75
Hanhoon Kang (2012)[42]	RestaurantReviews	Unigram, Bigram, Improved Naïve Bayes I and NB	81.4	--	--	--

Lin Y, Zhang J (2012)[43]	Product Reviews	PMI, Lexicons, SVM,Unsupervised approach	--	82.62	85.26	83.92
Federico Neri (2012)[44]	Facebook postsRai – the Italian public broadcastingservice	words, phrases, sentences, Bayesian method, K-Means	--	93	87	--
Prashant Raina (2013)[45]	News Articles.	MPQA corpus, semantic parser, Sentic Computing, ConceptNet and SenticNet	71	Pos-46.3 Neg-61.6 Neu-90.9	Pos-79.3 Neg- 70.5 Neu-69.8	Pos-58.5 Neg- 65.8 Neu-79
Seyed-AliBahrainian (2013)[46]	Twitter dataset	Unigram feature,SVM, NB, MaxEnt Hybrid Approach	89.78	--	--	--
Emitza Guzman (2014)[47]	US App Store and Google Play.	Topics, N-top features, LDA,Topic Model, lexical sentiment analysis	--	91	73	--
Emitza Guzman (2014)[47]	US App Store and Google Play.	Topics, N-top features, LDA,Topic Model, lexical sentiment analysis	--	91	73	--

4. PROPOSED METHODOLOGY

Figure 3 depicts the methods used to mine the Twitter dataset, with the following phases being particularly significant.

a. Data Collection

The R Foundation for Statistical Computing [27] supports R, a computer language and software environment for statistical computing and graphics. R contains a number of packages that may be used to obtain social media data, such as Twitter and Facebook, as well as packages for text/numerical data pre-processing and visualization, such as tm, stringr, and ggplot. The TwitteR package is used to retrieve tweets from Twitter's API.

b. Pre-Processing

Pre-processing is critical in data mining since it influences the accuracy of the results. The tm package is used in pre-processing to extract text from tweets and to conduct additional pre-processing tasks such as eliminating stopwords, spaces, punctuation, and URLs, as well as stemming (get the root of the words). Unstructured data is represented in Term-Document Matrix after this stage.

c. Data Analysis

We received TDM from the previous stage, and using TDM, it's simple to uncover association rules, more frequent phrases, and execute sentiment analysis using the lexicon-based technique, which employs a collection of positive and negative words. Bing Lui lexicons were used to determine the score of each tweet using the Scoring Function [4].

d. Visualization of Result

The R package Wordcloud aids in the representation of a word cloud that displays the most frequently occurring terms from text data. The frequency of words in the customer's tweets is displayed using this manner.

5. EXPERIMENTS AND RESULT

The trials included a variety of hashtag (#) tweets. The TwitteR package is used to access Twitter's live tweets. Using keyword search queries, the ROAuth, TwitteR, Rcurl, and other tools enable authentication and access to twitter messages [27].

Table 2. Twitter Dataset Used for Experimental Work

Topic Name	Positive tweets	Negative Tweets	Neutral Tweets
#Budget2016	3507	1555	4938
#RailBudget2016	3703	969	5328
#Freedom251	7564	581	1855
#MakeInIndia	3052	1012	5936
#Oscars2016	5539	694	3767
#startup	3907	799	5294
#InternationalWomensDay	15600	1724	7676
#AsiaCupT20Final	9742	2143	8115
#IndvsPak	16590	2606	10804
#ProKabaddi	3328	224	2204

It employed Unsupervised Learning techniques, which included the usage of lexicons. Different topics have been accessed on Twitter over time. In this methodology, lexicon-based methods have been applied, which is related to sentiment analysis and opinion mining. Figure 1 depicts the #MakeInIndia topic's Wordcloud. The opinion lexicons and scoring function are required for determining the score of each tweet. The scoring function works as follows:

Sentiment Score = Σ positive words – Σ Negative words (Eq. 1)

Positive polarity occurs when the number of positive words exceeds the number of negative words, resulting in a positive score. Negative polarity occurs when the quantity of negative words exceeds the number of positive words, resulting in a negative score. If the amount of positive and negative words in the text is equal or if there are no opinion words in the text, the score will be neutral.

Table 3 shows the distribution of positive, negative, and neutral tweets for each topic as a result of sentiment analysis. Only one topic has been plotted in a wordcloud, as seen in Figure 3. And Figure 4 depicts the Table 3 chart, which shows how polarity fluctuates with different events in the Twitter dataset.

Table 3 Result of Polarity of Twitter Dataset

Topic name	No. of Tweets	Period
#Budget2016	10000	29 Feb 2016
#RailBudget2016	10000	25 Feb 2016
#Freedom251	10000	18 Feb 2016
#MakeInIndia	10000	23 Feb – 29 Feb 2016
#Oscars2016	10000	29 Feb 2016
#startup	10000	28 Feb – 29 Feb 2016
#InternationalWomens Day	25000	08 Mar 2016
#AsiaCupT20Final	20000	07 Mar 2016
#IndvsPak	30000	20 Mar – 21 Mar 2016
#ProKabaddi	5756	28 Feb – 06 Mar 2016

Figure 3. Wordcloud of #MakeInIndia

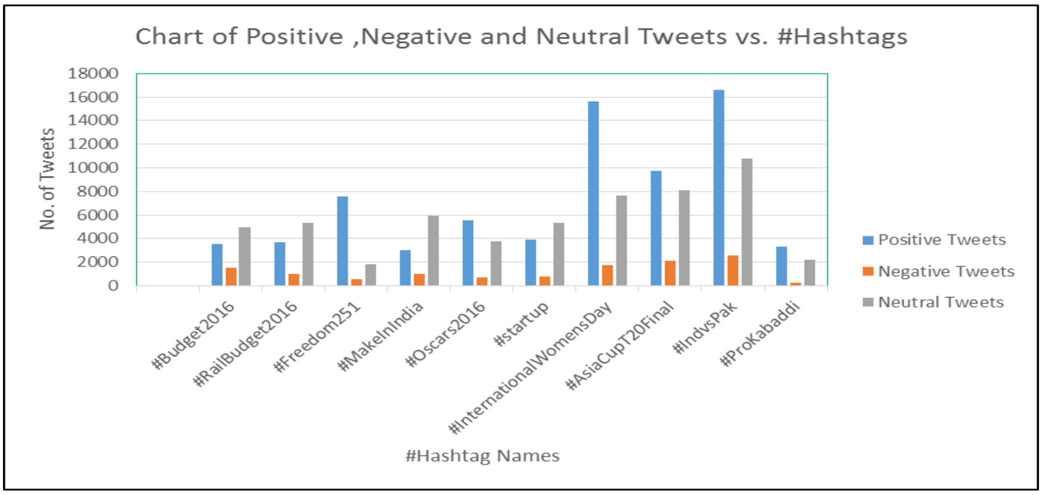

Figure 4. No. of Positive, Negative and Negative Tweets vs. #Hashtag

6. CONCLUSION AND FUTURE WORK

We used the Twitter API with the open-source R tool in this proposed project. Tweets from Twitter have been collected and fed into the tool's pre-processing activity. Text mining and crawling streaming data from social media sites such as Twitter and Facebook are both done with the R open-source programme. For sentiment analysis and opinion mining, the data from movie reviews was additionally pre-processed in the R tool. Diverse supervised and unsupervised methodologies, as well as different lexicons, dictionaries, and corpus-based methods, are all highly useful in Sentiment Analysis, as detailed in section 3.

Movie reviews, product reviews, Epinions datasets, and other datasets are all available. This algorithm calculates and counts the amount of good, negative, and neutral tweets for a given #Hashtag, allowing it to anticipate public opinion on a specific event. Individuals and industries can find the public opinion behind an event using the above analysis of distinct #Hashtags tweets for sentiment analysis. The used methodologies and dataset for each research group are listed in a summary table. The sentiment score is calculated using Bing Lui lexicons. Figure 1 depicts a Wordcloud that displays common event words on the plot and highlights those phrases that users have given tress to.

Future work on product review sentiment analysis will focus on identifying elements of the product and their polarity, which will aid consumers in making purchasing decisions on e-commerce sites. Aspect level sentiment analysis provides detailed information on a product or a movie review. For example, the director of a film should know exactly what the audience knows about that film, which is feasible with aspect based sentiment analysis. Hotel owners should know what goods people enjoy from their hotel and what other stuff customers require in hotel reviews, just as they should know what items people like from movies.

REFERENCES

[1]. Pang, Bo; Lee, Lillian; Vaithyanathan, Shivakumar (2002). "Thumbs up? Sentiment Classification using Machine Learning Techniques". Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 79–86.

[2]. Turney, Peter (2002). "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews". Proceedings of the Association for Computational Linguistics. pp. 417–424.

[3]. Wiebe, Janyce, Rebecca F. Bruce, and Thomas P. O'Hara. *Development and use of a gold-standard data set for subjectivity classifications*. In *Proceedings of the Association for Computational Linguistics (ACL- 1999)*. 1999.

[4]. Hu, Minqing and Bing Liu. *Mining and summarizing customer reviews*. In *Proceedings of ACM SIGKDD*

[5]. International Conference on Knowledge Discovery and Data Mining (KDD-2004). 2004.

[6]. Chau, M., & Xu, J. (2007). Mining communities and their relationships in blogs: A study of online hate groups. International Journal of Human – Computer Studies, 65(1), 57–70.

[7]. Martin, J. (2005). Blogging for dollars. Fortune Small Business, 15(10), 88–92.

[8]. Godbole, N.; Srinivasaiah, M. & Skiena, S. (2007), Large-Scale Sentiment Analysis for News and Blogs, in'Proceedings of the International Conference on Weblogs and social media (ICWSM)'.

[9]. H. Tang, S. Tan, and X. Cheng. A Survey on Sentiment Detection of Reviews. Expert Systems with Applications. Elsevier. 2009, doi:10.1016/j.eswa.2009.02.063.

[10]. Hu, and Liu, "Opinion extraction and summarization on the web", AAAI. (2006), pp. 1621-1624.

[11]. Long-Sheng Chen, Cheng-Hsiang Liu, Hui-Ju Chiu, "A neural network based approach for sentiment classification in the blogosphere", Journal of Informetrics 5 (2011) 313–322.

[12]. ZHU Jian , XU Chen, WANG Han-shi, " Sentiment classification using the theory of ANNs", The Journal

[13]. of China Universities of Posts and Telecommunications, July 2010, 17(Suppl.): 58–62 .

[14]. Bo Pang, Lillian Lee.A Sentimental Education: Sentiment Analysis Using Subjectivity. Proceedings of ACL, pp. 271--278, 2004

[15]. Bai, R. Padman, "Markov blankets and meta-heuristic search: Sentiment extraction from unstructured

[16]. text," Lecture Notes in Computer Science, vol. 3932, pp. 167–187, 2006.

[17]. Kennedy and D. Inkpen, "Sentiment classification of movie reviews using contextual valence shifters,"

[18]. Computational Intelligence, vol. 22, pp. 110–125, 2006.

[19]. Zhou, L. and P. Chaovalit (2008), Ontology-supported Polarity Mining, Journal of the American Society for Information Science and Technology, 59(1), 1-13.

[20]. Yulan He,"Learning sentiment classification model from labeled features. "Proceedings of the 19th

[21]. {ACM} Conference on Information and Knowledge Management, {CIKM} 2010, Toronto, Ontario, Canada, October 26-30, 2010.

[22]. Rudy Prabowo, Mike Thelwall, "Sentiment analysis: A combined approach." Journal of Informetrics 3 (2009) 143–157.

[23]. Rui Xia, Chengqing Zong, Shoushan Li, "Ensemble of feature sets and classification algorithms for sentiment classification", Information Sciences 181 (2011) 1138–1152.

[24]. Popescu, A. M., Etzioni, O.: Extracting Product Features and Opinions from Reviews, In Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, British Columbia, 2005, 339–346.

[25]. Qingliang Miao, Qiudan Li, Ruwei Dai, "AMAZING: A sentiment mining and retrieval system", Expert

[26]. Systems with Applications 36 (2009) 7192–7198.

[27]. Xing Fang and Justin Zhan."Sentiment analysis using product review data". Journal of Big Data 2015. DOI:

[28]. 10.1186/s40537-015-0015-2

[29]. Hatzivassiloglou, V., & McKeown, K.R. 1997. Predicting the semantic orientation of adjectives. Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL (pp. 174-181). New Brunswick, NJ: ACL.

[30]. Tong, R.M. 2001. An operational system for detecting and tracking opinions in on-line discussions. Working Notes of the ACM SIGIR 2001 Workshop on Operational Text Classification (pp. 1-6). New York, NY: ACM.

[31]. Wiebe, J.M. 2000. Learning subjective adjectives from corpora. Proceedings of the 17th National Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press.

[32]. Pang, Bo; Lee, Lillian (2005). "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales". Proceedings of the Association for Computational Linguistics (ACL). pp. 115–124.

[33]. Yang Liu, Xiangji Huang, Aijun An, Xiaohui Yu (2007). "ARSA: A Sentiment-Aware Model for

[34]. Predicting Sales Performance Using Blogs" SIGIR"07, July 23–27, 2007, Amsterdam, The Netherlands.

[35]. [27] R Core Team (2015). R: A language and environment for

statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[36]. Cataldo Musto, Giovanni Semeraro, Marco Polignano."A comparison of Lexicon-based approaches for Sentiment Analysis of microblog posts".8th International Workshop on Information Filtering and Retrieval Pisa (Italy) December 10, 2014

[37]. Andrea Esuli Baccianella, Stefano and Fabrizio Sebastiani. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Proceedings of LREC, volume 10, pages 2200-2204,2010.

[38]. George A Miller. WordNet: a lexical database for English. Communications of the ACM, 38(11):39{41, 1995.

[39]. Carlo Strapparava and Alessandro Valitutti. Wordnet affect: an affective extension of wordnet. In LREC,volume 4, pages 1083{1086, 2004.

[40]. Janyce Wiebe, Theresa Wilson, and Claire Cardie. Annotating expressions of opinions and emotions in language. Language resources and evaluation, 39(2-3):165-210, 2005.

[41]. Erik Cambria, Daniel Olsher, and Dheeraj Rajagopal. SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. AAAI, Quebec City, pages 1515{1521, 2014.

[42]. Erik Cambria and Amir Hussain. Sentic computing. Springer, 2012.

[43]. http://wordnet.princeton.edu/

[44]. Kamps, Maarten Marx, Robert J. Mokken and Maarten De Rijke, “Using wordnet to measure semantic orientation of adjectives”, Proceedings of 4th International Conference on Language Resources and Evaluation, pp. 1115-1118, Lisbon, Portugal, 2004.

[45]. Stone, P. J., Dunphy, D. C., Smith, M. S., and Ogilvie, D. M. (1966). The General Inquirer: A Computer Approach to Content Analysis. MIT Press.

[46]. Clayton J. Hutto, Eric Gilbert."VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text.".Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014, Ann Arbor, Michigan, USA, June 1-4, 2014.

[47]. http://www.sananalytics.com/lab/twitter-sentiment/

[48]. Kaiquan Xu , Stephen Shaoyi Liao , Jiexun Li, Yuxia Song, “Mining comparative opinions from customer

[49]. reviews for Competitive Intelligence”, Decision Support Systems 50 (2011) 743–754.

[50]. Long-Sheng Chen, Cheng-Hsiang Liu, Hui-Ju Chiu, “A neural network based approach for sentiment

[51]. classification in the blogosphere”, Journal of Informetrics 5 (2011) 313–322.

[52]. Hanhoon Kang, Seong Joon Yoo , Dongil Han,” Senti-lexicon and improved Naïve Bayes algorithms for

[53]. sentiment analysis of restaurant reviews”. Expert Systems with Applications 39 (2012) 6000–6010.

[54]. Lin Y, Zhang J, Wang X, Zhou A (2012) An information theoretic approach to sentiment polarity classification. In: Proceedings of the 2Nd Joint WICOW/AIRWeb Workshop on Web Quality,

[55]. WebQuality "12. ACM, New York, NY, USA.pp 35–40

[56]. Federico Neri , Carlo Aliprandi , Federico Capeci , Montserrat Cuadros , Tomas By, Sentiment Analysis on Social Media, Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), p.919-926, August 26-29, 2012

[57]. Prashant Raina, "Sentiment Analysis in News Articles Using Sentic Computing", ICDMW, 2013, 2013 IEEE 13th International Conference on Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on Data Mining Workshops (ICDMW) 2013, pp. 959-962, doi:10.1109/ICDMW.2013.27

[58]. Seyed-Ali Bahrainian, Andreas Dengel, "Sentiment Analysis and Summarization of Twitter Data", CSE, 2013, 2013 IEEE 16th International Conference on Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference on Computational Science and Engineering (CSE) 2013, pp. 227-234, doi:10.1109/CSE.2013.44

[59]. Emitza Guzman, Walid Maalej, "How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews", RE, 2014, 2014 IEEE 22nd International Requirements Engineering Conference (RE), 2014 IEEE 22nd International Requirements Engineering Conference (RE) 2014, pp. 153-162, doi:10.1109/RE.2014.6912257

[60]. Geetika Gautam and Divakar yadav, "Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis", 978-1-4799-5172-7/, pp.437-442 2014 IEEE.

[61]. Lars Kai Hansen, Adam Arvidsson, Finn Årup Nielsen, Elanor Colleoni,Michael Etter, "Good Friends, Bad News - Affect and Virality inTwitter", The 2011 International Workshop on Social Computing,Network, and Services (SocialComNet 2011).