



SENTIMENT ANALYSIS ON TWEETS USING MACHINE LEARNING TECHNIQUES

¹Shreelaxmi Kulkarni, ²Dr. Kiran K. Tangod

²Associate Professor

¹Department of Information Science & Engineering,
¹Gogte Institute of Technology, Belagavi, India

Abstract: . In recent days, invention of new platforms in social media has given lot of boost to the business development. In the business process social media is playing an important role as a deciding factor for success or failure of a business in a growing economy of the country. One such platform which helps people to understand and gauge the business prospectus is twitter. In this paper we are addressing the problem of sentiment analysis in twitter; which mainly deals with classifying tweets according to the sentiment expressed in them: positive, negative or neutral. Twitter is an online micro-blogging and social-networking platform which allows users to write short status updates of maximum length 140 characters. It is a rapidly expanding service with over 200 million registered users - out of which 100 million are active users and half of them log on twitter on a daily basis - generating nearly 250 million tweets per day. Due to this large amount of usage we hope to achieve a reflection of public sentiment by analyzing the sentiments expressed in the tweets. Analyzing the public sentiment is important for many applications such as firms trying to find out the response of their products in the market, predicting political elections and predicting socioeconomic phenomena like stock exchange. The aim here is to develop a functional classifier for accurate and automatic sentiment classification of an unknown tweet stream.

Index Terms – Artificial Intelligence, Sentiment Analysis, Classification, Data Mining, Lexical analysis.

I. INTRODUCTION

Introduction

Sentiment analysis is an area of research which is attracting young researchers, it mainly deals with processing of user estimations about the product, process, place or a political situation and it also covers enormous topics on which one can provides reviews or finding in real world scenarios.

Analysing sentiments of tweets comes under the domain of “Pattern Classification” and “Data Mining”. Both of these terms are very closely related and intertwined, and they can be formally defined as the process of discovering “useful” patterns in large set of data, either automatically (unsupervised) or semi automatically (supervised).

It would heavily rely on techniques of “Natural Language Processing” in extracting significant patterns and features from the large data set of tweets and on “Machine Learning” techniques for accurately classifying individual unlabelled data samples (tweets) according to whichever pattern model best describes them. The features that can be used for modelling patterns and classification can be divided into two main groups: formal language based and informal blogging based. Language based features are those that deal with formal linguistics and include prior sentiment polarity of individual words and phrases, and parts of speech tagging of the sentence. Prior sentiment polarity means that some words and phrases have a natural innate tendency for expressing particular and specific sentiments in general. For example: the word “excellent” has a strong positive connotation while the word “evil” possesses a strong negative connotation. So, whenever a word with positive connotation is used in a sentence, chances are that the entire sentence would be expressing a positive sentiment. Parts of Speech tagging, on the other hand, is a syntactical approach to the problem. It means to automatically identify which part of speech each individual word of a sentence belongs to: noun, pronoun, adverb, adjective, verb, interjection, etc. Patterns can be extracted from analysing the frequency distribution of these parts of speech (either individually or collectively with some other part of speech) in a particular class of labelled tweets. Twitter based features are more informal and relate with how people express themselves on online social platforms and compress their sentiments in the limited space of 140 characters offered by twitter. They include twitter hash tags, re-tweets, word capitalization, word Project Thesis Report 11 lengthening, question marks, presence of URL in tweets, exclamation marks, internet emoticons and internet shorthand/slang. Classification techniques can also be divided into two categories: Supervised vs. unsupervised and non-adaptive vs. adaptive/reinforcement techniques. Supervised approach is when we have pre - labelled data samples available and we use them to train our classifier. Training the classifier means to use the pre-labelled to extract features that best model the patterns and differences between each of the individual classes, and then classifying an unlabelled data sample according to whichever pattern best describes it. For example: if we come up with a highly simplified model that neutral tweets contain 0.3 exclamation

marks per tweet on average while sentiment-bearing tweets contain 0.8, and if the tweet we have to classify does contain 1 exclamation mark then (ignoring all other possible features) the tweet would be classified as subjective, since 1 exclamation mark is closer to the model of 0.8 exclamation marks. Unsupervised classification is when we do not have any labelled data for training. In addition to this Adaptive classification technique deal with feedback from the environment. In our case feedback from the environment can be in form of a human telling the classifier whether it has done a good or poor job in classifying a particular tweet and the classifier needs to learn from this feedback. There are two further types of adaptive techniques: Passive and active. Passive techniques are the ones which use the feedback only to learn about the environment (in this case this could mean improving our models for tweets belonging to each of the three classes) but not using this improved learning in our current classification algorithm, while the active approach continuously keeps changing its classification algorithm according to what it learns at real-time.

Literature Survey

Limitations of Prior Art

Sentiment analysis of in the domain of micro-blogging is a relatively new research topic so there is still a lot of room for further research in this area. Decent amount of related prior work has been done on sentiment analysis of user reviews [x], documents, web blogs/articles and general phrase level sentiment analysis. These differ from twitter mainly because of the limit of 140 characters per tweet which forces the user to express opinion compressed in very short text. The best results reached in sentiment classification use supervised learning techniques such as Naive Bayes and Support Vector Machines, but the manual labelling required for the supervised approach is very expensive. Some work has been done on unsupervised and semi-supervised approaches, and there is a lot of room of improvement. Various researchers testing new features and classification techniques often just compare their results to base-line performance. There is a need of proper and formal comparisons between these results arrived through different features and classification techniques in order to select the best features and most efficient classification techniques for particular applications.

Related Work

The bag-of-words model is one of the most widely used feature model for almost all text classification tasks due to its simplicity coupled with good performance. The model represents the text to be classified as a bag or collection of individual words with no link or dependence of one word with the other, i.e. it completely disregards grammar and order of words within the text. This model is also very popular in Project Thesis Report 14 sentiment analysis and has been used by various researchers. The simplest way to incorporate this model in our classifier is by using unigrams as features. Generally speaking, n-grams is a contiguous sequence of "n" words in our text, which is completely independent of any other words or grams in the text. So, unigrams are just a collection of individual words in the text to be classified, and we assume that the probability of occurrence of one word will not be affected by the presence or absence of any other word in the text. This is a very simplifying assumption but it has been shown to provide rather good performance. One simple way to use unigrams as features is to assign them with a certain prior polarity, and take the average of the overall polarity of the text, where the overall polarity of the text could simply be calculated by summing the prior polarities of individual unigrams. Prior polarity of the word would be positive if the word is generally used as an indication of positivity, for example the word "sweet"; while it would be negative if the word is generally associated with negative connotations, for example "evil". There can also be degrees of polarity in the model, which means how much indicative is that word for that particular class. A word like "awesome" would probably have strong subjective polarity along with positivity, while the word "decent" would although have positive prior polarity but probably with weak subjectivity.

There are three ways of using prior polarity of words as features. The simpler un-supervised approach is to use publicly available online lexicons/dictionaries which map a word to its prior polarity. The Multi-Perspective-Question-Answering (MPQA) is an online resource with such a subjectivity lexicon which maps a total of 4,850 words according to whether they are "positive" or "negative" and whether they have "strong" or "weak" subjectivity. The Sent WordNet 3.0 is another such resource which gives probability of each word belonging to positive, negative and neutral classes. The second approach is to construct a custom prior polarity dictionary from our training data according to the occurrence of each word in each particular class. For example, if a certain word is occurring more often in the positive labelled phrases in our training dataset (as compared to other classes) then we can calculate the Project Thesis Report probability of that word belonging to positive class to be higher than the probability of occurring in any other class. This approach has been shown to give better performance, since the prior polarity of words is more suited and fitted to a particular type of text and is not very general like in the former approach. However, the latter is a supervised approach because the training data has to be labelled in the appropriate classes before it is possible to calculate the relative occurrence of a word in each of the class. Kouloumpis et al. noted a decrease in performance by using the lexicon word features along with custom n-gram word features constructed from the training data, as opposed to when the n-grams were used alone.

The third approach is a middle ground between the above two approaches. In this approach we construct our own polarity lexicon but not necessarily from our training data, so we don't need to have labelled training data. One way of doing this as proposed by Turney et al. is to calculate the prior semantic orientation (polarity) of a word or phrase by calculating its mutual information with the word "excellent" and subtracting the result with the mutual information of that word or phrase with the word "poor". They used the number of result hit counts from online search engines of a relevant query to compute the mutual information. The final formula they used is as follows:

$$\text{Polarity(phrase)} = \frac{\log_2 \text{hits (phrase NEAR "excellent")}. \text{hits("poor")}}{\text{hits (phrase NEAR "poor")}. \text{hits ("excellent")}}$$

Where hits (phrase NEAR “excellent”) means the number, documents returned by the search engine in which the phrase (whose polarity is to be calculated) and word “excellent” are co-occurring. While hits (“excellent”) means the number of documents returned which contain the word “excellent”. Prabowo et al. have gone ahead with this idea and used a seed of 120 positive words and 120 negatives to perform the internet searches. So, the overall semantic orientation of the word under consideration can be found by calculating the closeness of that word with each one of the seed words and Project Thesis Report taking an average of it. Another graphical way of calculating polarity of adjectives has been discussed by Hatzivassiloglou et al. The process involves first identifying all conjunctions of adjectives from the corpus and using a supervised algorithm to mark every pair of adjectives as belonging to the same semantic orientation or different. A graph is constructed in which the nodes are the adjectives and links indicate same or different semantic orientation. Finally, a clustering algorithm is applied which divides the graph into two subsets such that nodes within a subset mainly contain links of same orientation and links between the two subsets mainly contain links of different orientation. One of the subsets would contain positive adjectives and the other would contain negative.

Objectives:

- To analyse the sentiment of twitter data feeds
- Read the opinions and user response to different products, brands or topics through user feeds on social media platform twitter.
- To provide top updated news feeds along with the sentiment
- Display the proportionality of the positive, negative and neutral news according to the most common user feeds.
- It allows to track the altitude of user opinion on any particular topic on twitter platform.

Relevance in Industry:

Going through millions of tweets manually would take far too much time. You’d miss out on valuable feedback that could help you instantly improve a customer’s experience with the latest feature (*bug issues, user experience*). By performing sentiment analysis with machine learning, you can quickly understand the tone and context of social mentions on Twitter.

The overall benefits of Twitter sentiment analysis include:

- **Scalability:** Analyse hundreds or thousands of tweets mentioning your brand and automate manual tasks. Easily scale sentiment analysis tools as your data grows and gain valuable insights on the go.
- **Real-Time Analysis:** Twitter sentiment analysis is essential for monitoring sudden shifts in customer moods, detecting if complaints are on the rise, and for taking action before problems escalate. With sentiment analysis, monitor brand mentions on Twitter in real-time and gain actionable insights.
- **Consistent Criteria:** Avoid inconsistencies that stem from human error. Customer reps won’t always agree on which tag to use for each piece of data, so you may end up with inaccurate results. Instead, machine learning models perform sentiment analysis using one set of rules, so you can ensure all your Twitter data is tagged consistently.

Twitter sentiment analysis provides many exciting opportunities. Being able to analyse tweets in real-time, and determine the sentiment that underlies each message, adds a new dimension to social media monitoring. Here are some of the most common business applications of Twitter sentiment analysis.

Social Media Monitoring

- Online reputation is one of the most precious assets for brands. A bad review on social media can be costly to a company if it’s not handled effectively and swiftly.
- Twitter sentiment analysis allows you to keep track of what’s being said about your product or service on social media, and can help you detect angry customers or negative mentions before they turn into a major crisis.
- At the same time, Twitter sentiment analysis can provide interesting insights. What do customers love about your brand? What aspects get the most negative mentions? This tweet, for example, indicates that fast shipping is one of the most valued aspects for this Amazon customer:



Loope

@upupandaway43

Follow

I love amazon prime. I ordered something today, and I'm already getting it tomorrow



5:31 pm - 22 May 2019

- Aspect-based sentiment analysis with Twitter can show you which aspects of your business need to be improved and what makes you stand out among your competitors.

Customer Service

- Twitter has become an essential channel for customer service. In fact, a growing number of companies have specific teams in charge of delivering customer support via this social media platform. Prompt replies are key since 60% of the customers that complain on social media expect a response within one hour.
- But how can you evaluate the performance of your customer support on Twitter? Twitter sentiment analysis allows you to track and analyze all the interactions between your brand and your customers. This can be very useful to analyze customer satisfaction based on the type of feedback you receive.
- This tweet, for example, shows a disappointed customer after an interaction with Southwest Airlines' customer support team:



Southwest Airlines @SouthwestAir · 46m

Replying to @meowyam

Please know that it's never our intention to disappoint, and we regret any inconvenience, Maryam. I assure you we are working diligently to get your family on their way as soon as possible. -Taylor



maryam @meowyam · 13m

nobody really intends to disappoint and this isn't "any" inconvenience. the fact of the matter is that now they're going to miss my graduation because they weren't notified in time. super gutted because i always take southwest and expect better, definitely not this



Market Research

- Twitter is a major source of consumer insight. In fact, people use it to express all sorts of feelings, observations, beliefs, and opinions about a variety of topics.
- You can use Twitter sentiment analysis to track specific keywords and topics to detect customer trends and interests. Understanding what things potential customers like, what their behaviours are, and how this change over time is essential if you are planning to launch a new product.
- Twitter sentiment analysis can also help you stay one step ahead of your competition. By identifying competitors' pain points, you can focus on these areas when promoting your business.

Brand Monitoring

- Whether you are launching a new feature on your platform, a site redesign, or a new marketing campaign, you may want to track customer reactions on Twitter. Taking action and making changes or improvements in real-time will help maintain customer loyalty.



12:05 pm - 27 Sep 2018

Political Campaigns

A huge part of Twitter conversation revolves around news and politics. That makes it an excellent place to measure public opinion, especially during election campaigns. Twitter Sentiment Analysis can provide interesting insights on how people feel about a specific candidate (and you could even track sentiment over time to see how it evolves).

Approaches and Techniques for Sentiment Analysis**Approaches**

Depending on the task at hand and perspective of the person doing the sentiment analysis, the approach can be discourse-driven, relationship-driven, language- model driven, or keyword-driven.

1. Knowledge-based approach

The main task in this approach is the construction of word lexicons that indicate positive class or negative class. The sentiment values of the words in the lexicon are determined prior to the sentiment analysis work. Lexicons can be created in different ways. It can be created by starting with some seed words and then using some linguistic heuristics to add more words to them, or starting with some seed words and adding to these seed words other words based on frequency in a text. SENTIWORDNET 3.0 is a publicly available lexical resource explicitly devised for supporting sentiment classification and opinion mining applications.

2. Relationship-based approach

In this approach the different relationships between features and components is analysed for sentiment classification task. Such relationships may be relationships between different participants, relationships between product features. For example, if one wants to know the sentiment of customers about a product brand, one may compute it as a function of the sentiments on different features or components of it.

3. Language models

In this approach the n-gram language models are built. Presence or frequency of n-grams can be used. In text classification, frequency of n-grams gives better results. Normally, the frequency is converted to TF- IDF to take terms importance for a document into account. However, Pang et al. [1] found that term- presence gives better results than term frequency. Their research of movie reviews shows that uni-gram presence is more suited for sentiment analysis.

4. Discourse structures and semantics

This approach uses discourse relation between text components for classification. In many reviews, the overall sentiment is usually expressed at the end of the text [1]. In this discourse-driven approach the sentiment of the whole review is obtained by determining sentiment between different discourse components and the discourse relations that exist between them. In such an approach, the last paragraph of the review might be given more weight in the determination of the sentiment of the whole review.

Techniques

Sentiment analysis can be implemented using both supervised and unsupervised methods of classification. Supervised methods have shown better performance than the unsupervised methods. However, the unsupervised methods are important too because supervised methods demand large amounts of labelled training data that very expensive whereas acquisition of unlabelled data is easy. Most domains except movie reviews lack labelled training data in this case unsupervised methods are very useful for developing applications.

1. Supervised Techniques

Supervised techniques can be implemented by building a classifier. This classifier is trained by examples which can be manually labelled. Mostly used supervised algorithms are Support Vector Machines (SVM), Naive Bayes classifier and Maximum Entropy. It has been shown that Supervised Techniques outperform unsupervised techniques in performance.

Cui et al. have argued that SVMs are more appropriate for sentiment classification because they can better perform when review contains both positive and negative words. However, when the set of training data is small, a Naive Bayes classifier might be more appropriate because SVMs requires a large set of data in order to build a high-quality classifier. One of the most important tasks in sentiment classification is selecting an appropriate set of features. The most commonly used features in sentiment classification are introduced below.

Term presence and their frequency:

These features include uni-grams or n-grams and their frequency or presence. These features have been widely and successfully used in sentiment classification. Pang et al. claim that uni-grams gives better results than bi-grams in movie review sentiment analysis, but researchers have also reported that bi-grams and tri-grams give better product-review polarity classification.

Part of speech information:

Part-of-speech is used to disambiguate sense which in turn is used to guide feature selection. Part-of- speech tagging is useful for identifying adjectives and adverbs in the sentences which identify the opinion words and nouns which are used to identify features of the products.

Negations:

Negation is also an important feature to take into account since it has the potential of reversing a sentiment. For example, the book is great and the book is not great, here the negation word not makes the second sentence negative.

Opinion words and phrases:

Opinion words and phrases such as like, nice, hate, I'd suggest that... are words and phrases that express positive or negative opinions. The main approaches to identify the semantic orientation (positive or negative) or polarity of an opinion words are statistical-based or lexicon-based.

The main limitation of supervised learning is that it is dependent on the amount and quality of the training data and may fail when training data are insufficient.

2. Unsupervised Techniques

In unsupervised technique, classification is done by comparing the features of a given text against word lexicons whose sentiment values are determined prior to their use. For example, start with positive and negative word lexicons, analyse the document for which sentiment need to find. Then if the document has more positive word lexicons, it is positive, otherwise it is negative. The most prominent work done using unsupervised methods for opinion mining and sentiment detection is by Turney [2]. He uses poor and excellent seed words as they appear more in web for calculating the semantic orientation of phrases, where orientation is measured by pointwise mutual information. The sentiment of a document is calculated as the average semantic orientation of all such phrases. He was able to achieve 66% accuracy for the movie review domain.

Ting-Chun Peng and Chia-Chun Shih [13] uses part-of-speech(POS) patterns for extracting the sentiment phrases of each review, they used unknown sentiment phrase as a query term and get top-N relevant phrases from a search engine. Next, sentiments of unknown sentiment phrases are computed based on the sentiments of nearby known relevant phrase using lexicons. Gang Li & Fei Liu [10] developed an approach for clustering documents into positive group and negative group based on the k-means clustering algorithm.

Approaches used:

1. First approach:

- Fetch tweets through Twitter API, using twitter account.
- Install tweepy, which is the python client for the official Twitter API
- Install textblob, which is the python library for processing textual data.
- Create app and copy 'Consumer Key', 'Consumer Secret', 'Access token' and 'Access Token Secret'.
- Authorize twitter API client.
- Make a GET request to Twitter API to fetch tweets for a particular query.
- Parse the tweets. Classify each tweet as positive, negative or neutral.

2. Second approach:

In this approach we tried to use Bert library to gain more accuracy.

BERT (Bidirectional Encoder Representations from Transformers) makes use of Transformer, an attention mechanism that learns contextual relations between words (or sub-words) in a text. In its vanilla form, Transformer includes two separate mechanisms — an encoder that reads the text input and a decoder that produces a prediction for the task. Since BERT's goal is to generate a language model, only the encoder mechanism is necessary.

As opposed to directional models, which read the text input sequentially (left-to-right or right-to-left), the Transformer encoder reads the entire sequence of words at once. Therefore, it is considered bidirectional, though it would be more accurate to say that it's non-directional. This characteristic allows the model to learn the context of a word based on all of its surroundings (left and right of the word).

We focus on using trained dataset by downloading them through a URL and classifying the sentences as positive, negative or neutral by training the model using this dataset. The data set comprises of various sentences (as of now) and the sentiment representing those sentences. Once the model is trained, the results are tested on a given sentence to check for the expected output (sentiment of the given sentence).

The same trained model can be used to analyse the tweets, instead of sentences, by extracting real time tweets from twitter.

We make use of TensorFlow as well as pandas during this process. Installing transformers to import Bert Tokenizer and TFBertForSequenceClassification to classify the sentiment as positive, negative or neutral.

Challenges for Sentiment analysis

Sentiment analysis classifies text as positive, negative or else objective, so it can be thought as text classification task. Text classification has many classes as there are many topics but sentiment analysis has only three classes. However, there are many factors that make sentiment analysis difficult compared to traditional text classification. The following are some of the factors.

1. Coreference resolution

Coreference resolution is the problem of identifying what a pronoun, or a noun phrase refers to. For example, "We watched the movie and went to dinner; it was awful." What does "It" refer to?

Coreference resolution may be useful for the topic/aspect based sentiment analysis. Coreference resolution may improve the accuracy of opinion mining.

2. Temporal Relations

The time of reviews may be important for sentiment analysis. The reviewer may think that Windows Vista is good in 2008, but now he may have negative opinion in 2009 because of new Windows 7. So assessing this kind of opinions that are changed with time may

improve the performance of the sentiment analysis system. This helps us to observe if a certain product gets improved with time, or people change their opinion about a product.

3. Sarcastic sentences

Text may have Sarcastic and ironic sentences. For example, What a great car, it stopped working in the second day. In such case, positive words can have negative sense of meaning. Sarcastic or ironic sentences can be hard to identify which can lead to erroneous opinion mining.

4. Requirement of World Knowledge

Knowledge about worlds facts, events, people are often required to correctly classify the text. Consider the following example, Casablanca and a lunch comprising of rice and sh: a good Sunday.

The system without world knowledge classifies above sentence as positive due to the word good, but it is an objective sentence because Casablanca is the name of the famous movie.

5. Domain Considerations

The accuracy of sentiment classification can be under the domain of the items to which it is applied. The reason is that the there are many words whose meaning changes from domain to domain. For example, Go read the book. This sentence has positive sentiment in book domain while it indicates negative sentiment for movie domain.

6. Grouping synonyms

Many times, text contains different words having same meaning. So, such word should be identified and group together for accurate classification. It is a difficult task to identify these words, as people often use different words to describe the same feature. For example, voice and sound both refer to the same feature in phone review.

7. Thwarted Expectations

Some text contains sentences starting with different context which has different context at the end. For example, the cast was not good, actors performed poorly, but I liked it. In above review the last sentence makes the whole review positive. If term frequency considered the above statements would classify as negative due to more negative words in review.

8. Negation

In traditional text classification small differences between two pieces of text don't change the meaning very much. In Sentiment analysis, however, "the movie was great" is very different from "the movie was not great. Negation handling is a difficult task in sentiment analysis as it reverses the polarity. Negation also expresses by sarcasm and implicit sentences which doesnt contain any negative words.

9. Review Spam Detection

On product review site, many people write fake reviews, called review spam, to promote their products by giving undeserving positive opinions, or defame their competitors' products by giving false negative opinions. The opinion spam indentation task has great impacts on industrial communities. If the opinion provided services contain large number of spams, they will affect the user's experience. Furthermore, if the user is cheated by the provided opinion, he will never use the system again.

Conclusion:

This work is done by Gathering Twitter Data and implementing python code which authorizes twitter API client, make a GET request to Twitter API to fetch tweets for a particular query. parse the tweets and classifies each tweet as positive, negative or neutral.

- Using sentiment analysis tools to analyse opinions in Twitter data can help companies understand how people are talking about their brand.
- It adds an extra layer to the traditional metrics used to analyse the performance of brands on social media, and provides businesses with powerful opportunities.
- Sentiment analysis helps you monitor your customer's emotions on Twitter and understand how they feel.

There is a huge need in the industry for such applications because every company wants to know how consumers feel about their products and services and those of their competitors Sentiment analysis can be developed for new applications. the techniques and algorithms used for sentiment analysis have made good progress, but a lot of challenges in this field remain unsolved. More future research can be done for solving these challenges.

References:

- [1] Bo Pang, Lillian Lee, Shivakumar Vaithyanathan, Thumbs up? Sentiment Classification using Machine Learning Techniques, Proceedings of EMNLP 2002, pp. 79–86.
- [2] Peter D. Turney, Michael L. Littman, Unsupervised Learning of Semantic Orientation from a Hundred-Billion-Word Corpus, arXiv preprint cs/0212012, 2002 - arxiv.org.
- [3] Alexander Pak and Patrick Paroubek. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In Proceedings of international conference on Language Resources and Evaluation (LREC), 2010.
- [4] Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner and Isabell M. Welp. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In Proceedings of AAAI Conference on Weblogs and Social Media (ICWSM), 2010.
- [5] Bo Pang, Lillian Lee and Shivakumar Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2002.
- [6] CHENHAO TAN, LILIAN LEE, JIE TANG, LONG JIANG, MING ZHOU AND PING LI. USER LEVEL SENTIMENT ANALYSIS INCORPORATING SOCIAL NETWORKS. IN PROCEEDINGS OF ACM SPECIAL INTEREST GROUP ON KNOWLEDGE DISCOVERY AND DATA MINING (SIGKDD), 2011.
- [7] Efthymios Kouloumpis, Theresa Wilson and Johanna Moore. Twitter Sentiment Analysis: The Good the Bad and the OMG! In Proceedings of AAAI Conference on Weblogs and Social Media (ICWSM), 2011.
- [8] Hatzivassiloglou, V., & McKeown, K.R.. Predicting the semantic orientation of adjectives. In Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL, 2009.
- [9] Johann Bollen, Alberto Pepe and Huina Mao. Modelling Public Mood and Emotion: Twitter Sentiment and socio-economic phenomena. In Proceedings of AAAI Conference on Weblogs and Social Media (ICWSM), 2011.
- [10] GANG LI & FEI LIU , CLUSTERING-BASED APPROACH ON SENTIMENT ANALYSIS. IN IEEE INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS AND KNOWLEDGE ENGINEERING
- [11] Luciano Barbosa and Junlan Feng. Robust Sentiment Detection on Twitter from Biased and Noisy Data. In Proceedings of the international conference on Computational Linguistics (COLING), 2010.
- [12] Abdullah Alsaedi, Mohammad Zubair Khan, A Study on Sentiment Analysis Techniques of Twitter Data, International Journal of Advanced Computer Science and Applications, Vol. 10, No. 2, 2019
- [13] Ting-Chun Peng, Chia-Chun Shih, Using Chinese part-of-speech patterns for sentiment phrase identification and opinion extraction in user generated reviews, Fifth International Conference on Digital Information Management (ICDIM), 2010.