Analysis of Behavior Extraction on Social Life Issues by Tweets using Machine Learning Sentiment Analysis

Adnan Beg[1], Shish Ahmad[2]

[1]U.G student, Department of Computer Science and Engineering, Integral University, Lucknow, India

[2]Associate Professor, Department of Computer Science and Engineering, Integral University, Lucknow, India

Abstract: ***At present time most of the world can be found on internet. The popularity of the social networking sites has been growing fast, parallel to emerging technologies, with increase in the number of users to the social networking sites that actively express their opinions on these sites. To forecast the sentiment analysis we have used the data stored on social sites stockpiling. Our goal is to retrieve data from social sites and analyze the emotion of a particular person on certain topic. For our analysis we have used the data available on Twitter, as Twitter has a large amount of data that people post, which gives the output beyond the polarity but those polarities can be used in product profiling, trend analysis and forecasting. This paper presents an overview of past and current research on twitter sentiment analysis and presents better ideas for future work.***

Keywords: Twitter; textmining; Microblogging; Sentiment Analysis

1. Introduction

Computers are beginning to learn to read between the lines of our tweets, Facebook updates, and email messages. Humans are fairly sophisticated when it comes to understanding the complex meanings beneath spoken or written word but computers are unable to understand the actual meaning of this type of sentence. For example, we can tell that a statement like, “I again missed the train. Brilliant!” is sarcastic, and not actually brilliant. And with the help of machine learning, computers are beginning to get better at the same, resulting in a new kind of analytics known as sentiment analysis. Sentiment analysis is also known as “opinion mining” or “emotion Artificial Intelligence” and alludes to the utilization of natural language processing (NLP), text mining, computational linguistics, and bio measurements to methodically recognize, evaluate, and examine emotional states and subjective information. Sentiment analysis is basically concerned with the voice in client materials; for example, surveys and reviews on the Web and web- based social networks. The basic aim of sentiment analysis is to determine the attitude of an individual or group regarding particular social issues. The sentiment or attitude may be a judgment, evaluation or emotional reaction of people. For example, Expedia in Canada used sentiment analysis to determine that the music accompanying one of their commercials was receiving an overwhelmingly negative response online, and they were able to respond to that sentiment appropriately: by releasing a new version of the commercial in which the offending violin was abruptly smashed. There have been several research projects that apply sentiment analysis to Twitter corpora to extract general public opinion regarding different political issues [1]. Most recently, the Indian government has abolished Article 370 in Jammu & Kashmir and after this abolition , Article 370 became a trending topic in India as well as in other countries. Some people are celebrating this by tweeting greets while some are against this abolition, so they are tweeting negative content [2]. But it is not confirmed how many are supporting this abolition and how many are opposing it. So to get an idea of majority review of people, we developed a model that retrieves tweets on a certain topic through the Twitter API and calculates the percentage of positive, negative and neutral tweets.

Of course, sentiment analysis is not yet 100% accurate, and still needs a human’s watchful eye to ensure that the nuances of human speech are being fully understood by the computer. In addition, it’s important to note that not all communications can be classified as positive, negative, or neutral; we’re just too complex for that.

Twitter API

Twitter is a service of micro-blogging with two main characteristics: its users end messages (tweets) of 140 characters usually compounds by keywords (in the form of hashtags), natural language and common abstracts; moreover, each user can follow other users so that his timeline is populated by their tweets. It is much easier to obtain user data because the profiles are public and can be viewed by anyone. As for Facebook, there is the ability to change the settings relating to privacy so that users can see their profile only after we have accepted their request. Even in this case, however, users who choose this route about privacy are few; moreover, also any person not registered on Twitter can access user profiles. Both social network provide APIs (Application Programming Interface) with the O Auth authentication mechanism. The purpose of this protocol is to provide a framework for the verification of the identity of the entities involved in secure transactions. There are at the moment two versions of this protocol: O Auth 1.0 [3] and O Auth 2.0. Both versions support two-legged authentication, in which a server is guaranteed about the user identity and three-legged authentication, in which a server is guaranteed by an application about the user identity. This type of authentication requires the use of the access token and it is currently implemented by the Social Network. Compared to Facebook, Twitter connections are bidirectional: there is an asymmetric network consisting of friends, that is the accounts that a user follows and followers, that is the accounts that follow the user. The timeline of a user that you can trace in the home consists of a real-time stream containing all the tweets of his friends. As well as Facebook provides the Graph API Explorer useful to explore the API, Twitter provides the Twitter Console; generally, Twitter offers an extensive collection of APIs, all based on HTTP. Twitter supports two authentication methods based on the O Auth protocol: the first one based on O Auth 1.0a related to the user and the second one based on O Auth 2.0 related to an application. The first mode defined application-user authentication includes an HTTP authorization request that communicates what application is making the request, on behalf of which use the application is making the request, if the user has authorized or not the application and if during transit the request has been tempered by third parties. In the second mode, defined application- only authentication, the application encodes its consumer key and its secret key in a set of encoded credentials and then performs an HTTP POST request to endpoint oauth2/token to exchange these credentials with a bearer token. The bearer token obtained is used to authenticate the application that it represents in the REST API. The latter approach is much simpler because it is not required that the call is signed. The typology of the Twitter API end-point is the following: https ://api.twitter.com/1.1/ <resource > / < action >. The Twitter API includes 16 resources: timeline, tweet, search, streaming, direct messages, friends and followers, users, user suggested, favorites, lists, saved search, places, trends, spam reporting, O Auth, help. The Twitter Search API allows the execution of real- time queries on recent tweets. In particular, the query must be simple, limited to a maximum of 1000 characters, including operators and it is always required some form of authentication. In this case, the only available resource is the tweet. The Twitter Streaming API allows a real-time update of information relating to specific resource, thereby eliminating the need to repeatedly call at regular intervals (polling) its REST end-point.

Behavior Analysis

Sentiment analysis aims to analyze people's sentiment, opinions, attitudes, emotions. Different techniques and software tools have been developed to carry out Sentiment Analysis. Most of the works in this research area focus on classifying texts according to their sentiment polarity, which can be positive, negative or neutral. Therefore, it can be considered a text classification problem, since its goal consists of categorizing texts within classes by means of algorithmic methods. The paper [4] offers a comprehensive review of this topic and compares some free access web services, analyzing their capabilities to classify and score different pieces of text with respect to the sentiments contained therein. In the last years, thanks to the increasing amount of information delivered through social networks, many researches have been focused on applying sentiment analysis to these data. Sentiment analysis aims at mining user opinion and sentiment polarity from the posted text on the social network. In [5], the authors apply data mining techniques to psychology, specifically to the field of depression, to detect depressed users in social network services. In fact, the main symptom of depression is severe negative emotions and lack of positive emotions.

2. Related Works on Sentiment Analysis

In recent years, a large number of techniques and enhancements have been proposed for the problem of sentiment analysis (SA) in different fields and for different tasks. Three types of techniques are used to classify opinions in SA, namely, lexicon-based approaches, machine learning approaches, and hybrid approaches. Lexicon-based approaches do not require any training dataset and use sentiment lexicons such as WordNet [6] and SentiWordNet [7], for classification purposes. These approaches give sentiment scores ranging from −1 to 1, but they do not classify the context-dependent opinion words appropriately. Moreover, there are also hybrid approaches that combine machine learning and lexical approaches [8].

In contrast, machine learning approaches can be grouped into two main categories, which are supervised and unsupervised techniques. Support Vector Machine (SVM) and Naive Bayesian (NB) classifications are examples of supervised sentiment analysis techniques that have achieved higher success in text classifications [9, 10, 12]. A primary concern of supervised approaches is that they depend on large training datasets, which can be time-consuming to collect for each domain. The supervised methods also depend on the selection and extraction of the appropriate set of features to be used for the detection of sentiments. For instance, uni-grams, bigrams, and part-of-speech tags are used as feature extractors [9, 11]. Feature selection enhances the sentiment classification method by combining syntactic features with semantic information from sources like SentiWordNet [13]. Feature selection methods, such as Principal Component Analysis (PCA) and Random Projection (RP), are aimed at eliminating irrelevant and redundant features to yield an improved classification accuracy for machine learning techniques and also to reduce data dimensionality [14]. Various feature selection approaches, such as information gain and the chi-square test, are employed to gain higher accuracy in sentiment analysis [15], but many of these studies mainly focused on document-level sentiment analysis.Authors [16] presented an efficient method of feature selection and ensemble learning for an aspect- based sentiment analysis. The algorithm is based on a single-objective PSO and basic learning algorithms, namely CRF, SVM and ME. Using the SVM+PCA method, an accuracy of 74.51% was obtained on a laptop, on restaurant review datasets. The aim of this works was to investigate the use of a hybrid sentiment classification to enhance the performance of a Twitter aspect-based sentiment classification. A General model for sentiment analysis is as follows,

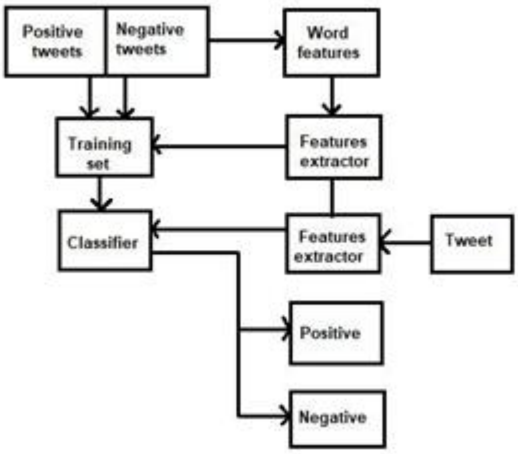

Figure 1: Sentiment Analysis Architecture

Following are the phases required for sentiment analysis of twitter data:

Pre-processing of the datasets

A tweet contains a lot of opinions of people about the data which are expressed in different ways by different users. The twitter dataset which is used in this survey work is already labeled into two classes viz. negative and positive polarity. Thus the sentiment analysis of the data becomes easy to observe the effect of various characteristics. The contradictory raw data is highly susceptible to inconsistency and redundancy. Preprocessing of tweet include the following points,

- Remove all URLs (e.g. www.xyz.com), hash tags (e.g. #topic), targets (@username)
- Correct the spellings; sequence of repeated characters is to be handled
- Replace all the emoticons with their sentiment.
- Remove all punctuations ,symbols, numbers
- Remove Stop Words
- Expand Acronyms(we can use a acronym dictionary)
- Remove Non-English Tweets

Table 1: Publicly Available Datasets for Twitter

HASH	Tweets	http://demeter.i nf .ed.ac.uk	31,861 Pos tweets 64,850 Neg tweets, 125,859 Neu tweets
EMOT	Tweets and Emoticons	http://twittersenti ment.appspot.c o m	230,811 Pos& 150,570 Neg tweets
ISIEVE	Tweets	www.i- sieve.com	1,520 Pos tweets,20 0 Neg tweets, 2,295 Neu tweets
Columbia univ.datas et	Tweets	Email: apoorv@cs.col um bia.edu	11,875 tweets
Patient dataset	Opinions	http://patientop in ion.org.uk	2000 patient opinions
Sample	Tweets	http://goo.gl/U Qv dx	667 tweets
Stanford dataset	Movie Reviews	http://ai.stanfor d. edu/~amaas/dat a/ sentiment/	50000 movie reviews
Stanford	Tweets	http://cs.stanfor d. edu/people/alec m go/training and test data.zip	4 million tweets categorized as positive and negative
Spam dataset	Spam Reviews	http://myleott.com /op_spam	400 deceptive and 400 truthful reviews in positive and negative category.
Soe dataset	Sarcasm and nasty reviews	http://nlds.soe. uc sc.edu/iac	1,000 discussion s, ~390,000 posts, and some ~ 73,000,00 0 words

2.3 Training

Supervised learning is an important technique for solving classification problems. Training the classifier makes it easier for future predictions for unknown data.

Feature Extraction

The preprocessed dataset has many distinctive properties. In the feature extraction method, we extract useful characteristics from the data, which in computer vision corresponds to calculating values from input images.

Later on, this aspect is used to compute the positive and negative polarity in a sentence which is useful for determining the opinion of the individuals using models like unigram, bigram [17]. Machine learning is a technique that requires representation of the key features of text or documents for processing. These key features are considered as feature vectors that are used for the classification task. Some examples of features that have been reported in the literature are

1) *Words And Their Frequencies*

Unigrams, bigrams, and n-gram models with their frequency counts are considered as features. There has been more research on using word presence to better describe this feature. Panget al. [18] showed better results by using a presence rather than using frequencies.

2) *Parts Of Speech Tags*

Parts of speech like adjectives, adverbs, and some groups of verbs and nouns are good indicators of subjectivity and sentiment.

3) *Opinion Words and Phrases*

Apart from specific words, idioms that convey sentiments can be used as features. e.g. cost someone an arm and leg.

4) *Position Of Terms*

The position of a term within a text can affect how much the term makes a difference in an overall sentiment of the text.

5) *Negation*

Negation is a very important but difficult feature to interpret. The presence of a negation usually changes the polarity of the opinion.

6) *Syntax*

Syntactic patterns like collocations are used as features to learn subjectivity patterns by many of the researchers.

Classification

Naive Bayes:
It is a probabilistic classifier and can learn the pattern of examining a set of documents that has been categorized [19]. It compares the contents with the list of words to classify the documents to their right category or class.

To train and classify using Naïve Bayes Machine Learning technique , we can use the Python NLTK library .

Maximum Entropy
The Max Entropy classifier is a probabilistic classifier that belongs to the class of exponential models and in this , assumptions are taken regarding the relationship between the features extracted from the dataset. The main task of this classifier is to maximize the entropy of the system by estimating the conditional distribution of the class label. Maximum entropy also handles overlap features and is the same as the logistic regression method which finds the distribution over classes.

Support Vector Machine:
The support vector machine analyzes the data, defines the decision boundaries and uses the kernels for computation which are performed in input space [20]. The input data are two sets of vectors of size n each. Then every data represented as a vector is classified into a class. Next, we find a margin between the two classes that are far from any document. The distance defines the margin of the classifier, maximizing the margin reduces indecisive decisions. SVM (support vector machine) also supports classification and regression which are useful for statistical learning theory and it also helps to recognize the factors precisely, that need to be taken into account, to understand it successfully.

3. Approaches for Sentiment Analysis

There are mainly two techniques for sentiment analysis for the twitter data:

Machine Learning Approaches
The machine learning-based approach uses a classification technique to classify text into classes. Machine learning techniques are mainly of two types:

Unsupervised learning:
It does not consist of a category and they do not provide with the correct targets at all and therefore rely on clustering.

Supervised learning:
It is based on the labeled dataset and thus the labels are provided to the model during the process. These labeled datasets are trained to get meaningful outputs when encountered during decision- making. The success of both these learning methods mainly depends on the selection and extraction of the specific set of features used to detect sentiment. The machine learning approaches applicable to sentiment analysis mainly belong to supervised classification. In machine learning techniques, two sets of data are needed:

- Training Set
- Test Set.

A lot number of machine learning techniques have been formulated to classify the tweets into classes. Machine learning techniques like Naive Bayes (NB), maximum entropy (ME), and support vector machines (SVM) have achieved great success in sentiment analysis.

Machine learning starts with collecting training dataset. Nextly we train a classifier on the training data. Once a supervised classification technique is selected, an important decision to make is to select a feature. They can tell us how documents are represented. The most commonly used features in sentiment classification are

- Term presence and their frequency
- Part of speech information
- Negations
- Opinion words and phrases

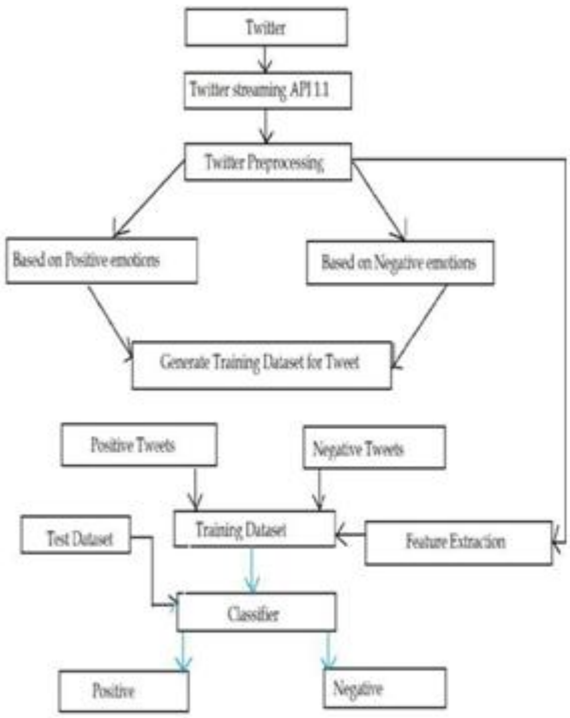

Figure 2: Sentiment Classification Based On Emoticons

With respect to supervised techniques, support vector machines (SVM), Naive Bayes, Maximum Entropy are some of the most common techniques used.

Lexicon Based Approaches
Lexicon based methods [20] use sentiment dictionary with opinion words and match them with the data to determine polarity. They assign sentiment scores to the opinion words describing how Positive, Negative and Neutral words contained in the dictionary are. This approach mainly depends on a sentiment lexicon such as the Opinion Finder lexicon;

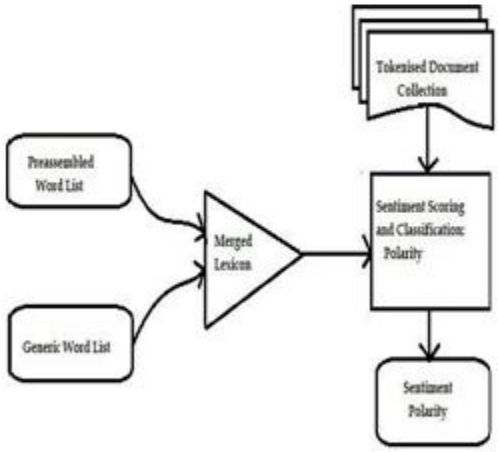

Figure 3: Lexicon-Based Model

There are two sub classifications for this approach:

- Methods based on statistics: Latent Semantic Analysis (LSA).
- Methods based on semantic such as the use of synonyms and antonyms or relationships from thesaurus like WordNet may also represent an interesting solution.

According to the performance measures like precision and recall, we provide a comparative study of existing techniques for opinion mining, including machine learning, lexicon-based approaches, cross-domain, and cross-lingual approaches, etc., as shown in Table 2.

Table 2: Performance Comparison of Sentiment Analysis Methods

	Method	Data Set	Acc.	Author
Machine Learning	SVM	Movie reviews	86.40%	Pang, Lee[23]
	Co Training SVM	Twitter	82.52%	Liu[14]
	Deep learning	Stanford Sentiment Treebank	80.70%	Richard[18]
Lexical based	Corpus	Product reviews	74%	Turkey

Dictionary-based:
It is based on the usage of terms that are usually collected and annotated manually. This set grows by searching the synonyms and antonyms of a Dictionary. An example of that dictionary is WordNet, which is used to develop SentiWordNet (a thesaurus).
Drawback: Can't deal with domain and context Specific-orientations.

Corpus-Based:
The corpus-based approach provides dictionaries related to a specific domain. A set of opinion terms that grows through the search of related words by means of the use of either statistical or semantic techniques are used to generate dictionaries:

	Dictionary	Amazon's Mechanical Turk	-----	Taboada[20]
Cross lingual	Ensemble	Amazon	81.00%	Wan,X[16]
	Co- Train	Amazon, ITI68	81.30%	Wan,X [16]
	EWG A	IMDb Movie review	>90%	Abassi,A.
	CLMM	MPQAN, TCIR, ISI	83.20%	Mengi

4. Sentiment Analysis Task

Sentiment analysis is an interdisciplinary task which includes natural language processing, web mining and machine learning. It is a complex task and can be decomposed into following tasks, viz:

a) Subjectivity Classification
b) Sentiment Classification
c) Complimentary Tasks
   - Object Holder Extraction
   - Object/Feature Extraction

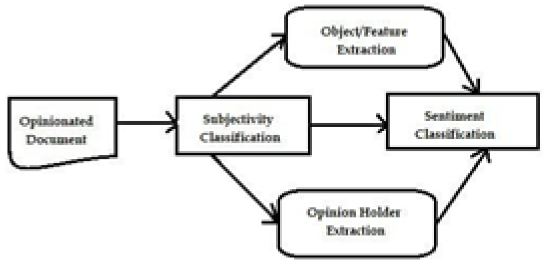

Figure 4: Sentiment Analysis Tasks

a) Subjectivity classification

Subjectivity classification is the task in which sentences are classified as opinionated or not opinionated. Let S = {s1, . . . , sn} be a set of sentences in the document d. The problem of subjectivity classification is to identify the sentences which are used to represent opinions and other forms of subjectivity(subjective sentences set Ss) from sentences used to objectively present factual information (objective sentences set So), where SsUSo=S.

b) Sentiment Classification

Once the task of finding whether a sentence is opinionated is done, we have to find the polarity of the sentence i.e., whether it expresses a positive or negative opinion. Sentiment classification can be a binary classification (positive or negative),multi- class classification(extremely negative, negative, neutral, positive or extremely positive), regression or ranking.
The way in which subtasks of opinion holder extraction and object feature extraction are treated as is totally dependent on the application of sentiment analysis.

c) Complimentary Tasks

- *Opinion Holder Extraction*
  In this opinion holders or sources are discovered. The detection of the opinion holders is to recognize direct or indirect sources of opinion.

- *Object /Feature Extraction*
  It is the discovery of the target entity

5. Levels of Sentiment Analysis

The tasks that are described in the previous section can be done at several levels of granularity.

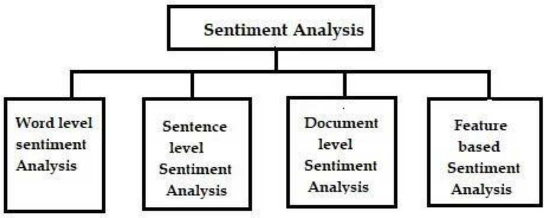

Figure 5: Levels of Sentiment Analysis

Document level
It basically deals with tagging individual documents with their sentiment. In this, the whole document is classified either into a positive or negative class.

General Approach:
Find the sentiment polarities of individual sentences or words and combine them to find the polarity of the document.

Other approaches:
Complex linguistic phenomena like co-reference resolution, pragmatics, etc.
Various Tasks involved in this are:
- Task: Sentiment Classification of whole document
- Classes: Positive, negative and neutral
- Assumption :Each Document focuses on a single object (not true in discussion posts, blogs, etc.) and contain opinion from a single opinion holder

Sentence or phrase level
Sentence-level Sentiment Analysis tag the individual sentences with their respective sentiment polarities. Sentence level sentiment classification classifies sentence three categories i.e, into positive, negative or neutral class.

General approach:
This approach finds the sentiment orientation of individual words in the sentence/phrase and then to combine them to determine the sentiment of the whole sentence or phrase.

Other approaches:
Consider discourse structure of the text

Various Tasks involved in this are:
- Task 1: Identifying Subjective/ Objective Sentences
  Classes: Objective and Subjective
- Task 2: Sentiment Classification of Sentences
  Classes: positive and negative **Assumption:** A sentence contains only one Opinion which may not always be true

Aspect level or Feature level
In this, each word is labeled with their sentiment and it also identifies the entity towards which the sentiment is directed. Aspect or Feature level sentiment classification is mainly concerned with identifying and extracting product features from the source data. Some techniques like dependency parser and discourse structures are used in this.

Various Tasks involved in this are:
Task1: Identify and extract object features that have been commented on by an opinion holder (eg. A reviewer)
Task2: Determining whether the opinions on features are negative, positive or neutral
Task 3: Find feature synonyms

Word Level
Most recent works have used the prior polarity of words and phrases for sentiment classification at sentence and document levels Word sentiment classification use mostly adjectives as features but adverbs, There are two methods of automatically annotating sentiment at the word level :
1) Dictionary-Based Approaches
2) Corpus-Based Approaches.

6. Evaluation of Sentiment Classification

The performance of sentiment classification can be evaluated by using four indexes calculated as the following equations:
Accuracy = (TP+TN)/(TP+TN+FP+FN)
Precision = TP/(TP+FP) Recall = TP/(TP+FN)
F1 = (2×Precision×Recall)/(Precision+Recall)

In which TP, FN, FP and TN refer respectively to the number of true positive instances, the number of false negative instances, the number of false positive instances and the number of true negative instances, as defined in the table 3.

Table 3: Confusion Matrix

	Predicted Positives	Predicted Negatives
Actual Positive	TP	FN
Actual Negative	FP	TN

7. Results and Discussion

We used the tweepy library i.e. a python library to extract the tweets from twitter. As the twitter data is not publicly available so we have used the consumerKey, consumerSecret, accessToken and accessTokenSecret to access the twitter's data. After accessing the tweets we have created a dataset containing userID, date & time and the user's tweets.

We used the pandas (python library) where the functions is applied to the raw sentences which make it more appropriate to understand. We used two csv files containing positive and negative words respectively. For comparison purpose, we have created two lists one for the positive words and another for negative words. Then we appended words from csv files into these two newly created lists. After that we have compared the content of each tweet with these two files to find accuracy of positivity and negativity of that particular tweet. In this way we have analysed the 1000 tweets and calculated the accuracy of positivity , negativity or neutral of the sentiments of the people on a particular topic.
The result of our analysis is shown in table:4

Table 4: Result table

Data Set	1000
Negative	350
Positive	200
Neutral	450

The visualization of result is represented by bar graph (fig.6) and pie chart (fig.7) which is created by matplotlib library.

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

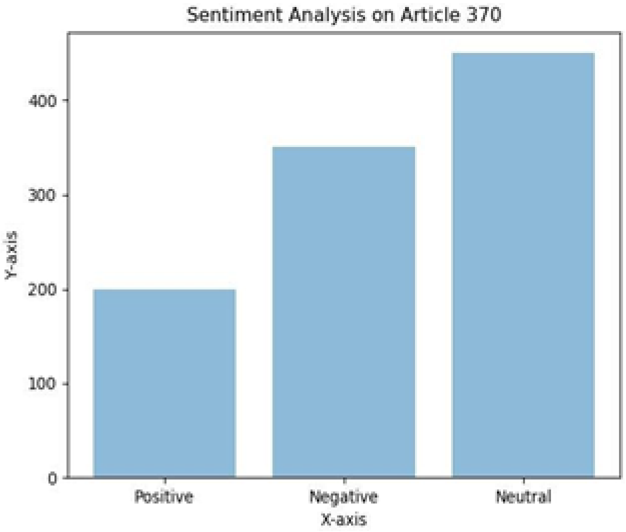

Figure 6: Bar Graph

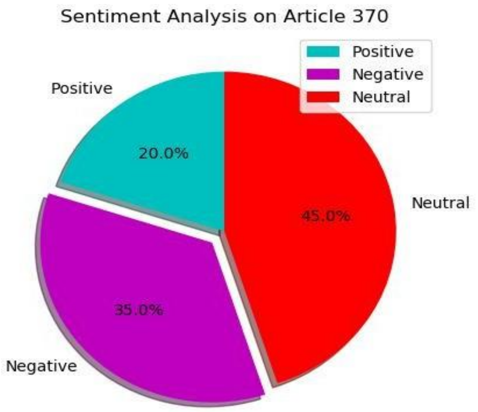

Figure 7: Pie chart

8. Conclusion

In this paper, we provide a survey and comparative study of existing techniques for sentiment analyses including machine learning and lexicon-based approaches, together with cross-domain and cross- lingual methods. We can conclude that cleaner data, more accurate results can be obtained. The use of the python libraries may provides better sentiment accuracy. We can focus on the study of combining the machine learning method into the opinion lexicon method in order to improve the accuracy of sentiment classification and adaptive capacity to a variety of domains and different languages.

References

[1] nlp.stanford.edu/courses/cs224n/2011/rep orts/patlai.pdf
[2] https://www.thehindubusinessline.com/news/centre-revokes-article-370-virtual-sloganeering-takes-over-twitter/article28824103.ece#
[3] Hammer-Lahav, E.,: The OAuth 1.0 Protocol. RFC 5849, April 201020. Hardt, D.,: The OAuth 2.0 Authorization Framework. RFC 6749, October 201222. Twitter API.
[4] https://dev.twitter.com/overview/documen tation
[5] Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., Herrera-Viedma, E.,: Sentimentanalysis: A review and comparative analysis of web services. Information Sciences.
[6] Wang, X., Zhang, C., Ji, Y., Sun, L., Wu, L.,: A Depression Detection Model Basedon Sentiment Analysis in Micro-blog Social Network. Trends and Applications inKnowledge Discovery and Data Mining. LNCS, vol. 7867, 201–213, (2013)
[7] Miller GA (1995) Wordnet: a lexical database for english. Commun ACM 38(11):39–41.
[8] Baccianella S, Esuli A, Sebastiani F (2010) Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Calzolari N, Choukri K, Maegaard B, Mariani J, Odijk J, Piperidis S, Rosner M, Tapias D (eds) LREC. European Language Resources Association. http://nmis.isti.cnr.it/sebastiani/ Publications/LREC10.pdf
[9] Appel O, Chiclana F, Carter J, Fujita H (2016) A hybrid approach to the sentiment analysis problem at the sentence level. Knowl-Based Syst 108:110–124. https://doi.org/10.1016/j. knosys.2016.05.040.http://www.sciencedir ect.com/science/article/ pii/S095070511630137X. New Avenues in Knowledge Bases for Natural Language Processing
[10] Pang B, Lee L, Vaithyanathan S (2002) Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on empirical methods in natural language processing - volume 10, EMNLP '02. Association for Computational Linguistics, Stroudsburg, pp 79–86
[11] Prabowo R, Thelwall M (2009) Sentiment analysis: a combinedapproach. JInform 3(2):143–157. https://doi.org/10.1016/j. joi.2009.01.003. http://www.sciencedirect.com/science/arti cle/pii/ S1751157709000108
[12] Tripathy A, Agrawal A, Rath SK (2016) Classification of sentiment reviews using n-gram machine learning approach. Expert Syst Appl 57:117–126.https://doi.org/10.1016/j.eswa.2016.0 3.0 28.
[13] http://www.sciencedirect.com/science/arti cle/pii/S0957417416301 18X
[14] Vinodhini G, Chandrasekaran R (2016) A comparative performance evaluation of neural network based approach for sentiment classification of online reviews. J King Saud University - Comput Inform Sci 28(1):2–12.
[15] https://doi.org/10.1016/j.jksuci.2014.03. 024.http://www.sciencedirect.com/science
[16] /article/pii/S131915781 5001020
[17] Abbasi A (2010) Intelligent feature selection for opinion classification. IEEE Intell Syst 25(4):75–79. https://www.scopus.com
[18] Sabbah T, Selamat A, Selamat MH, Ibrahim R, Fujita H (2016) Hybridized term-weighting method for dark web classification. Neurocomputing 173(3):1908–1926
[19] Zainuddin N, Selamat A (2014) Sentiment analysis using support vector machine. In: 2014 International Conference on Computer, Communications, and Control Technology (i4CT), pp 333–337
[20] Akhtar MS, Gupta D, Ekbal A, Bhattacharyya P (2017) Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis. Knowl-Based Syst 125:116– 135. https://doi.org/10.1016/j.knosys.2017.03.0

[21] 20. http://www.
[22] sciencedirect.com/science/article/pii/S095 070511730148X
[23] Socher, Richard, et al. "Recursive deep models for semanticcompositionality over a sentiment Treebank." Proceedings of theConference on Empirical Methods in Natural Language Processing(EMNLP). 2013
[24] Pang, B.and Lee, L. “A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts”. 42nd Meeting of the Association for Computational Linguistics[C] (ACL-04). 2004, 271-278.
[25] Pablo Gamallo, Marcos Garcia, “Citius: A Naive-Bayes Strategyfor Sentiment Analysis on English Tweets", 8th InternationalWorkshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland,Aug 23-24 2014, pp 171-175.
[26] Liu, S., Li, F., Li, F., Cheng, X., &Shen, H.. Adaptive cotraining SVM for sentiment classification on tweets. In Proceedings of the 22nd ACMinternational conference on Conference on information & knowledgemanagement (pp. 2079-2088). ACM,2013.