IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 64-72

Opinion Analysis of Twitter Data Using Machine Learning Technique

Bireshwar Ganguly[1], Devashri Raich[2]
[1]*Assistant Professor Dept. of IT, RCERT, Chandrapur*
[2]*Assistant Professor Dept. of IT, RCERT, Chandrapur*

Abstract: *Web based Micro blogging on informal organizations have been utilized for demonstrating opinions about certain substance in extremely short messages. Existing well known micro blogs like twitter, facebook etc, in which twitter achieves greatest measure of consideration in the field of research regions identified with item, film audits, stock trade and so forth. The examination on opinion analysis has been going for quite a while. Supposition analysis in present days turns into the serious issue in field of research and innovation. Because of step by step increment in the quantity of users on the long range informal communication sites, enormous measure of information delivers as content, sound, video and pictures. There is have to do estimation analysis as writings in type of messages or presents on discover the whether the opinion is negative, positive or unbiased. We had separated information from twitter for example movie surveys for feeling forecast utilizing machine-learning calculations. We connected administered machine-learning calculations like Support vector machines (SVM), maximum entropy and Naive Bayes to characterize information utilizing unigram, bigram and cross breed for example unigram + bigram highlights. Result demonstrates that SVM outperformed different classifiers with striking precision of 87% for film audits.*

Keywords: *Opinion mining, movie reviews, machine learning, SVM*

I. Introduction

Social networks are the gathering of people in explicit gatherings. It might be a political or religious gathering or a gathering of college understudies, youths, all sharing data about their interests, basically on the web. Twitter, MySpace or Facebook are a portion of the free and simple to- get to interpersonal interaction locales. Almost certainly, this connection incorporates kinship, families, aggregate connections, and opinional connections. Interpersonal organizations help individuals make new companions create individual connections and keep in contact with family very effectively. Due to the expansive number of individuals who interface with the system locales, the quantity of connections increments step by step.The long range informal communication highlights consolidated in a site are: client gatherings, most recent data about music gatherings, video and photograph areas, websites, individual profile and the sky is the limit from there. Interpersonal interaction locales additionally help individuals keep up and create business contacts with them. LinkedIn is the best model, as it very well may be a decent spot to talk about business and meet experts. It is simpler and quicker to engage with new business users. Web is the first and the primary correspondence innovation to change the social association of individuals. Since the mid 1990s, Web appropriation has developed quickly. For instance, in 2003, 63% of Americans had utilized the Web. During the 1990s, data innovation specialists anticipated that the Web should be thrown ever.

Web has turned into a basic piece of our lives. Numerous sites have offices for individuals to keep in contact as informal communities. Interpersonal interaction destinations are the best approach to communicate with new individuals and make associations, just as offer photographs, recordings and exercises with one another.

Opinion Mining: Opinion Mining is to decide the supposition of client identified with some occasion or the announcement portray the feeling of the client for example what he/she feel about it. Users share the things about their progressing life, talk about current issues and assortment of themes. Free to write in any configuration without following guidelines that makes this more famous than more established blogging locales. Films and item surveys effectively accessible now a days or contemplations on religious and political issues, so it winds up fundamental wellsprings of client notion and opinion. Data that we utilize in our analysis are from twitter, it contains huge number of messages by extensive number of users made without anyone else. Messages can change from popular conclusion to individual idea. For instance some post from twitter can be appeared Table 1.1.

These micro blogging destinations are immense wellspring of data and it is very simple to state that there is a need of computerizing the opinion examination process as there is excessively work engaged with

preparing this data physically. Different methodologies are drilled for the mechanization of this procedure like machine learning and Normal language handling. Users are expanding step by step as the populace and pattern of utilizing micro blogging destinations are expanding, so the information can be utilized in research reason for slant analysis and assessment mining.

For instance, film creators keen on following inquiries:-

- What is group of onlookers desire from our movie?(whether the film is agreeable or not)
- How the general population responded to our movie?
- Whether the movie is swing to be great or terrible?

In the season of race each news channel demonstrate the leave surveys of each ideological groups, so every ideological group willing to know what number of are in support and with the assistance of micro blogging destinations individuals will give the conclusions about preferences of the gathering. These assessments will help gatherings to build their voters.

The information we utilizing in this examination are movie audits. We have gathered around 17000 movie surveys from twitter. The movie survey contains audits of various movies. Surveys can be ordered in three different ways:

1. **Positive surveys**: messages in which individuals preferred the movie.
2. **Negative surveys:** messages in which individuals not preferred the movie.
3. **Neutral surveys:** messages in which individuals doesn't have any feeling or dependent on insignificant truth.

Fig 1: Opinion Mining of Social Media

II. Literature Survey

Opinion Analysis is the intensive research of how opinions and viewpoints can be identify with ones feeling and frame of mind appears in normal language admiration to an occasion. Ongoing occasions demonstrate that the opinion mining has come to upto extraordinary accomplishment which can outperform the positive versus negative and manage entire field of conduct and feelings for various networks and points. In the field of opinion mining utilizing distinctive strategies great measure of research has been done for expectation of social networks.

Pang and lee [1] proposed the framework where a feeling can be sure or negative was discovered by proportion of positive words to add up to words. Later in 2008 the creators created strategy in which tweet result can be chosen by term in the tweet. Contrast with baselines that are created by people, the outcomes are truly great when machine learning procedures are utilized. SVM gave best outcome as contrast with Naïve Bayes. Despite utilizing distinctive sorts of highlights the creators did not achieve wanted exactnesses over theme based order.

Jiang *et al.* [2] centre around target-subordinate opinion grouping. Here target-subordinate methods whether the opinion is certain, negative or unbiased relies upon nature of the inquiry that is inquired. The creators proposed to improve target- subordinate conclusion arrangement by joining highlights of target-subordinate and thinking about related tweets. The creators likewise recommended that there is need of thought current tweets to the related tweets by utilizing chart based streamlining. As guaranteed by creators trial results, the chart based streamlining builds the execution.

Tan et al. [3] said that the users that common comparative feelings are probably going to be associated. The creators proposed the model that were produced from either by following the system that has been made by labelling diverse client with the assistance of "@" or by examining the system of twitter supporter/follower. The creators clarified that by utilizing data of connection of twitter there will be improvement in client level supposition examination.

Chen et al. [4] utilized the feed-forward BPN system and utilizations assessment introduction to figure

the outcomes at every neuron. The creators proposed a system dependent on neural system. The proposed system is blend of machine learning classifiers and semantic introduction files. So as to get proficiency in strategy, semantic introduction files utilized as contributions for neural system. The proposed procedure beats other neural systems and customary methodologies by expanding proficiency in both preparing just as arrangement time.

Malhar and Ram [5] utilized managed machine learning procedures and counterfeit neural systems to order twitter information alongside contextual analysis of Presidential and Assembly races which results SVM beats every single other classifier. The creators proposed a technique to anticipate the result of race results by using the client impact factor. To complete decrease in measurement the creators consolidated the Principle Component Analysis with SVM.

Anton and Andrey [6] inspected the current strategies and built up a model for programmed estimation examination of twitter messages utilizing unigram, bigram and mutually for example half breed highlight. The motivation behind the creators is to investigate and deliver approaches for breaking down the emphasize of the messages in online networking. The creators looked into existing programmed conclusion examination approaches and so as to keep up the setting of developing techniques the character highlight of internet based life proclamations were considered.

Pak and Paroubek [7] perform phonetic analysis and manufacture a feeling classifier to decide positive, negative and nonpartisan estimations for a report. The creators built up an assumption classifier, which gives unbiased, negative and positive articulations of an archive. So as to prepare opinion classifier the creator proposed a methodology that gathers corpus consequently. So as to break down the uniqueness in dissemination among impartial, negative and positive sets, the creators utilized TreeTagger.

Kopel and Schler [8] clarify that it is imperative to utilize unbiased messages to get great information of extremity. The creator additionally expresses that positive and negative messages alone won't give legitimate comprehension about unbiased messages. Thinking about nonpartisan messages clear the distinction among positive and negative messages.The creators found that in one of the corpus having the majority of the unbiased reports gives no notion which can be utilized as counter to test both energy and antagonism of a record.

Go et al. [9] presented an approach for programmed estimation arrangement of twitter messages. Individual of question term messages were delegated negative or positive. Here creators utilize far off supervision to show the consequences of estimations of twitter posts with the assistance of the machine learning calculations. The calculations, for example, Maximum Entropy, SVM and Naïve Bayes are connected to preparing information which contains emojis, gave exactness above 80%. The creators likewise talk about pre-processing steps that was gotten higher precision. The creators concocted a thought for inaccessible managed picking up utilizing tweets that contain emojis.

Christianini and Taylor [10] distributed and shared the information about SVM which is machine learning calculation. The creators figure out how to give profound comprehension about calculation and how to approach the SVM calculation so as to actualize it to tackle the down to earth issues. The methodology will be hypothetical as when the book was distributed, the exploration was on going on each field.

Burger et al. [11] Since, In this period the PC have turned out to be sufficient amazing that can deal with extensive scale application which gives design acknowledgment and factual estimation of certifiable issues. The creators presented a n approach for measurable demonstrating dependent on most extreme entropy. By utilizing instances of issues in common language preparing, the creator demonstrates most extreme probability system for programmed development of greatest entropy models. Here the creators portrayed the standard of most extreme entropy. This rule chooses the model with most noteworthy entropy among all the reliable models. By augmenting the probability of preparing information we can acquire ideal estimations of given parameters.

Romero et al. [12] found that hash tags turns into the normal component of twitter utilized in each message and new terms are made and changing on everyday schedule which impacts the general importance of the first term. The creators likewise discovered basic contrast among issues and get familiar with the structure of generally utilized distinctive kinds of hash tags. The creators likewise created generative and recreation based models to examine the connection between plan of most recent adopters on which hash tag extends and appropriation elements.

III. Aims & Objectives

The main aim of the thesis is to compare the results that are implemented with the help of supervised classifier

The methodology followed is:

1. We‟ve compiled a body of positive, negative, and neutral tweets using the Twitter4j Twitter Twitter API. The size of our body can be extremely large.
2. Secondly, we have removed the empty words from the compiled corpus so that the content is free of commas, complete stops, and so on.

3. Next, we first apply the machine learning algorithms to our training set, and then we test and compare the results.
4. Using the results, we evaluate which automatic learning algorithm is best for classifying opinion analysis.

The objective of the thesis is discussed in the following points:
1. Explore, analyze and study the existing detection analysis techniques in the online micro blogging network.
2. Learn how to generate tweets from Twitter using the Python API.
3. Implement and analyze the results obtained after the application of the
4. Learning classifiers for the dataset.

IV. Methodology

1. **Collection of data:** We collect data from the Twitter API called Twitter4j using Netbeans. Searches are done using #Hashtag followed by the name of the movie, such as #FAN, #Gully Boy, # URI-The Surgical Strike, and so on. Some 23,000 tweets from several movie tweets have been compiled. Critics can also be found on #Hash labels, followed by movie stars, directors, production companies and respective record companies. On Twitter, hash tags become the symbol needed to find something and offer the user a limit of 120 words to express his point of view and attitude.
2. **Normalization:** We have discovered that to get the desired results from the classifier, we need to make sure that the tweets can be processed correctly. Since tweets can be in the user's language, we need to clean up unrelated data. The following items that may not be relevant to the data are:
   URL: The URLs of the message will be meaningless because they simply distract the result from the classifier.
   User Name: Deleting the user name may be necessary for cleaning purposes as it may falsely affect our results.
3. **Repetitive Words:** If the character is repeated more than twice, it may include a new word, but the meaning is the same. Therefore, we must eliminate this word and make it authentic. For example, you can write goooooood. If the message contains a word that has appeared more than twice in a row, it must be changed twice. For example, a good movie can become a good movie.
4. **Elimination of stop words**: Stop words are words such as "a", "is", "el", "etc.", etc. These words have nothing to do with emotion, so they must get rid of the message. The next step is to train the data using a supervised classifier.
5. **Machine learning techniques:** We use polarity- based classification methods that use a set of positive, negative, and neutral tweets provided by the Twitter4j API. Polarity is given by the probability relation that a word appears in a set of positive or negative affirmations that make the word positive or negative. The classifiers we use are based on the concept of polarity.

$$\text{Polarity} = \frac{P(\text{Postive\_Words})/P(\text{Total\_Words})}{P(\text{Negative\_Words})/P(\text{Total\_Words})}$$

If the feature is independent and based only on Standard English Dictionary then only this technique works. This method fails when we tried to record the opinion shown with respect to comparison. Further, the polarity based technique also fails to record query related opinion. In order to fulfill the requirement of classification we involved machine learning techniques.
The machine techniques comprised of following supervised classifier that are given below:-

- Naïve Bayes
- Support Vector Machines (SVM)
- Maximum Entropy (MaxEnt)

6. Supervised Classifiers:
Naïve Bayes*:* The Naïve Bayes classifier in one of the simplest probabilistic model works positively on text categorization and employed on Bayes rule with self-supporting feature collection [3] works positively on text categorization and employed on Bayes rule with self-supporting feature collection [3]. It is flexible in way of handling with any number of classes or attributes. For a given tweet d, C* is a class variable which defines the opinion given by $C* = argMaxcPNB(C|D)$
Bayes Probabilty PNB (C|D) described as

$$PNB\ (C|D) = \frac{(P(c)\ \square\ P(f|c)ni(d)\ mi\ =1\ )}{P(d)}$$

Support Vector Machines: SVMs are happening to be extremely accomplished at text categorization, widely outperforming Naïve Bayes (Joachims 1998). We examined big margin classifier to attain effective accuracy of classification process [9]. SVMs uses a function called kernel which are machine learning classification methodology in which the data is not separable linearly in the new area which it is to locate to area of data points, with allocation for classification of erroneous.

Support Vector Machines are the members of the family of classifiers which are linear. The main objective of the linear classifier is to find a hyperplane which is linear in nature of a feature area that divides all other entities in form of two classes. The main function of the SVMs is find out the hyporplane which is separating that has distance maximum from the nearest points to feature area in it.

Searching hyperplane in sample of linear separable, the equation can be consider as problem of optimization:

½ □ω □2 □ min (ω, b)

yi (wTxi + b) □ 1,j = 1, … , m,

Maximum Entropy*:* MaxEnt is another classification technique widely used lot in applications of natural language processing [10]. MaxEnt not always but sometimes outperforms the Naïve Bayes classifier for text classification [11]. MaxEnt is the most uniform model prefer for the classification purpose [12].

In the scenario of 2-class, to search for distribution over the both classes it is likely the same thing as using the logistic regression. Regarding independence of feature, it does not make any assumption. Due to this we can add features and phrases such as bigrams and to MaxEnt without affecting overlapping of the features. Let"s take an example in which we have two features such as "good" and other one is "very good" , then in case of Naïve Bayes their probabilities will be taken as independent even when the both of this are overlapping but not in case of MaxEnt. The equation for this model can be given as:

P(C|D) = 1 □Z (d)exp (□i λi,cFi,c(d,c))

Here, c indicates class, d indicates a single tweet, λ indicates vector of weight, normalization function Z(d), $Fi,c\varphi$ is aclass/feature function for class c and feature fi defined as follows :

fi,c (d, c □) = { 1, ni (d) > 0 and c □= c 0, ot erwise

Performance Measure: To calculate the accuracy of classifier we required measure on which accuracy can be obtained. There are two measures on which accuracy can be dependent:

- Precison
- Recall
- Accuracy

V. Implementation

1. Installation:

Tweepy: It is the python client for the official Twitter API. Install it using following pip command:

pip install tweepy

TextBlob: It is the python library for processing textual data. Install it using following pip command:

pip install textblob

Also, we need to install some NLTK corpora using following command:

python -m textblob.download_corpora (Corpora are nothing but a large and structured set of texts.)

2. Authentication:

In order to fetch tweets through Twitter API, one needs to register an App through their twitter account. Follow these steps for the same:

- Open the link and Create New App
- Fill the application details. You can leave the call-back url field empty.

- Once the app is created, you will be redirected to the app page.
- Open the „Keys and Access Tokens‟ tab.
- Copy „Consumer Key‟, „Consumer Secret‟,

„Access token‟ and „Access Token Secret‟.

We follow these 3 major steps in our program:

- 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.

3. Data Extraction:-

First of all, we create a Twitter Client class. This class contains all the methods to interact with Twitter API and parsing tweets. We use__init__function to handle the authentication of API client**.**

In get_tweets function, we use:

fetched_tweets = self.api.search(q = query, count = count)

- *to call the Twitter API to fetch tweets.*
- *In **get_tweet_opinion** we use textblob module.*

analysis = TextBlob(self.clean_tweet(tweet)) TextBlob is actually a high level library built over top of NLTK library.

4. **Pre processing: -** First, we call the clean_tweet method to remove links, special characters, and so on. Tweet using simple regular expressions.
5. Then, when we pass the tweet to create a TextBlob object, the following process is executed on the text using the textblob library:
6. Tokenize the tweet, that is, divide the words in the body of the text.
7. Eliminate the empty words of the chips. (Stop words are commonly used words that are not relevant in the analysis of a text, such as I am, you, are, etc.)
8. Label the tokens (part of the voice) and select only significant features / tokens such as adjectives, adverbs, etc.
9. Pass the tokens to an opinion classifier that classifies the opinion of the tweet as positive, negative, or neutral by assigning a polarity between -1.0 and 1.0.
10. Here's how the feelings classifier is created:
11. TextBlob uses a set of movie review data in which revisions have already been tagged as positive or negative.
12. The positive and negative characteristics are extracted from each positive and negative examination, respectively.
13. Training data now consists of positive and negative tagged features. These data are formed in a Naive Bayes classifier.
14. Then, we use **sentiment.polarity** method of **TextBlob** class to get the polarity of tweet between -1 to1.

Then, we classify polarity as:

{
if analysis. sentiment. polarity > 0: return 'positive'
else if analysis. sentiment. polarity == 0: return 'neutral'
else:
return 'negative'}

Finally, parsed tweets are returned. Then, we can do various type of statistical analysis on the tweets. For example, in above program, we tried to find the percentage of positive, negative and neutral tweets about a query.

VI. Results

We are using Python language for implementation. Python language offers maximum support when it comes to machine learning techniques. Machine learning techniques can be easily implemented in Python language.

***Fig: 1** Initial Console System Interface*

***Fig 2:** During Execution of the algorithm.*

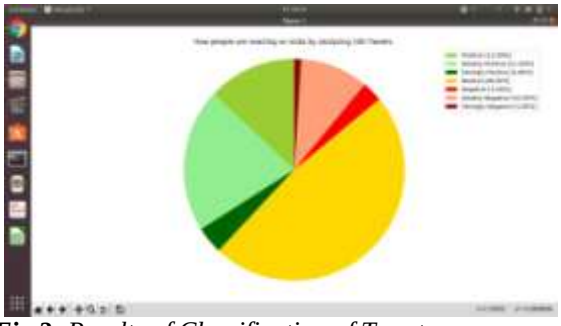

***Fig 3:** Results of Classification of Tweets*

***Table 1:** Precision and recall*

Algorithm	Performance Measure	
	Precision	Recall
Naïve Bayes	0.78	0.76
Support Vector Machines	0.87	0.73
Maximum Entropy	0.74	0.71

VII. Conclusion

In this thesis, we performed a comparative analysis of supervised classifiers such as Naïve Bayes, support vector machines, and maximum entropy using the unigram, bigram, and hybrid functions (unigram + bigram). It is necessary to conduct opinion analyzes in the form of texts in the form of messages or publications to determine whether the opinion is negative, positive or neutral. We had extracted data from Twitter, that is, movie reviews for the prediction of opinions using machine learning algorithms. We first extracted the twitter data using the twitter API. Then, during preprocessing, we clean up the data and make it available for training using classifiers. We compiled 15,000 tweets for the training set and 2,000 tweets for the test set. The SVM that uses the hybrid function surpasses all other classifiers and the selection function with an accuracy of 87%, while Max Ent exceeds Naïve Bayes with the bigrama function. MaxEnt, in some datasets, gives better results than Naïve Bayes. It is concluded that SVM performs better than other classifiers.

References

[1]. **B**. Pang, L. Lee and S. Vaithyanathan, "Thumbs up: opinion classification using machine learning techniques", *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pp. 79--86, 2002.

[2]. L. Jiang, M. Yu, M. Zhou, X. Liu and T. Zhao, "Target-dependent twitter opinion classification", *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, pp. 151--160, 2011.

[3]. C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou and P. Li, "User-level opinion analysis incorporating social networks", *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1397--1405, 2011.

[4]. L. Chen, C. Liu and H. Chiu, "A neural network based approach for opinion classification in the blogosphere", *Journal of Informetrics*, vol. 5, no. 2, pp. 313-322, 2011.

[5]. M. Anjaria and R. Guddeti, "Influence factor based opinion mining of Twitter data using supervised learning", *Communication Systems and Networks (COMSNETS), 2014 Sixth International Conference on*, pp. 1--8, 2014.

[6]. A. Barhan and A. Shakhomirov, "Methods for Opinion Analysis of Twitter Messages", *12th Conference of FRUCT Association*, 2012.

[7]. A. Pak and P. Paroubek, "Twitter as a Corpus for Opinion Analysis and Opinion Mining.", *LREc*, vol. 10, pp. 1320-- 1326, 2010.

[8]. M. Koppel and J. Schler, "THE IMPORTANCE OF NEUTRAL EXAMPLES FOR LEARNING OPINION", *Computational Intell*, vol. 22, no. 2, pp. 100-109, 2006.

[9]. A. Go, R. Bhayani and L. Huang, "Twitter opinion classification using distant supervision", *CS224N Project Report, Stanford*, vol. 1, p. 12, 2009.

[10]. A. Andrew, "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods20016Nello Christianini and John Shawe-Taylor. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, 2000. xiii + 189 pp., ISBN: ISBN 0-521-78019-5 Hardback: £27.50", *Kybernetes*, vol. 30, no. 1, pp. 103-115, 2001.

[11]. A. Berger, V. Pietra and S. Pietra, "A maximum entropy approach to natural language processing", *Computational linguistics*, vol. 22, no. 1, pp. 39--71, 1996.

[12]. K. Nigam, J. Lafferty and A. McCallum, "Using maximum entropy for text classification", *IJCAI-99 workshop on machine learning for information filtering*, vol. 1, pp. 61--67, 1999.

[13]. D. Romero, B. Meeder and J. Kleinberg, "Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on Twitter", *Proceedings of the 20th international conference on World wide web*, pp. 695--704, 2011.

[14]. [14]S. Tan and J. Zhang, "An empirical study of opinion analysis for Chinese documents", *Expert Systems with Applications*, vol. 34, no. 4, pp. 2622-2629, 2008.

[15]. [15]J. Martineau and T. Finin, "Delta TFIDF: An Improved Feature Space for Opinion Analysis", *Proceedings of the Third International ICWSM Conference*, vol. 9, 2009.

[16]. FA. Nielson. "A new ANEW: Evaluation of a word list for opinion analysis in microblogs", *arXiv preprint arXiv:1103.2903*, 2009.

[17]. SM. Mohammad, S. Kiritchenko and X. Zhu, "NRC- Canada: Building the State-of- the-Art in Opinion Analysis of Tweets", *arXiv preprint arXiv:1308.6242*, 2013.

[18]. [18]E. Kouloumpis, T. Wilson and J. Moore, "Twitter Opinion Analysis: The Good the Bad and the OMG!", *ICWSM*, vol. 11, pp. 538—541, 2011.

[19]. [19]K. Denecke, "Using SentiWordNet for Multilingual Opinion Analysis", *Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on IEEE*, pp. 507—512, 2008.

[20]. [20]B. Gokulkrishnan, P. Priyanthan, T. Ragavan, N. Prasath and A. Perara," Opinion Mining and Opinion Analysis on a Twitter Data Stream", *Advances in ICT for Emerging Regions (ICTer), 2012 International Conference IEEE*, pp. 182—188, 2012.

[21]. [21]F. Neri, C. Aliprandi, F. Capeci, M. Cuadros and T. By," Opinion Analysis on Social Media ", *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining*, pp.919—926, 2012.

[22]. [22]T. Wilson, J. Wiebe and P. Hoffman, "Recognizing Contextual Polarity: An Exploration of Features for Phrase- Level Opinion Analysis", *Computational Lingusitics, MIT Press*, vol. 35, no. 3,pp. 399—433, 2009.

[23]. [23]N. Godbole, M. Srinivasaiah and S. Skiena," Large-Scale Opinion Analysis for News and Blogs", *ICWSM*, vol. 7, no. 21, pp. 219—222, 2007.

[24]. [24]F. Benamara, C. Caserano, A. Picariello, DR. Recupero and VS. Subrahmanian," Opinion Analysis: Adjectives and Adverbs are better than Adjectives Alone", *ICWSM*, 2007. [25]D. Boyd and N. Ellison," Social Network Sites: Definition, History, and Scholarship", *IEEE Engineering Management Review*, vol. 3, no. 38, pp. 16—31, 2010.

[25]. [26]A. Agarwal, B. Xie, I. Vovsha, O. Rambow and R. Passonneau, "Opinion Analysis of Twitter Data", *Proceedings of the workshop on languages in social media, Association for Computational Linguistics*, pp. 30—38, 2011.

[26]. [27]T. Nasukawa and J. Yi, "Opinion Analysis: Capturing Favorability Using Natural Language Processing", *Proceedings of the 2nd international conference on Knowledge capture, ACM*, pp. 70—77, 2003.

[27]. [28]H. Wang, D. Can, A. Kazemzadeh, F. Bar and S. Narayanan," A System for Real- time Twitter Opinion Analysis of 2012 U.S. Presidential Election Cycle", *Proceedings of the ACL 2012 System Demonstrations, Association for Computational Linguistics*, pp. 115—120, 2012.

[28]. [29]T. Wilson, J. Wiebe and P. Hoffman,"Recognizing Contextual Polarity in Phrase- Level Opinion Analysis", *Proceedings of the conference on human language technology and empirical methods in natural language processing, Association for Computational Linguistics*, pp. 347—354, 2005.

[29]. [30]H. Kanayam and T. Nasukawa, " Fully Automatic Lexicon Expansion for Domain- oriented Opinion Analysis", *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics*, pp. 355—363, 2006.

[30]. [31]Y. Choi and C. Cardie, "Learning with Compositional Semantics as Structural Inference for Subsentential Opinion Analysis", *Proceedings of the Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics*, pp. 793—801, 2008.

[31]. [32]P. Melville, W. Gryc and RD. Lawrence, " Opinion Analysis of Blogs by Combining Lexical Knowledge with Text Classification", *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM*, pp. 1275—1284, 2009.

[32]. [33]G. Paltoglou and M. Thelwall, "A study of Information Retrieval weighting schemes for opinion analysis", *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics*, pp. 1386—1395, 2010.

[33]. [34]J. Fernandez, Y. Gutierrez, J. Gomez and P. Martinez- Barco, " GPLSI: Supervised Opinion Analysis in Twitter using Skipgrams", *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), number SemEval*, pp. 294—299, 2014.

[34]. [35]T. Mullen and R. Malouf, "A Preliminary Investigation into Opinion Analysis of Informal Political Discourse", *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, pp. 159—162, 2006.

[35]. [36]A. Harb, M. Plantie, G. Dray, M. Roche, F. Trousset and P. Poncelet, "Web opinion mining: How to extract opinions from

blogs?", *Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology, ACM,* pp. 211—217, 2008.

[36]. [37]SM. Kim and E. Hovy, "Determining the Opinion of Opinions", *Proceedings of the 20th international conference on Computational Linguistics, Association for Computational Linguistics,* 2004.

[37]. [38]A. Martalo, N. Novielli and F. De Rosis," Attitude Display in Dialogue Patterns",*Proc. AISB 2008 Symposium on Affective Language in Human and Machine,* 2008.

[38]. [39]E. AL-Daoud, "Integration of Support Vector Machine and Bayesian Neural Network for Data Mining and Classification", *World Academy of Science, Engineering and Technology,* vol. 64, 2010.

[39]. [40]K. Yessenov and S. Misailovic, "Opinion Analysis of Movie Review Comments", *Methodology,* pp. 1—17, 2009.