

Review: Sentiment Analysis using SVM Classification Approach

Shweta V. Raut¹, Madhu M. Nashipudimath²

Student, Department of Information Technology Engineering, Pillai College of Engineering, Panvel, India¹

Assistant Professor, Department of Computer Engineering, Pillai College of Engineering, Panvel, India²

Abstract: Recently, lots of attempts are done to work on social sites to examine of public sentiment. Most of the efforts are usable to give fine ideas of social public opinions from social media. Hence, there is a need of suitable approach to overcome this problem. Sentiment Analysis (SA) is an action of computationally diagnosing and grouping opinions represented in a particular bunch of text. It is used to recognize opinion of public as feedbacks depending upon the data/domain in social media. Information Gain (IG) is a measure used to identify most impactful words as features in the tweet to classify the opinions using some classification approaches. The purpose of this article is to discuss some approaches for extracting features from tweets and classifying it.

Keywords: Feature Selection, Support Vector Machine(SVM), Information Gain(IG), Sentiment Analysis, Twitter.

I. INTRODUCTION

The concept of Sentiment Analysis (SA) was evolved in 20th century [1]. It is a thoughtful stage to rescue user's concerns and desires from the tweets. With emerge of cyberspace based applications has escorted abundant personified reviews for various related knowledge on the Web [2]. These reviews occur in different forums like social applications, blogs and search engine websites. Both tourists and clients find the judgment in these reviews to be favorable for their understanding and planning actions. The explosion of search portals like Yahoo and Google has overwhelmed the users with ample amount of relevant reviews about particular intentions. SA postures as a supreme tool for users to extricate necessary information, as well as to combine the collective sentiments of reviews. Several approaches and tactics have been come into spotlight. It is hence a challenge to recognize which techniques are most potent at pre-processing, feature selection, feature extraction, classification stages and to also choose relevant dictionaries for correctly mapping terms from tweet to provide exact result. Dealing with abundant data gives rise to the term called "Big data".

Big data is a large quantity, pace and diversity of knowledge service that requires profitable, ingenious scheme of data processing to boost vision and decision making [3]. It is a matter of phase that brings enormous heterogeneous [4] information together. It aspired to build platforms and tools to ingest store and analyze data that can be characterized by high dimensionality, large sample size with variety and possibly fast (velocity) changing. The information technology industry, academia, and governments have made lot of efforts to analyze data sources. In today's world, Big data is used, studied and analyzed for many purposes such as system log analytics, social media analytics i.e. sentiment analysis for risk management, customer service, brand management, segmentation and prediction, Churn prediction, operational analytics, recommender system and target marketing, etc.

Big data, in general, are grouped into three types of data [5]: Regular data, Semi-Regular data, and irregular data (Fig. 1) [6]. This is called as structural diverseness in Big data. Only 5% of existing data is regular [7]. Regular data is tabular data which is represented in the fashion of relative data store and excel sheets. Data having no specific composition and schema is known as "Irregular", it can be in any structure such as text, sound, photocopy, promotional film, e.g., document files, portable document format files, context of tweets, forums, emails, search pages, audio files, video files, pictures etc. The online network is overwhelmed with irregular data. Around, 95% of existing data is irregular in the form of videos [8]. Regular data is easy to operate upon where as irregular data requires miscellaneous scrutiny [9]. In between regular and irregular data, there exists semi-regular data which has no authorized format e.g., Extensible Markup Language(XML) files, search logs, sensor logs, etc. XML allows interchanging the information over internet and is machine understandable since it comprises of user-defined data IDs/chips.

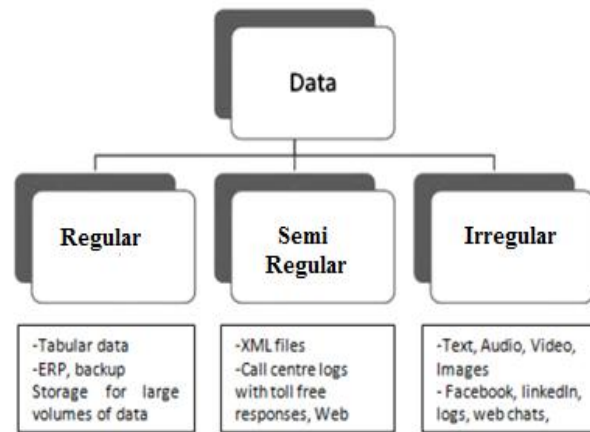


Fig. 1- Types of Big data

Sentiment Classification levels are of different levels: testimony/credential level, statement level, or factor/quality level [10].

Sentiment Analysis General Block diagram

Sentiment Analysis (SA) is a study of people's notions, philosophy and ecstasy towards a subject.

The intension of SA is to investigate viewpoints, observe the belief it intend, and then compute the terms lying in different classification classes; e.g. positive, negative, neutral, etc. as shown in Fig. 2.

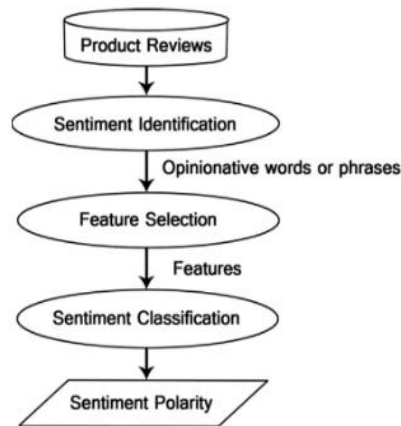


Fig. 2- Sentiment Analysis process on product reviews

Explanation:

1] **Product Reviews:** Firstly, when anyone posts a particular subject related Blog, then users using that particular application posts their reviews (opinions or comments) on it.

2] **Sentiment Identification:** The reviews posted are analyzed and the terms i.e. feature belonging to which domain is identified.

3] **Feature Selection:** Pre-processing algorithms will be applied in order to remove unwanted words (terms / features) like stopwords, prepositions, vovals, etc. from the different reviews posted. Different feature selection algorithms are Gini Index, Information Gain, etc. to find out most impactful word ie. term or feature to eventually form proper dictionary to obtain accurate result when provided as an input to classifier.

4] **Sentiment Classification:** Classification algorithms like Naïve Bayesian, SVM, KNN , etc. are used to obtain accurate and correct result ie. terms or words are classified in accurate positives or negatives classes to determine proper opinion about a particular subject from different reviews posted.

5] **Sentiment Polarity:** The result is presented in proper classes respectively like in positive, negative class, etc. after applying classification algorithm that can provide correct result.

II. RELATED WORKS

Mining social views and investigating them on twitter data has contributed to manifest way to unmask social desires, which comforts for decision making in heterogeneous subjects. Twitter is crucial and favored platform for users interplay/communication. By using twitter platform, number of people shares their intuitions, thoughts and judgments. For making crucial judgments, it is vital to drill out social print of views and to investigate root causes behind diversity of sentiment is worthy.

This article unfolds the analysis of the efficacious weight by Particle Swarm Upgrade and Ant Colony Upgrade tactics with discriminative classifier of Support Vector Machine (SVM) and Naïve Bayes (NB). Result shows SVM_PSO doesn't perform well as compared to that of NB, which is 86.29% correctness because NB iteratively change the threshold if weight is different for different keywords [11]. The correctness level can be optimized even more by taking into account of emoticons for the classification of input data as it can resolve a lot for correct classification categorization and also by using other optimization methods along with classifiers.

Aldo Hernandez-Suarez, et. al. [12], this paper proposed a Sociable Conviction Sensor in Twitter by gathering ancient tweets. After computing three different classification methods, the computed result shows that the Maximum Entropy provides most potent results than NB and SVM for negative, positive and safety-based tweets. The use of ℓ_1 regulation assisted to upgrade the assessment of excellent forecast of coefficients, which the Sociable Conviction Sensor could achieve about 80% of correctness between cyber-attack and Twitter communal ambience. The application of proposed method is not restricted to cyber-attacks, as it aims to reconstruct the proposed method to forecast events such as in epidemic, diagnosis, constitutional arrangement and retail scrutiny.

It is speculated that the machine learning method is very potent and simple than regular methods. Twitter data analysis is tough because it is very onerous to identify empathic words from tweets and also due to repetition of characters existence, blank spaces, argot words, miswriting, etc. To handle these issues feature vector is created. Before creating feature vector, pre-processing is carried out on each and every tweet. Then features are imitated in two phases: First phase is to deduce twitter-related words. Then these words are deduced from the text. Now deduced feature vector is converted into regular text. After that, features are deduced from tweet which are normal text without any hash tags. And the deduced features are then supplemented to model feature vector. There are heterogeneous machine leaning classifiers to categorize the tweets. From the results, it is demonstrated that NB, Maximum Entropy and SVM operates better and also gives immense correctness. Firstly, Maximum Entropy method operates excellent than other methods of having about 90% of accuracy. Secondly, SVM classifier operates better than NB of having accuracy of about 83.2%. Thirdly, NB operates least efficiently than other methods having accuracy of about 64.2%. By this paper, it is concluded that feature vector operates well for Game reviews tweets [13]. The other similar paper possessing same procedure produces 75% correctness from SVM and about 65% correctness from NB classifier when it is applied over Movie reviews tweets [14]. The correctness of classification for both researches can be upgraded by upgrading training data.

The article deals to envision which tweets are within the #metoo act that would be linked or related with by other clients [15]. It gathered about three thousand tweets, analyzing the description of each by word frequencies. Though it diagnosed with a regular lexicon and their own personalized vocabulary to the dataset. It experimented the results using NB and SVM, training the method to envision whether a given tweet would be retweeted more than 450 times. SVM outperforms NB slightly at 90.58% correctness on test data. It described discussions related to ministry, the enterprise, and the ladies who presented themselves to bring bond in the act. In future, the project can be upgraded in three ways: traversing working hours within the act, applying other techniques to the data and the own proposed technique on the other societal acts. In one-month of time, they collected data and interactions in the #metoo tag that transferred many times. It could help to preferably know the vast templates of it whereas to diminish the deviation of individual communications. Project can be extended by acquiring human population knowledge.

The paper examined the pursuance of SVM for SA [16]. The pursuance diagnosis of SVM is calculated using two datasets; individual driving motors and apple brand goods. Weka tool is used for pursuance examination and differentiation. Outputs are obtained in forms of precision, recall and f-measure. Firstly, the datasets obtains highest recalls. Secondly, f-measure is greater than calculated precision for first dataset and precision is greater than f-measure for second dataset. The outputs thus, presented the reliancy of SVM pursuance and other machine learning methods. This paper can be used as a basic standard for comparable diagnosis.

Ankit Pradeep Patel, et. al. [17], it explained the methods for preprocessing and information retrieval of tweets from Blogs. SA is significant element in judgment making measure in specific subject. Also, it presented a study of SVM that is used for text association which can be utilized to investigate correct classification of tweets. SVM yields some features like; Large Dimensional properties space that can exclude the ought of feature selection. The pursuance of

SVM can be measured utilizing precision and recall. This paper describes that SVM outperforms than ANN for text classification.

The article represents a set of tests on opinion categorization in societal network text using hybrid knowledge techniques [18]. Each initial beginner in combination used meta-level feature eradication. The features enclosed a fusion of various present vocabulary and mechanism assets for SA. Moreover, characteristics hashing was utilized for tweets description. The observation depicts that such fusion of categorization method can reduce the flaw estimation by bypassing low choice from the self-reliant classifiers that is an efficient action of assuring constancy. Utilizing meta-level characteristics minimized issues related with inadequacy of information. In that circumstance, the meta-level fusion technique can attain guaranteed results. It is concluded that the designed method can be pertinent to other societal forums scrutiny and either classifier can simply be merged to form designed architecture. The classification effort can be upgraded by assessing various fusion of techniques (polling patterns) or by taking into account other vocabulary assets and techniques in SA to increment classifier pursuance.

Pooja C. Sangvikar [19], concluded that range and opportunity to expand in field of information retrieval due to frequently developing and enlarging of cyberspace medias all over the world. Moreover, widening of SA and decision mining is volunteered as intriguing field of study because of its various operations over cyberspace medias. The audit describes decision mining methods that have lots of issues like correctness, ability to expand, format of information domain reliancy, etc. These issues should be handled and those solutions can be used to upgrade the techniques to do SA and sentiment categorization.

The article explained the use of maxEnt package in R which is uniquely invented to reduce storage utilization on so many huge datasets like the scattered record-term grid created in text mining [20]. To know which characteristics are significant in upgrading the correctness of designed MaxEnt classifier, it used an expulsion method. It showed the outcome after removing a particular engineered featured for e.g.- lexicon score or emoticon mapping or slang mapping on the accuracy of the classifier used. It find that tweaking the characteristics goes a long way in improving the accuracy. The accuracy for a classifier that uses all these features is significantly higher (97.35%) compared to a classifier that doesn't use any of these characteristics (96.9%). Lexicon score removal leads to the maximum decline in accuracy thus it can gauge that it is indeed an important feature for our analysis. It have only considered unigrams as it did not obtain any significant change in the accuracy when bigrams/trigrams been considered. This could be perhaps due to the restriction on the length of the tweet for these characteristics to be useful.

The SA process is improved by identifying influential user in social media network [21]. It is determined by clustering the words in the comments using SOM clustering. It clusters the words based on the consideration different measures like word frequency, SentiNet and network based properties. Then in the constructed network the information modules are pruned to identify the influential user through module average opinion and user average opinion measures. The influential users enact significant character in SA. The proposed method identifies the influential user effectively using different measures and techniques. The experimental results show that the designed method is superior to other method in forms of accuracy, precision and recall.

The research deals with better SA system with an improvement of Information Gain (IG) based feature selection and classification. The proposed feature selection method selects feature that possesses high information gain and high occurrence. As a result, it accomplished in providing feature that most probably appears in testing also. Proposed classifier utilized the positive and negative terms obtained from the IG calculation which is faster than the previous classifier (SVM, ANN, etc.). The hybrid method of information gain and document frequency in this study proposed feature selection; IGDFFS selects subfeatures that satisfy these criteria: (1) high relevance to the output class and (2) high occurrence in dataset. As a result, it constructs subfeatures that reach better performance in the classification. Compared to the present classifier, Information Gain Classifier (IGC) overcomes the recent high accuracy which belongs to EWGS (only 88.05%). It succeeded in avoiding overfitting problems in any condition. The performance of IGC is quite stable in both training and testing. We are considering to group the words based on their relevance to positive and negative reviews. Note that British domain consists of 171,476 words at present been utilized and 47,156 of ancient words (based on Oxford English Dictionary). At least a finite number of groups would be less than the total number of words.

It presented a design of categorizing tweets that was invented in the research of edible cost inflation in Indian merchandise [23]. Twitter API was utilized for scanning the tweets. The scanned tweets were filtrated for significant knowledge and hoarded in a datastore. The various phases of pre-processing were operated on it and the tweets were eliminated from special symbols, stopwords, tokens, etc. Stemming was performed to all terms to eliminate the base terms. TF-IDF calculation related method was been employed and was computed for every tweets. Characteristics can be chosen using Chi Square test and IG. The extricated characteristic constructs a record-term grid that is applied in categorization technique. The output obtained is excellent and is concluded that KNN is superior as compared to NB technique. A robotic setup is fabricated for decision mining in this research.

This paper [24], describes filtration of each characteristics utilizing IG. Best result is obtained using IG when each separate characteristic are considered than considering all characteristics. In addition, training SVM with 1% of

characteristics gives only 20% of duration necessary to train it when utilized all characteristics. Efficiency of Cognitive synonyms and categorizing-grammar is also shown. Further, the group of characteristics can be upgraded to involve distinguish of grammar based characteristics over n-gram related characteristics.

It presented technique used characteristics study and the Adaptable Framework for Artificial Neural Networks (AAN2) for quality-integrated Twitter SA [25]. The technique explained drawbacks related with the uncommon features of the Twitter language and the recall of gentle emotional descriptions of predominant attention to quality-integrated professionals. It exposed the potency of the method on a Starbucks quality-related Tweets dataset including of 5,526 tweets. The characteristics study designed a terminating tweet characteristic structure involving of only seven measurements, with high characteristics quantity. The analysis represented comparison of designed method with pursuance of two event-of-skill tweets examination structure of scholastic and financial subjects. The outcome showed the designed technique performs better as compared to the event-of-skill design in both three-class and five-class tweet emotions categorization correctness about 80% high and outstanding recall of gentle feelings. The AAN2 and SVM frameworks operated better as compared to event-of-skill framework, utilizing depiction reproduced from characteristic study. The AAN2 framework was remarkable than SVM framework when applied over all researches and emotional classes, in formats of categorization correctness and recall, with one expulsion of firmly negative class recall. AAN2 framework presented the awareness obligatory to correctly determine both gentle and firm emotional descriptions.

This article concludes the issue of review categorization [26]. It is investigated that unigram is a superior technique to extricate emotions from the tweets. For positive as well as negative class, unigram with stemming and without stopwords outperforms when applied over information gain (IG) method. In future, combination of feature selection methods can be applied over this work for more experimentation.

III. ISSUES IN SENTIMENT ANALYSIS

Abundant social information is growing day-by-day from online network and relates to predict stock price movements [27]. The information yields the beneficial knowledge like; vote-casting forecast, obtaining client's response related to any freeware, inspecting the retail prestige of distinct grade and attaining client idea prior to propel of a fresh invention, etc. Many techniques are accessible for SA like; related to vocabulary, machine learning and the composition of both.

People speaking variety of dialects always swaps when speaking over internet applications. Occasionally, initial alphabetical order of dialect is perpetuated, whereas utilizing a routine alphabetical order for all dialects is largely famous also because of usefulness. In this moments, diversified dialects need to be assorted along distinct regulations of grammar, utilizing similar alphabetical order that directs to confronting assignment for natural language processing also in case of correct emotion identification[28]. Experiments is executed on film reviews dataset possessing different dialects mixing function of two dialects like; English and Bengali, both typed in Roman alphabetical order that gives better results for English language but poor for both dialects when mixed while expressing one's vision.

The choice of characteristics can be upgraded to involve other n-gram characteristics like; bi-grams or tri-grams [29].

Criticism is a type of expression where users intend to speak unfavorable manner using affirmative or severe affirmative tweets in text. It is difficult to find effective solution for sarcastic reviews [30] in twitter to provide accurate classification result.

Result is not in presentable format. Using more no. of classes represents data in more precise manner [31].

IV. SUGGESTIONS TO SOLVE ISSUES

Following solutions open scope for further research:

- As huge amount of data is scattered over social networking sites, it becomes difficult it to handle. The solution for this problem is use of a unique grid-based confusion algorithm [32].
- Nearly all the setups are applicable for single-language but become unsuccessful while dealing with usage of mixture of two or more languages in the reviews like Bengali-English. The solution for this is using various calculation methods such as; Dialect-Mixed Index (DMI), Dialect-Mixed Factor (DF) involving many factors like, language and feelings [33].
- For more accurate results, n-gram feature can be expanded are explained along with examples for Support Vector Machine(SVM) Classifier [34] and Naïve Bayes(NB) Classifier[35].
- Sarcastic reviews on Twitter can be handled using a Motive-based method [36][37].
- To represent the obtained result in more precise format, more than two classes must be used [38][39].

- Proposed system [31] to be designed for Arabic language [40], work is still in process will end till 2022.
- In Airline system,[41] only texts need not be considered but also user who tweet them and also time of retweets. The solution for this is to use Twitter API [42].

There is a need to work on emoticons such as, ‘Happiness’, ‘Sadness’, ‘Anger’ and ‘Disgust’ [43]. Emoticon detection methods can be used as solution for such problem [44].

Table 1. Survey on Sentiment Analysis techniques

Literature	Solutions	Advantages	Constraints
LI Bing, Keith C.C. Chan, “A Fuzzy Logic Approach for Opinion Mining on Large Scale Twitter Data”, 2014, IEEE/ACM 7 th International Conference on Utility and Cloud Computing, Hong Kong.	It designed a Map-Reduce framework to parallelize a grid-based Confusion algorithm to extricate huge amount and types of Twitter emotions.	The framework attained a magnificent pursuance, including both rapid operating pace and great forecasting correctness as the number of records increases with time.	The designed framework is utilized to forecast only stock valuation act of record of enterprises, and has achieved high predictivity accuracy. The system can be used on other dataset reviews in future.
Soumil Mandal, Sainik Kumar Mahata, Dipankar Das, “Preparing Bengali-English Code-Mixed Corpus for Sentiment Analysis of Indian Languages”, arXiv:1803.04000v1 [cs.CL] 11 March 2018,Kolkata.	This paper provides a solution to deal with code-mixed language, i.e. Bengali-English with better results.	An aurous definitive Bengali-English dialect-mixed knowledge with language and categorization tag for SA is prepared.	The quality of the system can be improved by increasing the population size of the resources and training the classifiers on bigger data.
Vishal A. Kharde and S.S. Sonawane, “Sentiment Analysis of Twitter Data: A Survey of Techniques”, International Journal of Computer Applications (0975-8887), April 2016, Pune.	Paper presented to expand n-gram feature to provide better and accurate classifier results.	Variety of machine learning methods instructs the dataset with characteristics vectors and then the syntactical scrutiny produces a huge set of meanings and likeliness that gives out the different categorizations of tweets.	Designed framework can be upgraded using hybrid method of machine learning method and decision lexicon method to uplift the correctness of tweet categorization and flexible capability of various subjects and various dialects.
Mondher Bouazizi and Tomoaki Otsuki (Ohtsuki), (Senior Member, IEEE), “A Pattern-Based Approach for Sarcasm Detection on Twitter”, Japan, September 28,2016.	Designed a Motive-based method to discover criticism on Twitter.	Also designed an better way to upgrade with more group of irony templates utlizing previous training group of 6000 tweets, and hashtag “#sarcasm”.	The framework describes to obtain better outputs. In future, the pursuance of this framework can be upgraded using immense training set.
L. Almuqren and A. I. Cristea, “Twitter analysis to predict the satisfaction of telecom company customers,” in Late-breaking Results, Demos, Doctoral Consortium, Workshops Proceedings and Creative Track of the 27th ACM Conference on Hypertext and Social Media (HT 2016), Halifax, Canada, July 13-16, 2016. ---> work will be completed by 2022.	There is less work done on Arabic language for Airline reviews. Thus, proposed a system that provides a framework to work on Arabic language in Saudi Arabia.	Build a comprehensive Saudi dialect lexicon.	This work is in progress from Sept. 2015 to August 2022.

<p>Rucha Jadhavar, Agastya Kumar Komarraju, “Sentiment Analysis of Netflix and Competitor Tweets to Classify Customer Opinions”, Paper 2708-2018.</p>	<p>Analyzed tweets for Netflix and its competitors such as Amazon Prime Video, Hulu and HBO NOW.</p>	<p>Demonstrated the use of multiple SAS tools to analyze the tweets, generate summaries, identify different categories of tweets, and classify reviews.</p>	<p>The emoticons like: “Happiness”, “Sadness”, “Anger” and “Disgust”, can be identified using SAS Viya for integrating open source softwares to code in SAS environment. A daily visual report to identify daily trends, influencers and customer response to ad campaigns can be useful for marketing strategies. Can use multiple language dictionaries, scope can be increased to all languages used on twitter and capture a larger chunk of population.</p>
<p>Mamatha M, Thriveni J, Nenugopal K. R., “Techniques of Sentiment Classification, Emotion Detection, Feature Extraction and Sentiment Analysis: A Comprehensive Review”, International Journal of Computer Sciences and Engineering (IJCSE), E-ISSN: 2347-2693, 2018.</p>	<p>Provides a survey focuses on various methods for Feature Extraction and Emotion Detection.</p>	<p>Information regarding to require more work on multi-modal analysis and identification of emotions in music.</p>	<p>Can work on how emotional expressions changes over time or between genders or between ethnic groups. Can work on explicit features. Search techniques based on emotions is to be improved.</p>

Discussion highlights the capability of each techniques regarding SA. With this understanding, in future we are planning to propose an hybrid approach for feature selection and classification for twitter dataset. The model will attempt to provide an accurate result in positive and negative terms for sentiments. A basic Idea is to construct a Pattern Knowledge to work on code-mixed different domains tweets together and to provide better classification results. The tentative system is given below:

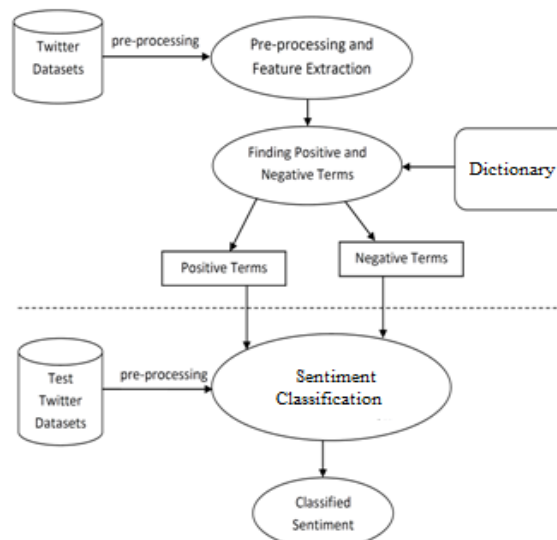


Fig. 3- Sentiment Analysis using SVM approach

V. CONCLUSION

Sentiment Analysis (SA) is transmuting the world in exciting ways, but it also brings new challenges in different level of data analysis. This discussion is focused on how to work and handle :

- 1] Huge data to day increasing data,
- 2] Multi-lingual code-mixing sentiments,
- 3] Dealing with n-grams features,
- 4] Dealing with sarcastic sentiments,
- 5] To deal with more than 2 classes as classifier gives more precise results.
- 6] Work on Arabic language, etc.

REFERENCES

- [1] Mika V. Mantyla, Daniel Graziotin, Miikka Kuutila, "The Evolution of Sentiment Analysis – A review of Research Topics, Venues, and Top cited Papers", Feb 2018, ISSN 1574-0137, <https://doi.org/10.1016/j.cosrev.2017.10.002>, University of Oulu, Finland.
- [2] Devika MD, Sunitha C, Amal Ganesh, "Sentiment Analysis: A Comparative Study On Different Approaches", Fourth International Conference on Recent Trends in Computer Science and Engineering, 2016, doi:10.1016/j.procs.2016.05.124, Chennai, Tamil Nadu, India.
- [3] Big data and data protection, 20140728 Version: 1.0
- [4] Michael, Katina, and Keith W. Miller. "Big data: New opportunities and new challenges [guest editors' introduction]." *Computer* 46.6 (2013): 22-24.
- [5] Mona Tanwar, Reena Duggal, Sunil Kumar Khatri, "Unravelling Unstructured Data: A Wealth of Information in Big Data", IEEE , 2015 , Noida, India.
- [6] Akash, Bryan, Nishi, Prashant, Subhadeep and Vaibhav,v "Emerging Marketing Analytics- Big Data." [Online]. Available: <http://www.slideshare.net/AkashTyagi8/big-data-marketinganalytics>
- [7] K. Cukier, "Data, data everywhere: A special report on managing information," *The Economist*, February 25, 2010.
- [8] Amir Gandomi and Murtaza Haider, "Beyond the hype: Big data concepts, methods, and analytics," *International Journal of Information Management*, vol. 35, Issue 2, pp 137-144, April 2015.
- [9] Bing Liu. *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, May 2012.
- [10] Ashish Katrekar, AVP, Big Data Analytics , "An Introduction to Sentiment Analysis", www.globallogic.com.
- [11] Prachi Bansal and Ramanjot Kaur, "Twitter Sentiment Analysis using Machine Learning and Optimization Techniques", Feb 2018 in *International Journal of Computer Applications*(0975-8887), Volume 179- No. 19, pp. 5-8, Doaba Institute of Engg. & Technology, Kharar, Punjab, India.
- [12] Aldo Hernandez-Suarez, Gabriel Sanchez-Perez, Karina Toscano-Medina, Victor Martinez-Hernandez, Hector Perez-Meana, Jesus Olivares-Mrcaso and Victor Sanchez, "Social Sentiment Sensor in Twitter for Predicting Cyber-Attacks Using ℓ_1 Regulation", March 2018 in *Preprints*(www.preprints.org), Instituto Politecnico Nacional, ESIME Culhuacan, Mexico City 04440, Mexico.
- [13] T.D.V. Kiran*, K. Gowtham Reddy, Jagadeesh Gopal, "Twitter sentiment analysis of game reviews using machine learning techniques", *Journal of Chemical and Pharmaceutical Sciences*, Print ISSN:0974-2115,2017, Tamil Nadu, India.
- [14] Akshay Amolik, Niketan Jivane, Mahavir Bhandari, Dr. M. Venkatesan, "Twitter Sentiment Analysis of Movie Reviews using Machine Learning Techniques", e-ISSN:0975-4024, *International Journal of Engineering and Technology (IJET)*, VIT University, Vellore-632014, Tamilnadu, India.
- [15] Ana Tarano (atarano) and Dana Murphy (dkm0713), "Tracking #metoo on Twitter to Predict Engagement in the Movement," 2017.
- [16] Munir Ahmad, Shabib Aftab and Ifikhar Ali, "Sentiment Analysis of Tweets using SVM", *International Journal of Computer Applications*, Nov 2017(0975-8887), Volume 177- No. 5, Pakistan.
- [17] Ankit Pradeep Patel, Ankit Vithalbhair Patel, Sanjaykumar Ghanshyambhai Butani and Prashant B. Sawant, "Literature Survey on Sentiment Analysis of Twitter Data using Machine Learning Approaches" in *International Journal for Innovative Research in Science & Technology*, ISSN:2349-6010, Volume 3, Issue 10, March 2017, Mumbai, India.
- [18] Alnashwan, Rana; O'Riordan, Adrian P.; Sorensen, Humphrey; Hoare, Cathal, "Improving sentiment analysis through ensemble learning of meta-level features", Sept. 2016, *Proceedings of the 2nd International Workshop on Knowledge Discovery on the WEB*, Cagliari, Italy, CEUR Workshop Proceedings, 1748.
- [19] Pooja C. Sangvikar, "A Survey on Sentiment Analysis and Opinion Mining", *International Journal for Scientific Research & Development*, Vol. 3, Issue 10, 2015, ISSN(online):2321-0613, Pune, India.
- [20] Ritwik Moghe, Pranita Khandelwal, Raju Bhattachayya and Sushant Rajput, "Studying Twitter Sentiment of Football Superstars", 2015, CDS Group-2.
- [21] P. Kalarani and Dr. S. Selva Brunda, "User Influential Level Based Sentiment Analysis using User Average Opinion (UAO) and Module Average Opinion(MAO)", *International Journal of Pure and Applied Mathematics*, Volume 118, No. 7, 2018, pp. 245-251, ISSN: 1311-8080(printed version); 1314-3395 (on-line version), Tamilnadu, India.
- [22] Asriyanti Indah Pratiwi and Adiwijaya, "On the Feature Slection and Classification Based on Information Gain for Document Sentiment Analysis", *Hindawi, Applied Computational Intelligence and Soft Computing*, Volume 2018, Article ID 1407817, Feb. 2018, Indonesia.
- [23] Shubham Goyal , "Sentiment Analysis of Twitter Dats using Text Mining and Hybrid Classification Approach", *International Journal of Advance Research, Ideas and Innovations in Technology*, ISSN: 2454-132X, Vol. 2, Issue 5, Bhwainigarh, Punjab, India, 2016.
- [24] Kim Schouten, Flavius Fransincar. And Rommert Dekker, "An Information Gain-Driven Feature Study for Aspect-Based Sentiment Analysis", Erasmus University Rotterdam, the Netherlands {schouten, fransincar, rdekker}@ese.eur.nl
- [25] David Zimbra, M. Ghiassi and Sean Lee, "Brand-related Twitter Sentiment Analysis using Feature Engineering and the Dynamic Architecture for Artificial Neural Networks", 2016, 49th Hawaii International Conference on System Sciences, Santa Clara University, IEEE Computer Society.
- [26] Shahana P.H, Bini Omman, "Evaluation of Features on Sentiment Analysis", *International Conference on Information and Communication Technologies (ICICT 2014)*, Kerala.Example when using et al.:
- [27] Munir Ahmad, Shabib Aftab, Muhammad Salman Bashir, Noureen Hameed, "Sentiment Analysis using SVM: A Systematic Literature Review", *International Journal of Advanced Computer Science and Applications* (www.ijacsa.thesai.org)", Pakistan, 2018.
- [28] Soumil Mandal, Dipankar Das, "Analyzing Roles of Classifiers and Code-Mixed factors for Sentiment Identification", 15 March 2018, arXiv:1801.02581v2 [cs.CL].

- [29] Harmani Mat Zin, Norwati Mustapha, Masrah Azrifah Azmi Murad and Nurfadhline Mohd Sharef, "The effects of pre-processing strategies in sentiment analysis of online movie reviews", 2017, American Institute of Physics, doi:10.1063/1.5005422.
- [30] Shruti Gupta, Ashutosh Pandey, Prof. K.K. Paliwal, "Sentiment Analysis of Twitter and Facebook Data Using Map-Reduce", IJEE, A UGC Recommended Journal, June 2017, Panipat, India, www.csjournals.com
- [31] Ali Mustafa Qamar, Suliman A. Alsubhany, and Syed Sohail Ahmed, "Sentiment Classification of Twitter Data Belonging to Saudi Arabian Telecommunication Companies", International Journal of Advanced Computer Science and Applications (IJACSA), 2017, Saudi Arabia.
- [32] LI Bing, Keith C.C. Chan, "A Fuzzy Logic Approach for Opinion Mining on Large Scale Twitter Data", 2014, IEEE/ACM 7th International Conference on Utility and Cloud Computing, Hong Kong.
- [33] Soumil Mandal, Sainik Kumar Mahata, Dipankar Das, "Preparing Bengali-English Code-Mixed Corpus for Sentiment Analysis of Indian Languages", Kolkata, arXiv:1803.04000v1 [cs.CL] 11 March 2018.
- [34] Gautami Tripathi and Naganna S., "Feature Selection and Classification Approach For Sentiment Analysis", Machine Learning and Applications: An International Journal (MLAIJ), June 2015.
- [35] Vishal A. Kharde and S.S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques", International Journal of Computer Applications (0975-8887), April 2016, Pune.
- [36] Mondher Bouazizi and Tomoaki Otsuki (Ohtsuki), (Senior Member, IEEE), "A Pattern-Based Approach for Sarcasm Detection on Twitter", Japan, September 28, 2016.
- [37] Mondher Bouazizi and Tomoaki Ohtsuki, "Opinion Mining in Twitter - How to make use of Sarcasm to Enhance Sentiment Analysis", 2015, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Japan.
- [38] David Zimbra, M. Ghiassi and Sean Lee, "Brand-Related Twitter Sentiment Analysis using Feature Engineering and the Dynamic Architecture for Artificial Neural Networks", 2016 49th Hawaii International Conference on System Sciences.
- [39] Daniel Jurafsky and James H. Martin, "Chapter 6: Naïve Bayes and Sentiment Classification", Speech and Language Processing, August 7, 2017.
- [40] L. Almuqren and A. I. Cristea, "Twitter analysis to predict the satisfaction of telecom company customers," in Late-breaking Results, Demos, Doctoral Consortium, Workshops Proceedings and Creative Track of the 27th ACM Conference on Hypertext and Social Media (HT 2016), Halifax, Canada, July 13-16, 2016. --> work will be completed by 2022.
- [41] Yun Wan and Dr. Qigang Gao, "An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis", Canada, 2015 IEEE 15th International Conference on Data Mining Workshops.
- [42] Anastasia Giachanou and Fabio Crestani, "Like It or Not: A Survey of Twitter Sentiment Analysis Methods", Svizzera Italiana, 12 April 2018, ACM Computing Surveys.
- [43] Rucha Jadhavar, Agastya Kumar Komarraju, "Sentiment Analysis of Netflix and Competitor Tweets to Classify Customer Opinions", Paper 2708-2018.
- [44] Mamatha M, Thriveni J, Nenugopal K. R., "Techniques of Sentiment Classification, Emotion Detection, Feature Extraction and Sentiment Analysis: A Comprehensive Review", International Journal of Computer Sciences and Engineering (IJCSSE), E-ISSN: 2347-2693, 2018.