

SENTIMENT ANALYSIS FOR THE NEWS USING MACHINE LEARNING

¹V.Jyothi,²P.Kalyani,³M.Santhosh Reddy,⁴A.Shashi Kumar,⁵D.Vamshi Krishna

¹ Assistant professor, Department of Computer Science and Engineering
Anurag Group of Institutions, Hyderabad, Telangana, India

^{2,3,4,5} Department of Computer Science and Engineering
Anurag Group of Institutions, Hyderabad, Telangana, India

Abstract: - Sentiment analysis is a technique through which you can analyze a piece of text to determine the sentiment behind it. Numerous machine learning methodologies reveals better accuracy than lexicon-based strategies, they take extra time in training and applying the algorithm and consequently are not suitable for huge facts units. A computing sentiment of news data is a significant component of the social media big data. The proposed method uses Naive Bayes and Levenshtein algorithm to determine the emotion into different categories from given social media news data, which provide better performance for real time news data on social media.

Keywords:- Sentiment analysis, Levenshtein distance, Naive Bayes algorithm.

I. INTRODUCTION

Sentiment Analysis is the use of Machine Learning and Natural Language Processing to identify, extract and categorise the text documents. Sentiment analysis for micro blogs is difficult due to use of brief length of the message, casual phrases, word shortening, spelling variation and emoticons. It combines machine learning and natural language processing (NLP) to achieve this. Using basic Sentiment analysis, a program can understand whether the sentiment behind a piece of text is positive, negative, or neutral. It is a powerful technique that has important business applications. Our goal is to determine the emotion into different categories from given social media news data, which provide better performance.

II. RELATED WORKS

[1] A. Kaushik, S. Naithani, ³A Study on Sentiment Analysis: Methods and Tools' International journal of Science and Research, vol. 4, 2015, pp 287-291.

Due to the sheer volume of opinion rich web resources such as discussion forum, review sites, blogs and news corpora available in digital form, much of the current research is focusing on the area of sentiment analysis. People are intended to develop a system that can identify and classify opinion or sentiment as represented in an electronic text. An accurate method for predicting sentiments could enable us, to extract opinions from the internet and predict online customer's preferences, which could prove valuable for economic or marketing research. Till

now, there are few different problems predominating in this research community, namely, sentiment classification, feature based classification and handling negations. This paper presents a survey covering the techniques and methods in sentiment analysis and challenges appear in the field.

[2] Abinaya. R, Aishwaryaa. P, Baavana. S, 'Automatic Sentiment Analysis of User reviews', IEEE International conference on Technological Innovations in ICT for Agriculture and Rural Development, 2016, pp 158-162.

The Sentiment Analysis is sometimes a technique to look at the information that is the form of text and determine opinions content from the text. It is also termed as emotion or feeling mining. On-line communication channels like Twitter, Facebook, YouTube, and so forth are these days a lot of passion into human life. People share their thoughts or feelings thereon. During this review paper, we tend to match on opinion mining or feeling assessment which is an area of web data mining and Machine Learning. This paper shows aftereffect of examination by utilizing different ML and Lexicon investigation methodologies. Outcomes are analysed to play out an evaluation study and check the estimation of the present composition. In this manner, it will help the future investigators with understanding present beginnings in the configuration of possibility examination.

[3] Paul Sharma, Teng-Sheng Moh, 'Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter', IEEE International conference on Big Data, 2016, pp.1966-1970.

Sentiment analysis is considered to be a category of machine learning and natural language processing. It is used to extricate, recognize, or portray opinions from different content structures, including news, audits and articles and categorizes them as positive, neutral and negative. It is difficult to predict election results from tweets in different Indian languages. We used Twitter Archiver tool to get tweets in Hindi language. We performed data (text) mining on 42,235 tweets collected over a period of a month that referenced five national political parties in India, during the campaigning period for general state elections in 2016. We made use of both supervised and unsupervised approaches. We utilized Dictionary Based, Naive Bayes and SVM algorithm to build our classifier and classified the test data as positive, negative and neutral. We identified the sentiment of Twitter users towards each of the considered Indian political parties. The results of the analysis for Naive Bayes was the BJP (BhartiyaJanta Party), for SVM it was the BJP (BhartiyaJanta Party) and for the Dictionary Approach it was the Indian National Congress. SVM predicted a 78.4% chance that the BJP would win more elections in the general election due to the positive sentiment they received in tweets. As it turned out, BJP won 60 out of 126 constituencies in the 2016 general election, far more than any other political party as the next party (the Indian National Congress) only won 26 out of 126 constituencies.

Sentiment Analysis is a new subject in Research and is useful in many other fields. In Modern World, A huge amount of textual data is collected using surveys, comments, and reviews over the web. All of the collected data is used to improve products and services provided by both private organizations and governments around the world. This Paper includes sentiment analysis of movie reviews using feature-based opinion mining and supervised machine learning. In this paper, the main focus is to determine the polarity of reviews using nouns, verbs, and adjectives as opinion words. Reviews will be Classified into

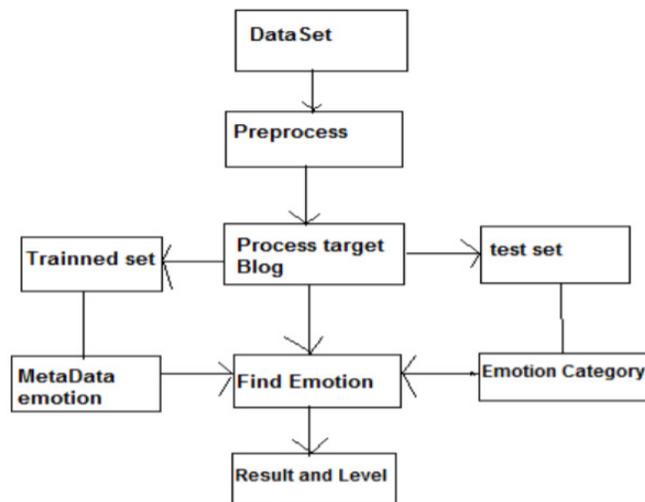
two different categories positive and negative. Reviews of Open Movie Database is used as source data set and Natural Language Processing Toolkit for Part of Speech Tagging. This paper also contains some facts about the classification of data on basis of polarity.

III. PROPOSED SYSTEM

In this paper, we proposed Our new technique idea is to do the sentiment computing for a news event by the assist of its semantics. The important part of the information occasion sentiment computing mission is the emotion computation, from the news event-based data on social media. Our major object is to discover the user emotion from news information. If we're method on new facts we just get-tough facts we need to system on statistics and discover which type of emotion own through these facts. We compute the sentiment for news data into different categories using Naïve Bayes algorithm and Levenstein distance, which provide better performance on real word dataset.

Merits of Proposed System

- Better performance.
- High in accuracy.
- SNMD approach that provide better accuracy compare with previous work.



IV WORKING OF PROPOSED SYSTEM

Levenstein Distance: -

- Levenshtein distance (LD) is a measure of the similarity between two strings, which we will refer to as the source string (s) and the target string (t).
- The distance is the number of deletions, insertions, or substitutions required to transform s into t. For example, x If s is "test" and t is "test", then $LD(s, t) = 0$, because no transformations are needed.
- The strings are already identical.
- If s is "test" and t is "tent", then $LD(s, t) = 1$, because one substitution (change "s" to "n") is sufficient to transform s into t.
- The greater the Levenshtein distance, the more different the strings are.

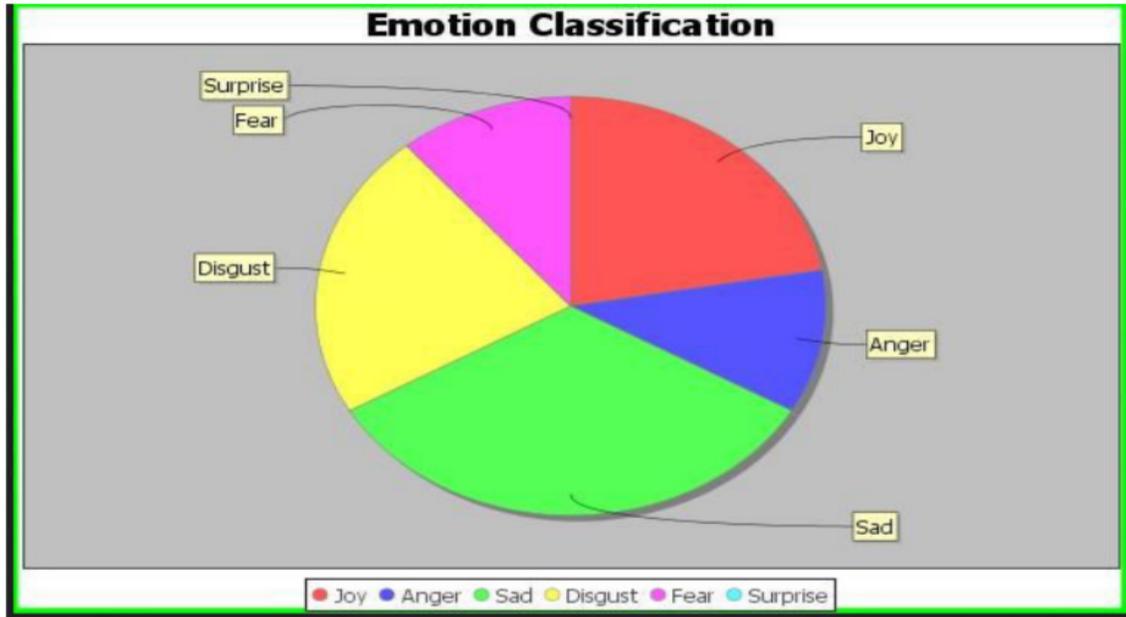
- Levenshtein distance is named after the Russian scientist Vladimir Levenshtein, who devised the algorithm in 1965.
- Levenshtein, the metric is also sometimes known as edit distance.

Naive Bayes Algorithm: -

- Naive Bayesian algorithm is a simple classification algorithm which uses probability of the events for its purpose.
- It is based on the Bayes Theorem which assumes that there is no interdependence amongst the variables.
- For example, if a fruit is banana and it has to be yellow/green in colour, in the shape of a banana and 1-2cm in radius. All of the properties stated above contribute individually towards that fruit being a banana and hence these features are referred to as “Naive”. As it considered the feature set to be Naive, the Naive Bayes algorithm can be trained using less training data and also mislabelled data.
- The Bayes Theorem is based on the following formula:
- $P(A/B) = P(A) \times P(B/A) P(B)$
- Here we are calculating posterior probability of the class A when predictor B is given to us i.e. $P(A/B)$. $P(A)$ is the prior probability of the class. $P(B/A)$ is the likelihood of predictor B given class A probability. $P(B)$ is the prior probability of the predictor B. Calculating these probabilities will help us calculate probabilities of the words in the text.

V. RESULTS

With all experiment and process we just find out the result of emotion of news event. Suppose any issue going on and social media are largely focus on that then we get dataset related this issue to find out positive and negative reaction of common people in form of emotions. With all test and manner we simply find out the result of emotion of news event. Suppose any difficulty taking place and social media are largely consciousness on that then we get dataset related this issue to find out positive and negative response of commonplace people in shape of emotions. We categories emotion in different level like anger, sad fear, joy, disgust and surprise.



VI. CONCLUSION

In this paper, we have developed an innovative method to do the sentiment computing for the news events based on the social media big data. Aim main target to process on News data and find out what reaction from this data in the form of emotion. We are able to find out this emotion level from current dataset. Our proposed method works on Levenshtein and Naïve Bays algorithm and effectively find a small emotion text. Emotions are very possessive and any emotion can use small text or large text like happy, hapy, hpy both are same level of emotion. With the help of new algorithm, we are easily distinguishing a emotion text from news events and data. Emotion are not easy to find from large data because user are uses number of text to define a emotion but it's not easy to define in category but our approach use a specific way to find this text from larger amount of data. So our conclusion is that Levenshtein algorithm provides a very easy way to text processing on data. Its work fast and provide maximum level of accuracy to processing large amount of data.

VII. REFERENCES

1. A. kaushik, S. Naithani, 'A Study on Sentiment Analysis: Methods and Tools' International journal of Science and Research, vol. 4, 2015, pp 287-291.
2. Abinaya. R, Aishwaryaa. P, Baavana. S, 'Automatic Sentiment Analysis of User reviews', IEEE International conference on Technological Innovations in ICT for Agriculture and Rural Development, 2016, pp 158-162.
3. Paul Sharma, Teng-Sheng Moh, 'Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter', IEEE International conference on Big Data, 2016, pp.1966-1970.
4. A.M. Reddy, V. V. Krishna, L. Sumalatha and S. K. Niranjan, "Facial recognition based on straight angle fuzzy texture unit matrix," 2017 International Conference on

- Big Data Analytics and Computational Intelligence (ICBDAC), Chirala, 2017, pp. 366-372, doi: 10.1109/ICBDACI.2017.8070865. (
5. A.M. Reddy, K. SubbaReddy and V. V. Krishna, "Classification of child and adulthood using GLCM based on diagonal LBP," 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Davangere, 2015, pp. 857-861, doi: 10.1109/ICATCCT.2015.7457003.
 6. Swarajya Lakshmi V Papineni, SnigdhaYarlagadda, HaritaAkkineni, A. Mallikarjuna Reddy. Big Data Analytics Applying the Fusion Approach of Multicriteria Decision Making with Deep Learning Algorithms *International Journal of Engineering Trends and Technology*, 69(1), 24-28, doi: 10.14445/22315381/IJETT-V69I1P204.
 7. Srinivasa Reddy, K., Suneela, B., Inthiyaz, S.,Kumar, G.N.S., Mallikarjuna Reddy, A." Texture filtration module under stabilization via random forest optimization methodology "International Journal of Advanced Trends in Computer Science and Engineering, Volume 8, No.3, May - June 2019
 8. W. Zhang, H. Xu, W. Wan, ³Weakness Finder: Find product weakness from Chinese reviews by using aspects-based sentiment analysis, *Expert Systems with Applications*, Elsevier, vol. 39, 2012, pp. 10283-10291.
 9. A. Khan, B. Baharudin, K. Khan; ³Sentiment Classification from Online Customer Reviews Using Lexical Contextual Sentence Structure[´] ICSECS 2011: 2nd International Conference on Software Engineering and Computer Systems, Springer, pp. 317-331, 2011.
 10. Ji Fang and Bi Chen, ³Incorporating Lexicon Knowledge into SVM Learning to Improve Sentiment Classification [´], In Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP), pages 94±100, 2011.
 11. A. Abbasi, H. Chen, and A. Salem, ³Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums, [´] In ACM Transactions on Information Systems, vol. 26 Issue 3, pp. 1-34, 2008.
 12. Mallikarjuna Reddy, V. Venkata Krishna, L. Sumalatha," Face Recognition based on Cross Diagonal Complete Motif Matrix", *International Journal of Image, Graphics and Signal Processing(IJIGSP)*, Vol.10, No.3, pp. 59-66, 2018.DOI: 10.5815/ijigsp.2018.03.07.
 13. Ayaluri MR, K. SR, Konda SR, Chidirala SR. 2021. Efficient steganalysis using convolutional auto encoder network to ensure original image quality. *PeerJ Computer Science* 7:e356 <https://doi.org/10.7717/peerj-cs.356>
 14. L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, ³Combining lexicon-based and learning based methods for twitter sentiment analysis, [´] HP Laboratories, Technical Report HPL-2011, vol. 89, 2011.
 15. D. Tang, B. Qin, and T. Liu, ³Learning semantic representations of users and products for document level sentiment classification, [´] in Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing, 2015.