

Analysis of Twitter Sentiments using Machine Learning Algorithms

Miss. Saraswati Bunge¹, Miss. Mijba Shaikh², Miss. Darshana Suryavanshi³, Ishrat Sayyed⁴
Prof. Monali Deshmukh⁵

Department of Information Technology^{1,2,3,4,5}

Matoshri Aasarabai Polytechnic, Eklahare, Nashik, Maharashtra, India

Abstract: *There exists plenty of data which prevails on the web available for internet users at the time of development of various web technologies. The internet has now grown into a place for online education, sharing ideas of each people and also for sharing one's opinion, the internet sites for instance google+, Facebook, Twitter are growing like a wind because they allow users in expressing their persuasion freely on diverse topics. It also acknowledges users to get into a conversation with other people or groups, and also share their messages across the world. The perusal on sentiment analysis in twitter data has driven a lot of people. It is helpful for classifying the tweets depending upon the polarity, which is positive negative and neutral. Analyzing the twitter data is tough because it has a large amount of data as the twitter contains data on diverse topics. There arises the need for simplified computing Machine learning algorithms. Therefore, for our Sentimental tweet prediction, supervised classification algorithms have been used. Using distinctive supervised machine learning algorithms such as K-Nearest Neighbor(KNN), Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Decision Tree, and Random Forest algorithms were employed for analyzing the sentiment of tweets. These machine learning algorithms has been compared, and the most effective method is chosen to predict the outcome.*

Keywords: Sentiments Analysis, Opinion Mining, Twitter Sentiments Analysis

I. INTRODUCTION

Twitter sentiments analysis make use of Natural Language Processing to evaluate a speaker's, writer's, or other person's mood and emotions through the piece of text. Through Sentiments Analysis, we can determine if a tweet of a user is positive negative or neutral. Social networking platforms such as Twitter, Facebook, Instagram, YouTube, etc. have been so popular for days now. They allow people to communicate, create networks, and share thoughts easily and promptly. Twitter has become an excellent medium for opinion creation and presentation. Twitter Sentiments Analysis can be used for real-time applications which can be a very helpful business. It can be used for people's sentiments on current political topics or trends. It can also be used for the review of movies using the trends on twitter.

As the audience on the media platforms grows continuously data from these sites can be used to analyses the sentiments of the people.

Manufacturers or developers of the products of the can review their product by analyzing the sentiments of the people. That is how people reacting to their products.

Marketing personal can see how people are reacting to their advertising campaign. They can analyze the sentiments related to this.

Political parties can see how their political campaign is running and how people reacting to it. They can analysis which issue to be raised to not.

Filmmakers can see how people are reacting to their newly released movie, by analyzing the sentiments of the people.

There are many reasons why we chose twitter data for sentiments analysis some of the reasons are given below:

Twitter is used varied from regular people to actors, politician, businessmen, and various religious and social leader to post their opinion. x The number of tweets on twitter daily is more than 500 million and that is enormous data for sentiments analysis.

Twitter users range from daily users to celebrities, from business executives to political figures. So twitter reflects views of all groups.

Twitter represents people of all age group and a high percentage of the business person is present on twitter. And people from many countries are present on twitter. x Twitter application has more than 50 million downloads and used by many on web browser. A popular use for this technology comes from its implementation in the social media field to explore how people feel about certain topics, specifically through the word-of-mouth of social media users in text posts or their tweets in the context of Twitter. In this paper, we use different machine learning methods to analyze the sentiments of the people.

Here we use machine learning methods like Naïve Bayes Classifier, Support Vector Machine method and Maximum Entropy method. Here we will compare these methods based on their accuracy and precision and see which method gives the best result. All the above methods are supervised learning methods. So, in all these cases we need to first train the data.

II. LITERATURE SURVEY

This section covers some research on Sentiment Analysis using Twitter social media regarding coronavirus disease. A study has been conducted in to identify reaction of people about coronavirus. Python is used with both Tweepy and Textblob libraries to analysis 500,000 generated tweets from 09-04-2020 to 15-04-2020. The result shows that 50% of tweets was neutral about coronavirus, 36% positive tweets and 14% negative tweets.[9]

Another study presented by Mourad and others to investigate the impact of social media using twitter during COVID-19 Pandemic. Tweepy library is used to analysis one million tweets that recorded over two months for 288K users. The result shows that 16.1% are misleading tweets, 83.9% are tweets with non-reliable medical information but 1% are reliable tweets from reliable sources such as journalist and doctors.[10]

A research presented by Dubey about COVID-19 sentiment analysis using twitter to analysis emotions of people during pandemic. The research included 12 countries from 11th March 2020 to 31st March 2020. The countries included are Australia, Belgium, China, France, Germany, India, Italy, Netherland, Spain, Switzerland, UK, and USA. The result shows that majority of 12 countries had more than 50% positive tweets compared with the negative and neural tweets. However, China, France, Switzerland, Netherland, and United States of America had more than 50% negative tweets.[11]

Samuel and others studied classified tweets using machine learning to find public sentiment. The study implemented in R with R packages and applied two methods for classifying the tweets. The first method is Naïve Bayes and the second method is Logistic Regression. The accuracy using Naïve Bayes with short tweets (less than 77 words) was 91.4% but for long tweets was 57.1%. Logistic Regression accuracy with short tweets 74.2% but 52% for long tweets.[12]

Drias and Drias analyzed tweets to identify people emotion and find association rules. There are more than 600,000 tweets from 23 February 2020 to 03 March 2020. The result show that negative tweets are more than positive tweets with fear as a highest word tweeted. Also, the higher confidence 0.93 for wash hand association rule and selloff economy simultaneously for 3 frequent patterns with 0.99 as the highest confidence.[13]

III. METHODOLOGY

In this thesis, both approaches have been combined, namely Lexicon-based and Machine learning for sentiment analysis on Twitter data. The algorithms were implemented for preprocessing of data set for filtering as well as reducing the noise from the data set. Therefore, the core linguistic data processing algorithm using Natural Language Processing (NLP) has been developed and implemented and discussed in Chapter 5, and assigned sentiment polarity to the tweets using lexicon-based approach. Finally, the data set is trained using machine learning algorithm: Naïve Bayes and SVM for measuring the accuracy of the training data set, and have compared results of both algorithms in Chapter 6. The most abstract view of derived approach that combines the lexicon-based and machine learning for sentiment analysis is shown in Figure 1.

Encryption process

- K-N Secret Sharing Algorithm

Copyright to IJAR SCT

www.ijarsct.co.in

DOI: 10.48175/IJAR SCT-15761



355

- Enveloping
- Digital watermarking
- Decryption process

System Architecture:

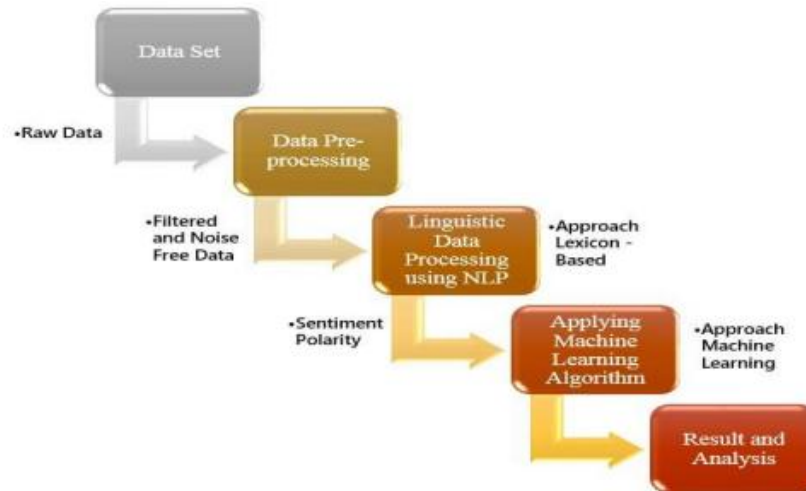


Figure 1: Overview on approach for Sentiment Analysis

Advantages:

- Sentiment analysis is an automated Machine Learning process that helps identify, classify, and analyze subjective information in any given text.
- It can glean the emotions, opinions, or judgments of people about a certain topic from the given text.

Disadvantages:

- In accurate Prediction.
- Difficult to Interpret.
- Expensive

IV. CONCLUSION

The different machine learning technique of data analysis of twitter are discussed like Naïve Bayes, SVM and Maximum Entropy Method. The analysis of twitter data is being done in various aspects to mine the sentiments. This study defines the concept of opinion in sentiment analysis of Twitter. Sentiment analysis deals with opinion classified into positive, negative and neutral. The study shows that the machine learning method such as Naïve Bayes has the highest accuracy and can be consider as the baseline learning methods as well as in some cases Maximum Entropy methods are very effective. More work in future is needed to improve the performance measures.

REFERENCES

[1] J. . I. M. Carpendale and C. Lewis , "Constructing an understanding of mind: The development of children's social understanding within social interaction," Published online by Cambridge University Press, vol. 27, no. 1, pp. 79-96, 2004.

- [2] J. Bjornestad, C. Moltu, M. Veseth and T. Tjora, "Rethinking Social Interaction: Empirical Model Development," J Med Internet Res, vol. 22, no. 4, 2020.
- [3] H. Wang, D. Can, A. Kazemzadeh, F. Bar and S. Narayanan, "A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle," in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, Republic of Korea, 2012.
- [4] S. Nadeem, "Coronavirus COVID-19: Available Free Literature Provided by Various Companies, Journals and Organizations around the World," journals and organizations around the world journal of ongoing chemical research, vol. 5, no. 1, pp. 7-13, 2020.
- [5] G. H. Guanche, "COVID-19. A challenge for healthcare professionals," Bimestral publication of the universidad de ciencias médicas de la habana , no. 1729, pp. 1-4, 2020.
- [6] Anon, "Severe Acute Respiratory Syndrome," Department of Health and Human Services, 2004.
- [7] D. S. Alehegn, "Document designed to create awereness for people covid 19," Jigdan college research and community service, Ethiopia, 2020.
- [8] World Health Organization, "Weekly Epidemiological Update," 2020.
- [9] K. H. Manguri, R. N. Ramadhan and P. R. Mohammed Amin, "Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks," Kurdistan Journal of Applied Research (KJAR), pp. 54-63, 2020.
- [10] A. Mourad, A. Srour, H. Harmanani, C. Jenainatiy and M. Arafah, "Critical Impact of Social Networks Infodemic on Defeating Coronavirus COVID-19 Pandemic: Twitter-Based Study and Research Directions," Computer Science, 2020.
- [11] A. D. Dubey, "Twitter Sentiment Analysis during COVID-19 Outbreak," 2020.
- [12] J. Samuel , G. G. .. N. Ali, M. M. Rahman , E. Esawi and Y. Samuel , "COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification," 2020.
- [13] Y. Drias and H. Drias, "Mining Twitter Data on COVID-19 for Sentiment analysis and frequent patterns Discovery," medRxiv, 2020.