IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, November 2022
Impact Factor: 6.252

A Survey on Automated Sentimental Analysis of Twitter Data using Supervised Algorithm

Mr. V. Chandra Sekhar Reddy[1] , K. Manvith Reddy[2] ,CH .Vachan Sai[3], K. Suraj[4], A. Abhinash[5]
Associate Professor, Department of Computer Science and Engineering[1]
IV B.Tech Students, Department of Computer Science and Engineering[2,3,4,5]
ACE Engineering College, Hyderabad, Telangana, India

Abstract: *Now a days as most of the people express their opinions and views on a particular topic, product or any other subject via social media, identifying the sentiment behind the statement has gained much significance. It is important to identify the statements that matter the most while someone is trying to know about a specific issue or product. In case of purchase of a product the buyer has the requirement of reading reviews that are sorted according to his requirements. While looking for cons of a product he wants to see all the negative reviews first and while looking for the pros of the product he wants start with positive reviews. Similar is the case with some national topic or a political scheme. As the data is enormous in size it cannot be analysed manually. We need a model that could perform sentimental analysis without human involvement. The Sentimental analysis involves few algorithms. In this model we are going to perform analysis using multiple algorithms such as Naivebayes, SVM, K-MEANS, KNN, Decision Treeand displaying performance comparisons between them (both accuracy and time taken). The GUI in this model in made much user friendly, the model is made flexible to be trained using any dataset as we can upload any dataset using the GUI developed and there is also a module where we can test the sentiment of custom input statement. The model can be used to analyse sentiment behind the statements and hence can be used in many real-time scenarios. One of application of this model is product reviews on shopping platforms so that the amount of negative and positive feedbacks can be seen by the new buyers. Another application may analyse sentiments that include political topics. As more and more people express their views through social media platforms, we can determine how the public's opinion is towards a particular scheme or a particular political Leader..*

Keywords: Algorithms, gui, analysis, product, dataset, model, platforms, topic, review, SVM Naïve Bayes, K-Means, K Nearest Neighbour, Decision Tree

I. INTRODUCTION

Most of the social media platforms such as Google, Facebook, twitter and e-Kart platforms like Flip kart, Myntra and Amazon have become popular as they allow users to express their opinions and views on the particular topic or product .Hence the collection of reviews is unstructured and with the polarities varying from extreme positive to extreme negative, analysing the sentiment behind the statement has gained importance. Sentimental analysis needs to be performed computationally as it is impossible for a human to classify millions of statements based on their polarities. This paper is aimed to provide a model that is trained on a dataset in 80-20 split ratio and performs analysis using many algorithms such as Naïve Bayes , SVM, K-Means, KNN, Decision tree and the most accurate one is picked . the accuracy varies with the size of training data set. The model also provides statistics of the comparison of performances of all the algorithms using a mat plot and the best algorithm which is suitable to perform analysis on the intended topic is picked. Accuracies of all the algorithms are calculated based on the 20% testing data. The algorithm that is more accurate may vary based on the topic that we choose to perform analysis on. Hence we have the option of uploading a data set on our own and then the more precise algorithm is picked to perform the sentimental analysis on a customised statement. Hence it is very important to train the model with the data set that is more relevant to the are that we are going to perform sentimental analysis in.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, November 2022
Impact Factor: 6.252

1.1 Challenges

Therearenumerousconcernsandobstaclesindeterminingthesentimentbehind a statement accurately.

A. Subjective Parts Identification

Thesubjectivepartinsentencemightbeanobjectinotherstatement.Itisquite complicated to identify how a particular word acts in a particular sentence.

Example:

Statement 1: There are several benefits from crude oil. Statement 2: His language was very crude.

In first statement, the word "crude" is the name of an oil and in statement 2 , crude is an opinion .

B. Domain Dependence

The words or tokens that provide a positive sentiment in a particular domain might have a negative sentimental impact in other domain.

Example:

Statement 1: The movie was unpredictable.

Statement 2: The shares of coke were unpredictable.

The word "unpredictable" in statement 1 describes about a film and has a very positive impact as the movies that are unpredictable are interesting. Whereas in statement2 the word "unpredictable" has a negative impact as unpredictable shares are always risky to invest in.

C. Sarcasm Detection

Many statements can be expressed in a sarcastically opposite way. Sarcasm cannot be detected during the analysis, so the model predicts the statement without considering the sarcasm.

Example:

Nice perfume, you should shower in it.

Here the statement contains all the words that contribute to predicting the statement as positive but the statement was intended in a sarcastic way.

D. Thwarted Expression

In some statements the sentiment of a part of the statement entirely determines the sentiment of the entire statement.

Eg: The film should be good. It sounds like a great film with all the elements required.

II. LITERATURE SURVEY

Application of Machine Learning Techniques to Sentimental Analysis [1] The paper was published by P. Dandannavar in 2016. An approach to sentimental analysis using naive bayes algorithm. Naïve bayes algorithm is a probabilistic classifier which mean it analyses the sentiment of the given statement based on the probability that is determined. A large dataset that is big enough to train the model for better accurate results is used to train the model and then the analysis is performed on the sentence where pre-processing such as tokenizing and stemming are done along with the generation on tf-idf matrix and count vectorizer

Optimisation of Sentimental Analysis using Machine Learning Techniques [2]: The paper have being published by Jaspreet Singh, Rajinder Singhand GurvinderSingh . It used the algorithms such as Naïve bayes and one R It would have been better if more algorithms were taken into consideration.

Sentiment Analysis of Product Reviews Using Machine Learning [3]The paper has divulged by Rajkumar S. Jag Dale, S Vishal . Shira sat ,N Sachin Deshmukh .It used the comparison of some of the algorithms like naïve bayes and support vector machine and found SVM was more accurate and precise. Many reviews are generated about a product or topic on the internet. It is difficult to categorize reviews which are generated on the internet. the analysis process consists includes natural language processing (NLP), computational linguistics, text analytics and classifying the polarity of the opinion. Naïve bayes and SVM are used in this analysis. This analysis also reveals that Support vector machine (SVM) let have high precision contrast to Naïve

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, November 2022
Impact Factor: 6.252

Sentimental Analysis of Movie Reviews Data using Machine Learning [4] The model targets analyzing sentiment of movie reviews. Opinion mining is done and the model uses naïve bayes and K Nearest Neighbor to perform the sentimental analysis and providethestatisticalreportsofthesentimentofaparticulartopicormovie.Manyother algorithmssuchasDecisiontreeandclusteringhaveemergedsothemodelcanbemade more precise if other algorithms are also used and the best one ispicked.

Investigate the Impact of Age and Gender on Sentimental Analysis using Machine Learning Classification Techniques [5] The paper had Advertised by Sudhanshu Kumar, Monika and Debi Prasad. The model uses the algorithms such as SVM Technology,convolutionalneuralnetworks,Extremityentropyandanmagnificent concept of abiding Temporary term memory to store the main features such era and gender identity of people and perform analysis based on the age. The paper mainly focuses on the impact of age and gender on sentimental analysis.

Automated for Sentiment Analysis of Political Tweets [6] The sentimental analys is of twitter data paper was published by Artik Das, Kushal Sai Gunturi, Aditya Chandrashekhar, Abhinandhan Padhi, Qian liu. This process avail oneself of all the NLP piece of work to clean and pre-processing of the dataset to make it ready for the categorization tasks from data scraping. The predictions are anticipated via word clouds and a colour map coded to disclose the sentiments of key nations around the world concerning the political events.

Sentiment Analysis of Twitter chatter Using Flask Environment [7]Flask environment based grading is carried out by Astha modi , Khelan Shah, shreyshah, Samir Patel , Manan Singh Flask environment comes up with various intrinsic functionalities to inspect the sentiments of text into three different class: positive, negative, and neutral. Also, it makes API calls to the Twitter Inaugurators account to fetch the Twitter data. After captivating need to analysing the data, then results get put on view on a webpage containing the percentage of positive, negative, and neutral tweets for a expression in a pie chart. It presents the language analysis for the same Expression.

Year	Technique/Methodology	Pros	Cons
2016	Analysis using Naive Bayes Classifier.	It classifies the data more relevantly	There are some limitations that should be addressed and to worked.
2017	Naïve Bayes and oneR	Results are predicted using social media parameters	No. of Limitations (very less Accuracy)
2018	Naïve Bayes and Support Vector Machine (SVM)	Detection can only be done in the plain text.	unprecedented levels of accuracy and fidelity
2019	Naïve Bayes and K-Nearest Neighbour Algorithm(K-NN)	Classifies the data accordingly on the dataset	Accuracy differs for different parameters.
2020	Introducing new attribute in sentimental Analysis. Support vector machine and maximum entropy	This predicts the data is positive or negative	Not Suitable for large data, data has more noise.
2021	Preprogrammed all the NLP tasks from data scraping to cleanse and pre-processing the dataset to make it ready for the classing tasks.	Usage of both Supervised and unsupervised algorithms	Vulnerable Detection, privacy issues
2022	Analysis using flask environment	Classifies the data using only flask environment	Works only for optimized data

III. CONCLUSION

The reason for the study is to build up an accurate supervised Algorithm that can classify the twitter data .Twitter is a free social networking site where the users broadcast short posts known as tweets. These tweets can contain text, videos, photos, links. Twitter is most commonly used by all the people nowadays Also Politicians, Celebrities; Higher Officials use the twitter to tweet the real time issues. This paper explains how the twitter data can be classified based on the different type of machine learning algorithm.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, November 2022
Impact Factor: 6.252

ACKNOWLEDGEMENT

We would like to thank our guide Mr. V. Chandra Shekhar Reddy and Mrs. Soppari Kavitha for dedicating their valuable time and guidance. Also, we are extremely grateful to Dr. M. V. VIJAYA SARADHI, Head of Computer Science and Engineering Department for his invaluable time, support, Ace Engineering College.

REFERENCES

[1]. P. Jain and P. Dandannavar, "Application of machine learning techniques to sentiment analysis of data" 2016 2nd global meetingfor Applied and Theoretical Computing and Communication Technology (iCATccT), 2016, pp. 628-632, doi: 10.1109/ICATCCT.2016.7912076.

[2]. Singh, J., Singh, G. & Singh, R. Optimizing sentiment analysis using machine learning Techniques. Hum. Cent. Comput. Inf. Sci. 7, 32 (2017). https://doi.org/10.1186/s13673-017-01163

[3]. Jagdale, Rajkumar & Shirsath, Vishal & Deshmukh, Sachin. (2019). Sentiment Analysis on Product Reviews Using Machine Learning Techniques: Proceeding of CISC 2017. 10.1007/978-981-13-0617-4_61.

[4]. Rahman and M. S. Hossen, "Sentiment Analysis on Movie Review Data Using Machine Learning Approach," 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), 2019, pp. 1-4, doi: 10.1109/ICBSLP47725.2019.201470.

[5]. Exploring Impact of Age and Gender on Sentiment Analysis Using Machine Learning. EISSN 2079-9292, Published by MDP Under the Electronics Organization.

[6]. Das, K. S. Gunturi, A. Chandrasekhar, A. Padhi and Q. Liu, "Automated Pipeline for Sentiment Analysis of Political twitter data," 2021 GlobalMeeting on Data Mining Workshops (ICDMW), 2021, pp. 128-135, doi: 10.1109/ICDMW53433.2021.00022.

[7]. Modi, A., Shah, K., Shah, S. et al. Sentiment Analysis of Twitter Tweets Using Flask Environment: A Superior Application of Data Analysis. Ann. Data. Sci. (2022). https://doi.org/10.1007/s40745-022-00445-1