IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

Automated Sentimental Analysis of Twitter Data

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: *In these days, many intellectuals in the society provide their opinions and perspectives on various topics, products and thoughts through social media. Because of this in the market, the requirement of analysing the sentiment of data became more important and increasing day by day. When people using the media need information about a specific issue in the society and products available, it is very crucial to recognize the statements which make sense. For example, when purchasing a product, the consumer may want to know the details of the product, reviews that are provided about it. If they want to look for the drawbacks of the product, they prefer to get through the negative reviews, whereas when wanting to know the advantages, they prefer to check with positive reviews. This pattern can also be applied to national problems, political areas also. Due to large amount of data, manually analysing the data becomes almost impossible. Therefore, we require a model which has the ability to perform sentiment analysis in an automated manner and without any human intervention. Sentiment analysis can be done with the help of different algorithms. In this model, we use the Natural Language Processing (NLP) for conduction of the analysis and present performance comparisons in terms of accuracy and time taken. The model's graphical user interface (GUI) has been designed to enhance user-friendliness. It offers flexibility by allowing for training with any type of dataset of different size and types of data with the help Graphical user interface (GUI) provided and it also includes a module for testing the sentiment of input statements which are custom. The model shows its essence by providing the sentiment and nature of behind statements, making it to be applicable in various real-time scenarios. For example, it can be employed in to assess the product reviews in shopping platforms, providing an overview for the new consumers to understand the positives and negatives about the products. Additionally, it can be utilized for analysing the sentiments related to political areas. Since, there is lot of growth in number of people expressing their views on political issues through social media beyond print media we can grab lot of opinions on political leaders from the common people regarding their manifestos, schemes implemented, behaviour etc.*

Keywords: Algorithms, Gui, analysis, product, dataset, model, platforms, topic, review, Natural language Processing

I. INTRODUCTION

Many Social media platforms such as Instagram, Facebook,Google, Twitter and e-commerce platforms like Myntra, Flipkart and Amazon gained more popularity since they offer a platform to users for expressing their opinions and views on specific and variant topics and products.

Because of this free spirit to people, the reviews found and collected becomes unstructured, with many polarities ranging from extremely positive to extremely negative. Therefore, the sentimental analysisbeyond these statements and analysing depth of the statements has become most important. It is technically, computationally infeasible for a human being to manually classify trillions of statements based on their polarities, necessitating the need for computational sentiment analysis. Natural Language Processing (NLP) is used to perform this analysis, with the most accurate algorithm being selected.

The accuracy of sentiment analysis varies depending on the size of the training dataset, which has to be suited for the intended topic to be analysed. The accuracies of almost every algorithm are calculated based on a 20 percent of testing dataset, which provides a reasonable measure of accuracy. However, accuracy may vary depending on the specific topic

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

being analysed. To address this, the model allows users to upload their own dataset, ensuring more precise sentiment analysis tailored to their specific needs.

Therefore, it is crucial to train the model using a dataset that is relevant to the area in which sentiment analysis will be performed.

The objectives of this model include training it on an 80-20 split ratio dataset and providing performance comparison statistics.

II. CHALLENGES

There are numerous concerns and obstacles in determining the sentiment behind a statement.

Subjective Parts Identification

Determining the subjective part of a sentence can be complex, as a word may serve as an object in one statement but convey a different meaning in another. Understanding how a specific word functions within a sentence in the passage or anywhere on web.

Examples:

Statement 1: He is good at English.

Statement 2: English people are smarter in some ways.

In the statement 1, the word "English" described as a Language, However, in the second statement, "English" is pointing to the group of people or a society and the meaning of English changes totally between two sentences.

Depending on Domains

The words or tokens may portray different sentiment according to the formation of sentences like the may be positive to think and some are with same words but provides different meaning which may negative also

For example:

Statement 1: Standing out became like a hobby to him in the college.

Statement 2: It is going to stand out as the best product.

In the First Statement, the word "Standing out" describes a like a person doing some work or taking a work as hobby, but which is negative in sense because it is a punishment given to him in the college. Whereas, in Statement 2, the word "Out Standing" carries a Positive sense provided to a product when compared to the other products. It is becoming one of the most demanded products in the market.

Detecting Sarcasm

Many statements can be formed in a sarcastic manner which contradict each other. However, detecting the sarcasm is difficult during the process of analysis, so the model we providedstarts predicting the statements without considering the sarcasm into account of prediction.

For Example:

Statement: "Nice Car, you should post in Social Media."

Here in this example, the statement has words which are like depicting as a positive intent of the sayer or a person. However, the provided statement is showing sarcasm on the other person is to make a negative intent on the person.

Thwarted Expression

In certain types of statements, a part of sentence portrays one meaning whereas whole sentence as a bunch of words gives the different essence to the sentence.

For Instance:

Statement: "He should be a good person. The great works done by him, and his team says it all."

Here in this scenario, the sentiment expressed in the statement hinges entirely on the sentiment conveyed in the mentioned part. The use of phrases such as "should be good," "sounds like a great person," and "great works done by him" suggests a positive sentiment towards the person.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

III. EXISTING SYSTEM

NAÏVE BAYES:

This way of approach uses the Naive Bayes algorithm to perform the analysis of sentiment. The Naive Bayes algorithm is a type of probabilistic classifier which tries to determine the sentiment of the provided statements based on some calculated probabilities. To train the model and achieve more accurate results, a sizable data set is used. The process involves some steps of preprocessing like tokenization and stemming, and also generating a good Tf-Idf matrix, count vectorizer also.

DECISION TREES:

Decision trees are a type of supervised learning method, which means they require training on labeled data. The underlying concept is like other text classification approaches. Given a collection of documents, typically represented as TF-IDF vectors, along with their corresponding labels, the algorithm determines the correlation between each word and a specific label. This process involves calculating the extent to which each word is associated with a particular label.

LOGISTIC REGRESSION

The concept behind this approach involves segregating the training set into positive and negative tweets. The words in the tweets are then counted, and a Python dictionary is created to store the frequencies of these words in both positive and negative tweets.

To analyze each tweet, a vector is generated consisting of a bias unit and the sums of the positive frequencies of the words present in positive tweets, as well as the sums of the negative frequencies of the words. Further details about this process will be discussed in the subsequent paragraph.

SUPPORT VECTOR MACHINES:

A comparison was conducted among various algorithms such as Naïve Bayes, Support Vector Machine (SVM), Maximum Entropy, Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM). It was found that some algorithms, like Maximum Entropy and CNN, demonstrated higher accuracy and precision.

In addition, an interesting concept of using LSTM was introduced to store and analyze the age and gender of individuals, enabling sentiment analysis based on these demographic factors.

Every day, a massive number of reviews and opinions are generated on the internet regarding products or various topics. Due to the sheer volume of reviews, it becomes infeasible to manually categorize and analyze such a large dataset.

K NEAREST NEIGHBORS:

The focus of the model is to analyze the sentiments showed in different areas like movie reviews, comments on social media. For the evolvement of analysis Opinion mining procedures also used, and also, the model uses Naïve Bayes, K Nearest Neighbor algorithms for performing of sentimental analysis. It generates statistical reports that reveal the sentiment associated with a specific topic or movie.

As the field of sentiment analysis continues to evolve, other algorithms like Decision trees and clustering have emerged. Incorporating these additional algorithms can enhance the model's precision, allowing for the selection of the best algorithm based on performance and accuracy.

IV. DRAWBACKS OF EXISTING SYSTEM"s

- Subjective part identification
- Domain dependence
- Order dependence
- Sarcasm detection
- Thwarted expression
- Entity recognition

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

V. PROPOSED SYSTEM

In this project we have used NATURAL LANGUAGE PROCESSING and WEB SCRAPING concept and performs analysis using data set provided and combines computational linguistics and then parses the raw data from web. A concept might be best in terms of accuracy in analyzing the statements that are related to politics and other algorithms might be best in terms of analyzing tweets related to business. So, the model we designed is more scalable and easier to be trained on a new data set. The data set is split into 80-20 and used for training and testing respectively and the algorithm that is best is picked to be used in customized sentiment testing process.

We will make use of Natural Language Tool Kit and Different Frameworks for exchanging the structured data. In advanced terminology we designed in such a way that which displays Polarity Score, Subjectivity Score, Word Count, Average no. of Sentences, Average no. of Words, Fog Index, Syllable for word, Personal Pronouns etc.

VI. ADVANTAGES OF PROPOSED SYSTEM

- It gives more accurate results.
- Easy to implement.

The process of finding the sentiment in a given or provided text is called opinion mining also can be popularly called as sentimental analysis. Here we use the combination of NLP (Natural Language Processing) techniques and with the help of web scraping we perform sentimental analysis on textual data gathered from the web.

Here's a general overview of the steps involved:

- Web Scraping: Web scraping is the process of extracting data from websites. You can use tools like Beautiful Soup or Scrapy in Python to scrape relevant textual data from websites. This could include reviews, comments, social media posts, or any other text data that contains sentimental information.
- Data Preprocessing: Once you have collected the text data, it's important to preprocess it. This step involves removing any unnecessary information, such as HTML tags or special characters, and converting the text to a consistent format. Additionally, you may want to perform tasks like tokenization (splitting text into individual words or tokens) and removing stop words (common words like "the," "is," etc.) to improve analysis accuracy.
- Sentiment Analysis: After preprocessing the data, you can apply various NLP techniques to determine sentiment.

VII. CONCLUSION

The research paper on sentiment analysis presents an in-depth analysis of the application of natural language processing techniques and web scraping for sentiment analysis. Through the exploration of various methods and approaches, the study has shed light on the potential of these techniques in extracting sentimental information from textual data gathered from the web.

The findings of the research indicate that sentiment analysis, when combined with NLP and web scraping, can provide valuable insights into the sentiment or emotion expressed in online content. The application of lexicon-based methods, machine learning-based methods, and deep learning-based methods has demonstrated promising results in predicting sentiment with varying degrees of accuracy.

The research also emphasizes the importance of data preprocessing in sentiment analysis. Properly cleaning and formatting the text data significantly improves the quality of sentiment analysis outcomes. Techniques such as tokenization, removing stop words, and handling special characters or HTML tags contribute to more accurate sentiment classification.

Furthermore, Adequate training data with sentiment labels ensures the model's ability to generalize well and accurately classify sentiment in new, unseen text. The evaluation of the sentiment analysis models using appropriate metrics has provided insights into the model's performance and effectiveness. Evaluation metrics such as accuracy, precision, recall, can be utilized to measure the model's success and identify areas for improvement.

While the research paper demonstrates the potential and effectiveness of sentiment analysis using NLP techniques and web scraping, it also acknowledges some limitations. Challenges such as domain-specific sentiment analysis, sarcasm or irony detection, and the influence of context on sentiment understanding require further investigation and improvement.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

In conclusion, the research paper contributes to the existing body of knowledge by showcasing the value of sentiment analysis in understanding the sentiment expressed in textual data from the web. It establishes the significance of NLP techniques and web scraping in extracting sentimental information and highlights avenues for future research and development in this field. The findings provide a foundation for the application of the sentiment analysis in many types of domains which includes social media messages and monitoring them, analysing the customer feedback .

ACKNOWLEDGEMENT

We are grateful and thankful to Mr. V. Chandra Shekhar Reddy, Associate Professor, Internal Guide, for his immense support and guidance. Also, we are thankful to our project coordinators Dr. Prem Kumar and Mrs. Soppari Kavitha, and we show our gratitude to Dr. M. V. Vijaya Saradhi, Head Of The Department (CSE), ACE Engineering College, for his constant support and providing valuable time for us.

REFERENCES

[1]Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1-2), 1-135.
[2]Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.
[3]Cambria, E., & Hussain, A. (2012). Sentic computing: Techniques, tools, and applications. Springer.
[4]Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the conference on empirical methods in natural language processing (EMNLP) (pp. 1631-1642).
[5]Kim, Y. (2014). Convolutional neural networks for sentence classification. In Proceedings of the conference on empirical methods in natural language processing (EMNLP) (pp. 1746-1751).
[6]Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the conference of the North American chapter of the Association for Computational Linguistics: Human language technologies (NAACL-HLT) (pp. 4171-4186).
[7]Wang, X., Li, L., & Deng, H. (2015). Sentiment analysis: Methods, applications, and challenges. IEEE Intelligent Systems, 30(4), 74-80.
[8]Hutto, C. J., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International AAAI Conference on Web and Social Media (ICWSM) (Vol. 8, No. 1, pp. 216-225).
[9]Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1(12), 2009.
[10]Kiritchenko, S., & Mohammad, S. M. (2018). Examining gender and race bias in two hundred sentiment analysis systems. Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics (*SEM 2018) (pp. 255-262).
[11]Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. In Proceedings of the Workshop on Languages in Social Media (pp. 30-38). Association for Computational Linguistics.