International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-12, Special Issue-1, March-2024
https://doi.org/10.55524/CSISTW.2024.12.1.15
Computing for Sustainable Innovation: Shaping Tomorrow's World (CSISTW-2024) on 16th-17th February, 2024
Organized by B.N. College of Engineering and Technology, Lucknow, India

Comment Analyzer for Sentiment Analysis in Social Media and E-Commerce Platforms

Ruchita Kumari[1], Anupriya[2], Soni Singh[3], and Harish Shukla[4]

[1,2,3]B.Tech Scholar, Department of IT, B.N. College of Engineering & Technology, Lucknow, UP, India
[4]Assistant Professor, Department of IT, B.N. College of Engineering & Technology, Lucknow, UP, India

Correspondence should be addressed to Ruchita Kumari; ruchitakumari1977@gmail.com

ABSTRACT- This research paper introduces a Comment Analyzer designed for sentiment analysis in the dynamic environments of social media and e-commerce platforms. The project focuses on harnessing natural language processing (NLP) and machine learning techniques to analyze and interpret sentiments expressed in user comments on social media and e-commerce platforms. The Comment Analyzer aims to enhance user experience, improve customer relations, and facilitate data-driven decision-making processes in the realms of social media and e-commerce.

KEYWORDS- Negative Comments, Neutral Comments, Positive Comments of Social Media and e-Commerce.

I. INTRODUCTION

In the vast and dynamic landscape of online communication, the sheer volume of user-generated content can be overwhelming. From social media platforms and forums to product reviews and blog comments, the opinions, sentiments, and discussions expressed by users are invaluable sources of information. However, manually sifting through this vast ocean of data is impractical, prompting the need for sophisticated tools known as comment analyzers.

A comment analyzer is a specialized software or algorithm designed to process, interpret, and extract insights from textual comments. Its primary purpose is to automatically analyze the sentiment, intent, and content of comments, providing a structured understanding of the vast array of user-generated text on the internet.[2]

Comment analyzers employ a combination of natural language processing (NLP) techniques, machine learning algorithms, and data analysis methods. These tools can categorize comments based on sentiment (positive, negative, neutral), identify key topics, recognize entities, and even understand the context in which comments are made[4,5].

II. OBJECTIVE

The primary objectives of the Comment Analyzer project include:

- Developing a robust sentiment analysis model specialized for social media and e-commerce comments.
- Implementing scalable and efficient algorithms capable of handling the vast volume of user-generated data.
- Enhancing user engagement and customer relations by understanding and responding to sentiments expressed in comments.
- Providing valuable insights for businesses to make informed decisions and improve their products or services based on user feedback.

III. METHODOLOGY

The purpose of the method is defined in this section. We used the annotated data set. We used python based machine learning libraries. First of all our system reads the data stored in the file. After reading, pre-processing phase is applied to clean and prepare the data for the use of machine learning algorithm. Directly text data cannot be given to machine learning algorithm it should be converted into a suitable type.[8] The text first goes through some language pre-processing, before it goes through NLPbased method to get data set. After that the data set analyzed with a sentiment classifier to calculate scores per comment and finally apply the quality deviation to get rating results. Goal of this section is to gather comment from social media. We perform some pre-processing on these unstructured remarks so as to make data set. Remove all expression that are not associated with suggested technique like date, link, special character .The comment collecting and pre-processing module first pules data from different sources like comment, review, feedback and process it.[12]

IV. ALGORITHM

Sentiment analysis, also known as mindfulness analysis, is a natural language processing (NLP) project that seeks to identify the emotions expressed in a text Emotions can be positive, negative, or neutral. Many frameworks and techniques have been used for sensitivity analysis, from rule-based methods to more sophisticated machine learning techniques Here are some of the most common concepts and approaches.

A. Rule-Based Methods:

• **Dictionary-based methods:** These methods use a dictionary of emotions or a dictionary of words related to positive and negative emotions. The meaning of a text depends on the presence and polarity of these words.[11]

• **N-gram Models:** Analysis of adjacent word sequences (n-grams) to capture context and emotion. For example, the one-gram model considers individual words, while the two-gram model looks at sequences of two words. [5]

B. Machine Learning Methods:

• **Naive Bayes Classifier:** A likelihood classifier using Bayes theory. Based on the availability of a set of words, it calculates the probability of reading a particular group of emotions.[6]

• **Support Vector Machines (SVM):** SVM classifiers separate data points into classes by searching for hyperplanes that maximize the difference between classes.

• **Logistic Regression:** A linear model that estimates the probability of a particular sentiment category based on a weighted sum of input characteristics. [8]

• **Random Forests and Decision Trees:** There are classes of learning methods that are predicted by combining multiple decision trees. It can be used to categorize emotions. [11]

• **Neural N:** Deep learning models, such as recurrent neural networks (RNNs) and long-term and short-term memory networks (LSTMs), can be trained on large data sets for sensory analysis tasks. [7,9]

C. Pre-Trained Language Models:

• **BERT (Bidirectional Encoder Representations from Transformers):** A transformer-based language model capable of capturing contextual information and pre-trained on large-scale tasks Optimization of BERT for sentiment analysis tasks has proven effective . [12]

• **GPT (Generative Pre-trained Transformer):** Models such as GPT-3 can be used for sensitivity analysis to understand sensitive content and generate through their speech generation capabilities.[2,6,8]

D. Decision Tree:

At different points in the text classification Decision tree classification can be used identified and analyzed. Its popularity depends on nature classification rules that make it interesting for NLP The researcher. Choice makes a decision Random data from a data set. It's useful—there are reasonable predictive rules Creates the fastest built from the training data tree Creates a short tree only as long as it is necessary to have sufficient property All data is partitioned and there are errors: The data can be overloaded or over-segmented, if it is small sample is tested, only one attribute is tested at a time To decide, the numbers don't solve it lost attributes and values To prevent overfishing. We optimize the hyperparameters of the Decision Trees such as max_features, min_samples_split,max_depth, and so on [6].

- **Random Forest:**A random forest classifier and compared its performance and other classifiers said random The forest algorithm makes it more efficient and smarter Consequently, segmentation is considered interesting .The one who divides things.[5]
- **K-th Nearest Neighbour:** KNN is a simple and efficient classifier. Called lazy learner because its training phase contains nothing but storing all the training examples as classifiers. KNN requires a lot of memory while storing the training values. The K-nearest neighbor algorithm essentially said that for a given value of K algorithm will find the K nearest neighbor of unseen data point and then it will assign the class to unseen data point by having the class which has the highest number of data points out of all classes of K neighbors.[11]
- **Natural Language Processing (NLP):** NLP, a branch of AI, is dedicated to the dynamic between machines and humans in their use of natural language. This encompasses comprehension, decoding, and production of human speech. Some commonly used methods include tokenization, part-of-speech tagging, named entity recognition, and sentiment analysis.[4,10]
- **Sentiment Analysis:** Sentiment analysis is a vital component of NLP that focuses on gauging the emotional tone of a given text. It is commonly used to decipher whether a comment conveys a positive, negative, or neutral sentiment. By utilizing sentiment analysis, comment analyzers can accurately comprehend the sentiment portrayed in a text.[1,8]
- **Feature Extraction:** Features play a crucial role in data used for training machine learning models. In comment analysis, they can range from analyzing word frequencies and sentiment scores to examining various linguistic aspects for better context and understanding of a comment.[4,5,8,12]
- **Deep Learning:** Deep learning has become a popular approach for analyzing comments. Remarkably, neural networks, in particular, have proven to excel in this task. Building on this foundation, models like Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), and BERT (Bidirectional Encoder Representations from Transformers) have proven to be remarkably adept at comprehending context and unraveling intricate linguistic nuances within text..[7,10]

V. APPLICATION

Feedback analysts find roles in a variety of industries and industries where insights and extraction from user content are invaluable. Here are some of the key activities:

A. Social Media Management:

Analyze comments on social media to gauge public sentiment, track trends, and understand how users perceive products, services, or events.

B. Customer Response Survey:

Soliciting feedback and surveys on e-Commerce websites or research forums to understand customer satisfaction, identify areas for improvement, and respond to customer concerns.

C. Brand Name:

Monitor and analyze feedback to control and improve brand reputation. This includes identifying potential PR issues, managing negative sentiment, and making the most of positive feedback.

D. Manufacturing and Marketing:

Analyzing feedback to gather insights into product development and marketing strategies. Understanding consumer mindsets helps companies tailor their products and campaigns to user expectations.

E. Content Management:

Using feedback analytics to automate and monitor user-generated content on websites and forums. This helps prevent the spread of inappropriate or harmful information.

F. Journalism and Media Analysis:

Analysis of comments in news reports and news forums to understand public opinion on current events, politics, and other topics of interest.

G. Market Analysis:

Using comment analytics to gain insights from online discussions, forums, and market research surveys. This can give businesses insight into market trends, competitor strategies and customer preferences.

H. Staff Feedback and Participation:

Use of feedback surveys for internal purposes, such as employee feedback survey

VI. FUTURE WORK

The future work of comment analyzers is likely to focus on advancing existing techniques and addressing emerging challenges. Here are some potential areas of future development in detail

- There is a need to improve the quality of claims analysis to better understand the context. This includes considering broader conversations, identifying comments on earlier comments, and identifying sarcasm or humor. Models that can better capture the context can be explored, such as long-range recursive roots or transformer-based architectures.
- Other features such as incorporating image and video analysis not only into stories but also into narratives. The integration of information from multiple sources provides a comprehensive understanding of emotion and intent.
- To move beyond simple labels of positive, negative, or neutral emotions to more fine-grained emotional analysis. This may involve recognizing specific emotions, understanding emotional intensity, or perceiving emotions in multiple dimensions within a single syllable.
- Creating annotation analyzers that can be easily customized for specific projects or projects. Customizable models allow companies to adapt sentiment analysis tools to their specific needs and terminology.
- Increasing the visibility and interpretation of claims analysis models. Users and entrepreneurs generally need to examine how models meet their ends. Reliable modeling is essential for widespread validation, particularly in applications where decisions based on data analysis can have significant impact.

VII. CHALLENGES AND LIMITATIONS

A. Brief Emotional Analysis:

Sentiment analysis, also known as mind mining, uses natural language processing and machine learning to analyze data and determine whether expressed sentiment is positive, negative or neutral. - Difficulties related to linguistic subtlety, context, and data quality pose limitations in sentiment analysis accuracy. [2,6]

B. Subtle understanding of language:

Subtleties of language, such as sarcasm, sarcasm, irony and dismissiveness, can confound accurate emotional assessment. - Improving the accuracy of sentiment analysis algorithms requires a deep understanding of linguistic nuances[6].

C. Contextual Challenges:

Sentiment analysis algorithms often struggle to capture contextual variation in sentiment expression. - It can oversimplify sentiment analysis, especially where sentiment relates to particular aspects of the text.[5]

D. Effects of data quality and bias:

Biased or poor quality training data can skew sentiment analysis results. - Over-reliance on user content leads to problems such as spam, false reviews, and conversions. [3,4]

E. Addressing constraints:

Advances in natural language processing and machine learning for promising solutions to overcome constraints. - Training for different data types, using aspect-based analysis, and refining algorithms can increase the accuracy of sentiment analysis.[3]

VIII. RELATED WORK

Many studies, in the field of sentiment analysis, have been carried out by different authors around the world. However, some notable papers on sentiment analysis have been published on Twitter comments and YouTube comments or social networks.(Thelwall et al., 2012) Sentiment analysis It has to do with the extraction of emotional information from text. The Sentiment analysis deals with marketing functions but sentiment analysis has played a role Increase in social network especially Twitter.(Prabowo and Thelwall,2009) Emotions Research plays an important role in the present study. The emotions are seen here feedback or criticism. Comment and can be useful for many purposes.[4] These are the emotions divided into positive and negative, or on an n-point scale. Recently (Jianqiang, 9). &Xiaolin, 2017) Twitters sentiment analysis empowers an organization Observe public sentiment about objects and events in the person's real time company[7]. In this paper, the authors conduct a study on the impact of text preprocessing method sentiment analysis in twitter posts. Research also reveals that as Twitter is accurate, Improved sensitivity analysis by extending the preprocessing method. But as a Grandfather Naive Bayes and random forest classifiers are superior when using the preprocessing method sensitive. (Bouazizi, Ohtsuki and Tomoaki2017) published a paper in which the authors explain that mind mining and sentiment analysis is the most popular topic today[6]. Most of them are Recent work has involved sensory analysis and mining texts. The authors in this study It was the first time SENTA had introduced an open sensitivity analysis tool. This is the solution .It is very accurate in binary ternary classification. It is a useful tool for sentiment analysis and sentiment terminology The study of emotions. This study was a sentiment analysis of Twitter posts. (Mike, et al., 2010) extracts emotional strength as informal English using SentiStrength software. (Vilares, Thelwall, & Alonso, 2015) analyzed the sentiment of 2,704,523 tweets He engaged with Spanish politicians and political parties from one month in 2014-15 through SentiStrength. state That's YouTube, a popular website with both commercial and non-commercial videos. Previous research (Bhuiyan, Hanif, Ara ,Jinat , Bardhan , Rajon& Islam , Md. Rashedul (2017) [12]. YouTube was reported to be the most popular site for watching videos, and videos. They are constantly putting it on. Users share videos and rate them based on the quality of the videos It's about rating. But popular videos don't always get a lot of likes or shares back Studies present the use of natural language processing for sentiment analysis or data retrieval is created by YouTube. This method helps in achieving relevance, popularity and high quality videos.(Thelwall et al., 2013) Sentiment analysis algorithms are now used to find Longitudinal use of emotional processes in online correspondence and machine-controlled support frameworks and user's interact well.[4]

IX. CONCLUSION

This project aims to summarize, detect, and classify different types of user generated text, including tweets, comments, and other forms of text. Machine learning techniques will be used to classify the text into various types. The process of analyzing digital text to determine if the emotional tone of the message is positive, negative, or neutral. The insights obtained from the analysis will be presented in an interactive dashboard for easy interpretation. Proper citation and referencing will be used throughout the project to ensure originality. [4,8]

In this study, we presented a case Analysis process for you-tube content. We have some of it Machine learning classifiers have been used a next to SVM, DT, LR, KNN and RF A variety of data processing and processing techniques Results of distribution. There are tests created in the dataset . The dataset is partitioned in appropriate training and test units . The accuracy of the distributions is as calculated by different analytical metrics F-score, and Accuracy score. The results show that SVM performs better than other classifiers. after SVM Naïve Bayes performs well. In the case of the macro average, SVM classification performance F score, accuracy and statistics are the best consider when a random forest is optimal in a given situation micro averages. The n-grams method together with Lemmatization algorithm for data dimension reduction . In this paper, we used six categories SVM with LR, DT, KNN, and RF.

CONFLICT OF INTERREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] Asghar, M. Z., Ahmad, S., Marwat, A., &Kundi, F. M. (2015). Sentiment Analysis onYouTube: A Brief SurveyMAGNT Research Report,3(1), 1250–1257.

[2] Bhuiyan,H., Ara, J., Bardhan, R., & Islam, M. R. (2017). Retrieving YouTube Video by Sentiment Analysis on User Comment. In prof. of the 2017 IEEE international Conference on Signal & Image Processing Applications (IEEE ICSIPA 2017), Malaysia, September 12-14,2017 (pp. 474–478). IEEE.

[3] Statistical Cybermetrics Research Group. (2018).Webometrics Analyst. Wolverhampton, UK: University of Wolverhampton

[4] Thelwall, M., Sud, P., & Vis, F. (2012). Commenting on YouTube Videos: From Guatemalan Rock to EI big bang.Journal of the American Society for Information Science & Technology, 63(3), 616–629.

[5] Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. Information Processing and Management, 52(1), 5–19. https://doi.org/10.1016/j.ipm.2015.01.005.

[6] Bouazizi, M., Ohtsuki, T., & Member, S. (2017). A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter.Rapid Review, 5, 20617–20639.

[7] Jianqiang, Z., &Xiaolin, G. U. I. (2017). Comparison Research on Text Preprocessing Methods on Twitter Sentiment Analysis. Rapid Review, 5, 2870–2879.

[8] Kharde, V. A., &Sonawane, S. S. (2016). Sentiment Analysis of Twitter Data : A Survey of Techniques.International Journal of Computer Applications,139(11), 5–15.

[9] Liu, B. (2012).Sentiment Analysis & opinion Mining.Morgan & Claypool Publishers.

[10] Pak, A., &Paroubek, P. (2011). Twitter for Sentiment Analysis : When Language Resources Are Not Available. In 2011 22nd International Workshop on Database and Expert systems Applications (pp. 111–115).IEEE. https://doi.org/10.1109/DEXA.2011.86.

[11] Pandey, A. C., Rajpoot, D. S., & Saraswat, M. (2017). Twitter sentiment analysis using hybrid cuckoo search method. Information Processing and Management, 53(4), 764–779. https://doi.org/10.1016/j.ipm.2017.02.004.

[12] Saif, H., He, Y., & Alani, H. (2012). Alleviating Data Sparsity for Twitter Sentiment Analysis. In MSM2012, 2nd Workshop on Making sense of Microposts,April 16 2012, (pp. 2–9). Lyon, France.