SENTIMENTAL ANALYSIS and SOCIAL MEDIA

Sowmya K
Professor, *dept. of Computer Science and Engineering*
Canara Engineering College
Mangalore
sowmya@canaraengineering.in

Sathvik
Student, *dept. of Computer Science and Engineering*
Canara Engineering College
Mangalore
sathvikmohan809@gmail.com

Shashank K N
Student, *dept. of Computer Science and Engineering*
Canara Engineering College
Mangalore
shashankkn2002@gmail.com

Shreejesh T
Student, *dept. of Computer Science and Engineering*
Canara Engineering College
Mangalore
tshreejesh@gmail.com

Sudeep M Shetty
Student, *dept. of Computer Science and Engineering*
Canara Engineering College
Mangalore
sudeepmshettysms@gmail.com

Abstract— **In an era characterized by the digital revolution and the widespread adoption of social media platforms, analysing user sentiments online has become essential for grasping public opinion, consumer preferences, and emerging trends. This project delves into sentiment analysis within social media, employing web scraping techniques to gather and evaluate data from diverse online platforms.**

The central aim of this project is to uncover valuable insights and trends by systematically examining user sentiments. With social media's pervasive influence, individuals and organizations now communicate, share information, and express opinions in unprecedented ways. This abundance of user-generated content presents a unique opportunity to access a vast repository of data that can offer meaningful insights. Sentiment analysis, also referred to as opinion mining, provides a means of understanding and categorizing the emotions, attitudes, and opinions expressed in online content.

This project showcases the potential of data-driven approaches in making informed decisions and gaining a competitive advantage in today's digital landscape.

Keywords—Sentiment analysis, Social-media, User sentiments, Data analysis, Web scrapping.

I. INTRODUCTION

In an age defined by the digital revolution and the widespread use of social media platforms, analysing user sentiments online has become crucial for understanding public opinion, consumer preferences, and emerging trends. This project focuses on sentiment analysis within social media, using web scraping techniques to collect and assess data from various online platforms.

The main goal of this project is to uncover valuable insights and trends by carefully examining user sentiments. With the widespread influence of social media, people and organizations now communicate, share information, and express opinions in new and extensive ways. The abundance of user-generated content presents a unique opportunity to access a vast pool of data that can provide meaningful insights. Sentiment analysis, also known as opinion mining, offers a method to understand and classify the emotions, attitudes, and opinions expressed in online content.

This project demonstrates the potential of data-driven approaches in making well-informed decisions and gaining a competitive edge in today's digital world. The methodology presented here can be adjusted to different domains, enabling various applications that utilize sentiment analysis effectively.

II. RELATED WORK

[1] The paper focuses on using text mining and sentiment analysis techniques to monitor product safety, particularly in the context of drug and cosmetic products, by analysing social media data such as Facebook comments and Twitter data. The authors aim to identify early signs of counterfeit products, allergies, adverse events, and product sentiment trends to aid decision-making for stakeholders like manufacturers, regulators, and enforcement agencies. They develop a machine learning classifier to predict sentiment orientation, providing a tool to track brand or product sentiment and take appropriate action in response to negative sentiment spikes. The study highlights the potential of social media data for combating counterfeit products and ensuring public safety.

[2] The research provides a comparative analysis of sentiment analysis in airline customer reviews employing Bidirectional Encoder Representations from Transformers (BERT) and its variants. It showcases the efficacy of transformer architecture-based NLP models, particularly emphasizing RoBERTa, in precisely categorizing sentiment within airline customer feedback. The results underscore the capacity of these models to augment operational efficiency and strategic planning within the airline sector. Additionally, the study recommends future research directions, including the investigation of more balanced datasets and the exploration of alternative large language models. Ultimately, this research establishes a benchmark for customer sentiment analysis in the airline industry and potentially other domains.t.

[3] The research delves into the exploration of the metaverse concept and its impact on the digital landscape.

Employing content and sentiment analysis techniques, including Natural Language Processing (NLP) methods and Artificial Intelligence (AI), the study aims to comprehensively understand the metaverse concept and its interconnectedness with related concepts. Through an analysis of metaverse articles sourced from The Guardian website, the research uncovers a spectrum of sentiment characteristics associated with the metaverse, ranging from positive to negative. Additionally, key themes and topics related to the metaverse, such as Facebook, games, platforms, marketing/advertising/PR, recreational business, and science & technology events, are identified.

[4] The research examines sentiment analysis techniques, with a specific focus on assessing the efficacy of different approaches in comprehending customer perceptions. It underscores the significance of feature engineering, particularly emphasizing the utilization of n-grams and semantic resources, to enhance the accuracy of sentiment classification. Furthermore, the study assesses the influence of various natural language processing (NLP) techniques, such as lemmatization and stopwords removal, on sentiment analysis models. Additionally, the research methodology involves the experimentation with machine learning algorithms such as SVM, NB, LR, and RF classifiers to categorize sentiment reviews and compare their accuracy. The authors, Dr. Mohammed Maree and Dr. Mujahed Elyat, possess extensive expertise in the fields of information technology and computer science, respectively.

[5] This research delves into the sentiment analysis of Twitter posts amidst the COVID-19 pandemic, with a particular emphasis on employing multinomial logistic regression. The primary objective is to comprehend public emotions during lockdown periods, recognizing their psychological ramifications. The research framework encompasses several stages, including pre-processing, polarity scoring, and feature extraction. Moreover, the study compares the proposed multinomial logistic regression method with decision tree, naive Bayes, and K-nearest neighbour models. The findings indicate that the multinomial logistic regression framework exhibits superior performance in terms of accuracy compared to the alternative models. Key terms: COVID-19, lockdown, sentiment analysis, machine learning algorithms.

[6] The research paper centers on the analysis of public sentiment regarding urban regeneration, utilizing sentiment knowledge enhanced pre-training and latent Dirichlet allocation techniques. It categorizes public sentiment into various topics associated with urban regeneration, such as development, investment, industry, and improvement. The study observes that public sentiment towards urban regeneration generally leaned more positively in 2022 compared to 2020, influenced by factors including project implementation and the impact of the SARS-CoV-2 epidemic. Furthermore, the research underscores the significance of analyzing public commentary over an extended period to comprehensively assess urban regeneration projects and effectively guide future initiatives.

[7] This article presents a novel method for evaluating airline service quality through the analysis of user-generated content and sentiment on Twitter. It investigates both positive and negative aspects of service quality across selected airlines, including Ryanair and Southwest, utilizing freely available analytical tools. The study's findings yield valuable insights into passenger feedback, highlighting its significance for service quality assessment and airline planning. The article suggests employing word clouds as a visual aid for sentiment analysis to facilitate decision-making in service quality enhancement. This cost-effective approach allows for comparative assessments between airlines, offering advantages over traditional methods such as onboard surveys. Moreover, the article discusses the intricacies of evaluating service quality in air transport, acknowledging the diverse range of services provided by airlines. It underscores the pivotal role of passenger satisfaction and various factors influencing airline choice. Additionally, the research explores sentiment analysis and user-generated content as a viable and cost-effective alternative for evaluating service quality, particularly advantageous for smaller airlines operating with limited financial resources.

[8] This research investigates various approaches for sentiment analysis on Twitter data, with a specific focus on classification techniques such as Naive Bayes and Support Vector Classification (SVC). The study identifies SVC as the most effective technique, achieving a notable accuracy of 86% through tenfold cross-validation. Moreover, the research underscores the escalating demand for social media sentiment analysis attributed to the proliferation of user-generated content. It also references additional studies that explore different classification techniques across diverse datasets, including movie reviews, airline Twitter data, and word embeddings.A significant challenge highlighted in sentiment analysis is the presence of imbalanced data, which can skew results. To address this issue, the study discusses various approaches employed to mitigate imbalance, enhancing the robustness and reliability of sentiment analysis models.

[9] The paper sought to analyze sentiments and emotions during the COVID-19 pandemic based on tweets posted between March 11th and March 31st, 2020. It demonstrated a consistency in people's mindsets globally during this period. The analysis utilized various sentiment analysis techniques, including BERT, renowned for its bi-directional reading capability and high accuracy. The study's conclusion highlighted that while the majority of people worldwide exhibited a positive outlook, there were instances of fear, sadness, and disgust expressed. The sentiment analysis conducted unveiled a prevalent negative outlook towards COVID-19 on a global scale, with individuals expressing concerns and negative sentiments regarding the pandemic.

[10] The paper introduces SAR-MCMD, an innovative approach that integrates multi-criteria decision-making with a hybrid deep learning method tailored for sentiment analysis within recommender systems. By combining expert evaluations, sentiment analysis techniques, and advanced deep learning models, the system aims to enhance recommendation accuracy and elevate user satisfaction levels. In light of the overwhelming volume of online information, the study underscores the critical role of sentiment analysis in refining recommendation quality and facilitating user navigation through a multitude of choices. The incorporation of sentiment analysis into the decision-making process with SAR-MCMD presents a promising solution for addressing the limitations of expert evaluations expressed in natural language, ultimately aiming to optimize user experience and decision-making outcomes. Additionally, the research extends beyond technical discussions to underscore the practical implications of enhancing user satisfaction through more tailored and precise recommendations. Leveraging sentiment analysis and deep learning techniques, the study endeavors to accurately predict user preferences and generate personalized recommendations based on multiple criteria, laying the groundwork for the development of advanced

decision-making models that adapt to the evolving needs of users in the digital realm.

III. PROPOSED SOLUTION

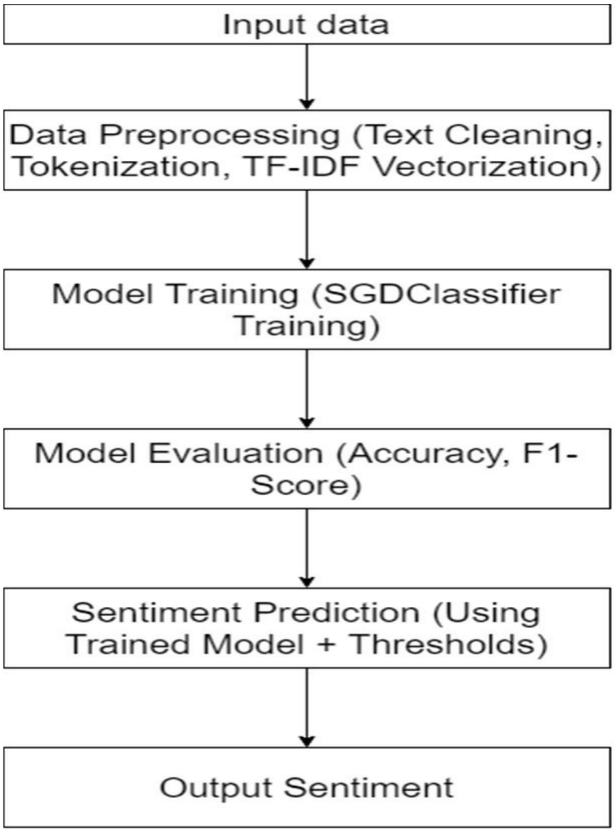

Figure 1: Architecture Diagram

Application Workflow

User Interface: This is the front-end part of the system where users interact with the sentiment analysis tool. It where users can input text to provide data for sentiment analysis.

Input Processor: This module is responsible for handling the input data provided by the user. It could be designed to accept various forms of input, such as raw text, files, or even live data streams, and prepare them for further processing.

Preprocessing: Before sentiment analysis can be effectively performed, the input data often needs to be cleaned and prepared. This preprocessing step might include tasks such as tokenization (breaking text into individual words or phrases), removing stop words (common words that don't contribute to sentiment, like "the" or "and"), stemming (reducing words to their base or root form), and handling of special characters or irrelevant formatting.

Sentiment Analysis Module: This is the core of the system where the actual sentiment analysis takes place. Here the SGD algorithm is used to determine the sentiment expressed in the input data. It may classify the sentiment as positive, negative, neutral and provide the accuracy.

Display Result: After the sentiment analysis is completed, the results need to be communicated back to the user. This module would take the output from the sentiment analysis module and present it in a user-friendly format, such as a visual representation (graphs, charts), a numerical score, or a text summary.

SGD Classifier

SGD is a popular optimization algorithm commonly used in machine learning for training linear classifiers.

SGD optimizes a differentiable objective function by updating the model parameters iteratively. At each iteration, SGD computes the gradient of the objective function with respect to the parameters using a single training example (hence "stochastic"). It then updates the parameters in the direction that minimizes the objective function, scaled by a learning rate. SGD is well-suited for large datasets, making it efficient for sentiment analysis tasks where the dataset size can be substantial. Due to its iterative nature and ability to process one sample at a time, SGD can handle datasets that may not fit into memory. In this implementation, the SGD Classifier utilizes the logarithmic loss function('log_loss'). Log loss is commonly used for binary and multiclass classification problems, measuring the performance of a classifier by penalizing incorrect classifications.

SGD Classifier, when combined with TF-IDF vectorization for text data, learns to classify text samples into sentiment categories based on the features extracted from the text. By minimizing the log loss function, the classifier optimizes its parameters to better predict the sentiment labels of the text data.

The SGD Classifier offers a versatile and efficient approach for sentiment analysis tasks, leveraging stochastic gradient descent optimization to learn from large-scale text data and make accurate predictions.

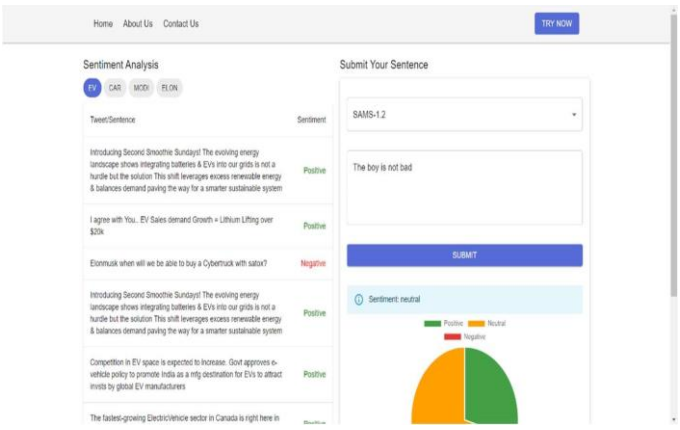

Figure 2: User Interface

IV. RESULTS AND DISCUSSION

The "Sentimental Analysis and Social media" project utilizes the text provided by the user and gives the output as either positive, negative or neutral. The accuracy along with F1-score is also displayed. The pie chart shows the degree of sentiment present in the text.

The Data preprocessing techniques like web scrapping, text cleaning, text lower-casing and TF-IDF along with SGD algorithm helps to revolutionize in obtaining the text from online platforms and use them to analyze the sentiments. This application through testing, has demonstrated high accuracy rate for the sentiment analysis. "Sentimental Analysis and Social Media" improves decision making factor which the user may be in need of and give them the clarity for their sentiments.

V. CONCLUSION

Sentiment analysis is widely utilized across social media platforms due to its versatility and broad applicability. It serves to understand public opinion on diverse topics and discern sentiments expressed by users. This tool serves as a pivotal component in understanding people's viewpoints and facilitating decision-making processes. Looking ahead, it is imperative to conduct additional research aimed at creating a universal sentiment analysis model that can effectively analyze various types of data. Moreover, expanding the exploration to include additional social networking platforms for collecting user feedback and diversifying the applications of sentiment analysis could yield invaluable insights, thus augmenting its efficacy across different contexts.

REFERENCES

[1] Isah, H., Trundle, P., & Neagu, D. (2014, September). Social media analysis for product safety using text mining and sentiment analysis. In 2014 14th UK workshop on computational intelligence (UKCI) (pp. 1-7). IEEE.

[2] Li, Z., Yang, C., & Huang, C. (2023). A Comparative Sentiment Analysis of Airline Customer Reviews Using Bidirectional Encoder Representations from Transformers (BERT) and Its Variants. Mathematics, 12(1), 53.

[3] Tunca, S., Sezen, B., & Wilk, V. (2023). An exploratory content and sentiment analysis of the guardian metaverse articles using leximancer and natural language processing. Journal of Big Data, 10(1), 82..

[4] Maree, M., Eleyat, M., Rabayah, S., & Belkhatir, M. (2023). A hybrid composite features based sentence level sentiment analyzer.

[5] Raheja, S., & Asthana, A. (2023). Sentiment analysis of tweets during the COVID-19 pandemic using multinomial logistic regression. International Journal of Software Innovation (IJSI), 11(1), 1-16.

[6] Chen, K., & Wei, G. (2023). Public sentiment analysis on urban regeneration: A massive data study based on sentiment knowledge enhanced pre-training and latent Dirichlet allocation. Plos one, 18(4), e0285175.

[7] Badanik, B., Remenysegova, R., & Kazda, A. (2023). Sentimental Approach to Airline Service Quality Evaluation. Aerospace, 10(10), 883.

[8] Tandon, V., & Mehra, R. (2023). An Integrated Approach For Analysing Sentiments On Social Media. Informatica, 47(2).

[9] Vijay, T., Chawla, A., Dhanka, B., & Karmakar, P. (2020, December). Sentiment analysis on covid-19 twitter data. In 2020 5th IEEE international conference on recent advances and innovations in engineering (ICRAIE) (pp. 1-7). IEEE.

[10] Angamuthu, S., & Trojovský, P. (2023). Integrating multi-criteria decision-making with hybrid deep learning for sentiment analysis in recommender systems. PeerJ Computer Science, 9, e1497.