SENTIMENT ANALYSIS USING HYBRID MACHINE LEARNING APPROACH

Soham Dalvi[1], Suyash Gharge[2], Prasad Gorhe[3], Saurabh Dev[4]

Prof. S.P. Jadhav[5]

Department Of Information Technology, Sinhgad College Of Engineering, Pune, India

---

Abstract - Sentiment Analysis (SA) is a rapidly evolving research field that faces challenges in keeping pace with its growth. SA explores how individuals express opinions and attitudes toward various subjects, events, and products through textual evaluations on platforms like blogs and forums. Amazon, a leading online retailer, allows users to openly rate products and submit reviews, vital for consumer decision-making. Analyzing these reviews for sentiment, positive or negative, aids decision-making from purchases to film reviews. Our project employs a hybrid machine learning approach to refine accuracy in sentiment analysis. Our project endeavors to categorize reviews according to their sentiments. Leveraging diverse machine learning models, our goal is to accurately classify the provided reviews into distinct groups representing positive, negative, or neutral sentiments. Moreover, our endeavor extends beyond mere classification; we also intend to conduct a comparative analysis of various machine learning models to assess their performance and accuracy in sentiment analysis.

Key words: Sentiment Analysis, Sentiment Classification, Opinion Mining, Machine Learning, Polarity Classification, Supervised Algorithms, Amazon Classification, Ensemble Learning, Hybrid Machine Learning.

1.INTRODUCTION

Sentiment Analysis (SA) has emerged as a crucial research area in recent years, driven by the increasing importance of understanding consumer sentiments in various domains. With the exponential growth of online platforms like Amazon, where users openly share their opinions and experiences through product reviews, the need for effective sentiment analysis techniques has become more pronounced. These reviews not only serve as valuable feedback for product manufacturers and retailers but also influence the purchasing decisions of potential buyers.

In this context, our research focuses on employing a hybrid machine learning (ML) approach for sentiment analysis of Amazon reviews. Traditional sentiment analysis techniques often rely solely on either supervised or unsupervised ML algorithms, which may have limitations in accurately capturing the nuances of human sentiment expressed in textual data. By combining the strengths of both supervised and unsupervised ML techniques, our hybrid approach aims to overcome these limitations and improve the accuracy of sentiment classification.

The primary objective of our research is to develop a robust sentiment analysis framework specifically tailored for Amazon reviews. We aim to accurately classify the sentiments expressed in these reviews as positive, negative, or neutral, thereby providing valuable insights for product manufacturers, retailers, and consumers alike. Additionally, we seek to contribute to the broader field of sentiment analysis by introducing a novel model that emphasizes the reduction of input data covariance during pre-processing, further enhancing the precision of sentiment classification.

Through this research paper, we aim to present our findings, methodologies, and insights gained from applying a hybrid ML approach to sentiment analysis of Amazon reviews. By doing so, we hope to contribute to the advancement of sentiment analysis techniques and facilitate informed decision-making in various domains influenced by consumer sentiments.

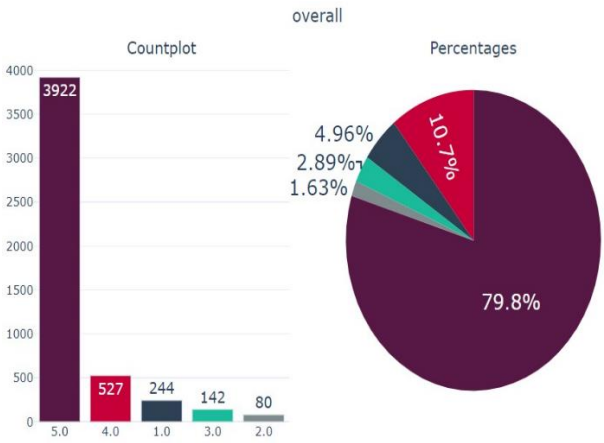
2. METHODOLOGIES

1. Data Preparation:

A dataset of Amazon customer reviews was meticulously curated, encompassing a broad spectrum of product categories. Each review within the dataset is labeled with its corresponding sentiment, delineating whether it embodies a positive, negative, or neutral

sentiment. This meticulous labeling facilitates the training and evaluation of sentiment analysis models.

In preparation for model implementation, a series of intricate text pre-processing steps were meticulously executed. Tokenization, the foundational process of breaking down the text into discrete tokens or words, was meticulously performed to provide a granular understanding of the textual data. Subsequently, the removal of stopwords, including common, non-informative words, was undertaken to streamline the analysis process. Furthermore, lemmatization, a critical step aimed at reducing words to their base or root form, was conducted to normalize the textual data, ensuring consistency across the dataset.



3.MODELS

1. Logistic Regression

Logistic Regression is a linear classification algorithm that is widely used for binary classification tasks. In sentiment analysis, it works by modeling the probability that a given input belongs to a particular sentiment class (positive or negative) based on the input features. Despite its simplicity, logistic regression often performs well in text classification tasks, particularly when the relationship between the input features and the target variable is linear or nearly linear. It learns a linear decision boundary by minimizing a logistic loss function and can handle high-dimensional feature spaces efficiently.

2. Naive Bayes

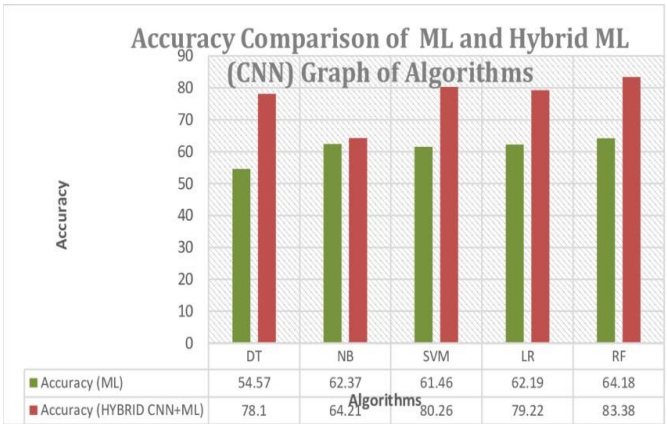
Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem with the "naive" assumption of feature independence. Despite its simplifying assumption, Naive Bayes often performs surprisinglywell in text classification tasks, including sentiment analysis. In sentiment analysis, it calculates the posterior probability of each sentiment class given the input features and selects the class with the highest probability as the predicted sentiment. Naive Bayes classifiers, particularly the multinomial variant, are well-suited for handling text data represented as word frequency vectors (e.g., TF-IDF), making them popular choices for sentiment analysis tasks.

3. Random Forest

Random Forest is an ensemble learning method based on decision trees. It constructs multiple decision trees during training and combines their predictions through a voting mechanism to make the final prediction. Random Forest is known for its robustness and ability to handle complex relationships in the data. In sentiment analysis, Random Forest builds decision trees based on features extracted from the text data and aggregates their predictions to classify the sentiment of the input text. Random Forest classifiers are less prone to overfitting compared to individual decision trees and can capture nonlinear relationships in the data effectively.

Each of these base models contributes unique strengths to the hybrid approach, and their combination through ensemble learning allows for improved performance and robustness in sentiment analysis tasks.

Table -1: Accuracy Report



4.BACKEND

Backend Implementation:

In developing the backend of our sentiment analysis web application, we focused on integrating the sentiment analysis model and handling user requests efficiently. The backend, powered by Python and Flask, manages the processing of user-inputted Amazon product reviews and returns sentiment analysis results to the frontend interface.

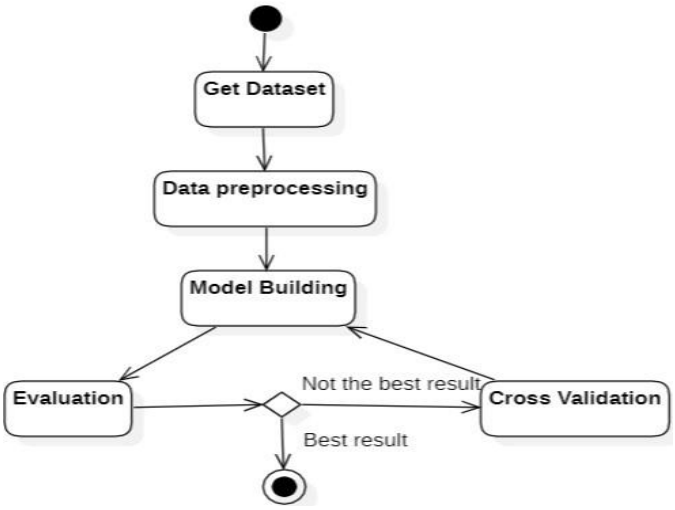
We incorporated our sentiment analysis model into the Flask backend to perform sentiment analysis on user-provided Amazon product reviews. The model, built using a hybrid machine learning approach as described in the methodologies section, processes the textual data and predicts the sentiment (positive, negative, or neutral) of each review.

We defined Flask routes to handle different user interactions and requests within the web application. These routes include functionalities such as receiving user-inputted reviews, passing them to the sentiment analysis model for processing, and returning the sentiment analysis results to the frontend interface.

The backend processes and validates user-provided Amazon product reviews before passing them to the sentiment analysis model. This ensures that the input data is clean and suitable for sentiment analysis, enhancing the accuracy of the results.

The backend interacts seamlessly with the frontend interface, exchanging data and responses through HTTP requests and responses. User actions on the frontend, such as submitting a review for analysis, trigger corresponding actions in the backend, which processes the requests and returns the results accordingly.

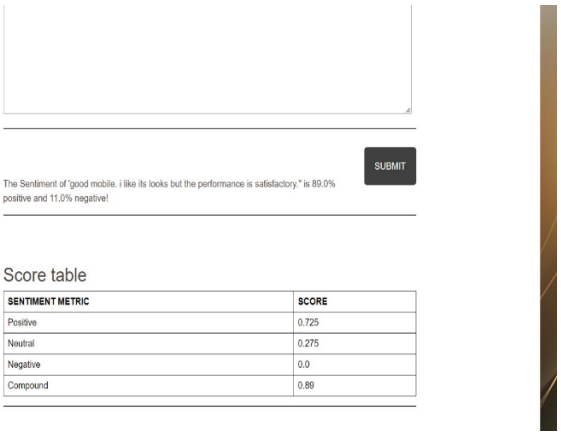
Once the backend implementation is complete, we deploy the Flask application on a local server or a cloud hosting platform to make it accessible to users. This enables users to access the sentiment analysis web application from their web browsers and analyze Amazon product reviews in real-time.



5.CLIENT

On the client-side of our sentiment analysis web application, users input sentences or reviews into a designated text field. Upon submitting the input sentence, the backend processes the text using the sentiment analysis model integrated into the Flask application. The model calculates the sentiment polarity scores, representing the percentage of positivity and negativity in the input sentence based on the hybrid machine learning approach described in the methodologies section. These scores are then dynamically displayed to the user in a tabular format within the frontend interface, providing immediate insights into the sentiment expressed in the input. This interactive and user-friendly approach enables users to quickly assess the sentiment of their input and make informed decisions based on the analysis results, enhancing their overall experience with the application.

Fig -1: Figure



6. CONCLUSION

In conclusion, our research highlights the efficacy of a hybrid machine learning approach for sentiment analysis of Amazon reviews, offering a nuanced understanding of consumer sentiments. By integrating supervised and unsupervised ML techniques, we've developed a robust framework capable of accurately classifying sentiments expressed in textual data, addressing the limitations of traditional methods. Through the implementation of a user-friendly web application using Flask, we've provided an interactive platform for users to analyze sentiment polarity scores of their input sentences, enhancing their decision-making process. Future research could focus on refining the model and expanding its application across different platforms and domains, contributing to the advancement of sentiment analysis techniques and providing valuable insights into consumer behavior.

7. FUTURE SCOPE

In the future, we aim to enhance our sentiment analysis framework by integrating advanced machine learning techniques like deep learning, refining its accuracy and granularity. Expanding the application to various e-commerce platforms beyond Amazon could broaden its utility, while real-time analysis capabilities could provide timely insights. Improving result interpretability through visualization and exploring applications in diverse domains like customer feedback analysis represent promising avenues for future development, ensuring our framework remains relevant and valuable across different contexts.

ACKNOWLEDGEMENT

We are deeply grateful to Prof. S. P. Jadhav, our guide, for her invaluable guidance and support throughout this endeavor. Her direction has been indispensable, and we cannot thank him enough for her advice, assistance, and collaboration. We would also like to extend our appreciation to every member of our department for their contributions. Additionally, we would like to acknowledge Dr. S. R. Ganorkar, Head of the Department of Information Technology, for his assistance during the project. Without their support, this work would not have been possible.

REFERENCES

[1] Meenu, Sunila Godara, "Sentiment Analysis using Decision Tree," IJEE, 2019.

[2] Anjali Dadhich, Blessy Thankachan, "Sentiment Analysis of Reviews Using Hybrid Rule-based Approach," MECS, 2021.

[3] Kian Long Tan, Chin Poo Lee, Kian Ming Lim, "Analysis With Ensemble Hybrid Deep Learning Model," IEEE, 2022.

[4] Amjad Iqbal, Rashid Amin, Javed Iqbal, Mudassar Hussain, "Sentiment Analysis of Consumer Reviews Using Deep Learning," MDPI, 2022.

[5] Alim Al Ayub Ahmed, Sugandha Agarwal, "Business Boosting Through Sentiment Analysis Using Artificial Intelligence Approach," Springer, 2021.