Enhancing Product Recommendation Systems through Text Classification of Customer Reviews

Abdullah Al Fahim
Department of CSE
Brac University
Dhaka, Bangladesh
abdullah.al.fahim1@g.bracu.ac.bd

Golam Rabbi
Department of CSE
Brac University
Dhaka, Bangladesh
golam.rabbi@g.bracu.ac.bd

Ananna Acharjee
Department of CSE
Brac University
Dhaka, Bangladesh
ananna.acharjee@g.bracu.ac.bd

Arannya Dhar Deb
Department of CSE
Brac University)
Dhaka, Bangladesh
arannya.dhar.deb@g.bracu.ac.bd

***Abstract*—This study focuses on the urgent requirement to improve product recommendation systems by utilizing sophisticated text categorization methods applied to user evaluations. Recognizing the importance of user-generated material, namely in the form of reviews, we provide a system that combines advanced natural language processing and machine learning methods to extract relevant insights. The goal is to enhance the accuracy and customization of the recommendation system by collecting the attitudes and preferences of users that are expressed in textual data. An assessment of the framework will be carried out using a wide range of datasets, guaranteeing its adaptability across different product areas. The expected results have the potential to revolutionize tailored product suggestions in e-commerce, providing platforms with a tool to better understand and meet user requirements. This research enhances customer pleasure and engagement, leading to a more informed and user-centric online buying experience in the growing field of e-commerce.**

I. Introduction

In the dynamic realm of online commerce, the vast array of items requires advanced recommendation algorithms to assist consumers in making well-informed decisions. In the midst of this digital marketplace, customer evaluations are seen as a useful collection of user experiences and opinions. They provide a wealth of information that may greatly improve the effectiveness of recommendation systems. This study aims to connect unorganized textual data, particularly customer reviews, with the enhancement of product suggestions using sophisticated text categorization methods. Current recommendation systems, however useful, sometimes encounter difficulties in precisely capturing the subtleties of human preferences and moods. This study aims to provide a complete framework that smoothly combines text categorization models with recommendation systems by utilizing natural language processing and machine learning. The purpose of this integration is to analyze the hidden information in customer reviews with the goal of improving the accuracy, relevance, and customization of product suggestions.Our objective in this research is to analyze the complexities involved and pinpoint the shortcomings of existing systems. Additionally, we aim to introduce a fresh strategy that enables e-commerce platforms to offer personalized and user-focused suggestions. This project aims to analyze the complex user attitudes included in textual data. Its findings contribute to the continuing discussion on improving recommendation systems, ultimately leading to a more user-friendly and enjoyable online shopping experience.

II. Related Works

Natural language processing (NLP) is being used in e-commerce to automate product evaluation and sentiment analysis. This study examines sentiment prediction using machine learning and deep learning models like BERT, Glove, Elmo, and Fast Text word embedding. The research highlights how natural language processing (NLP) transforms market analysis, staff well-being, and consumer satisfaction forecasts. It also shows how NLP can manage massive textual data and derive valuable insights. It also acknowledges the difficulties of analyzing consumer feedback and the changing NLP landscape across several sectors.

The author evaluates numerous e-commerce recommender systems algorithms and highlights the drawbacks of using simply user ratings, such as sparsity and unreliability. The suggested method generates ideas using user evaluations, sentiment analysis, and item-based collaborative filtering. It acknowledges limits such as a heavy reliance on sentiment analysis, limited comparability with other review-based systems, and a limited Amazon dataset assessment scope. The work also ignores cold-start, data sparsity, and scalability issues. It also does not analyze their approach's statistical significance or computing cost. To further comprehend the review-based recommendation system's efficacy and applicability, the book advises more extensive examinations and comparisons with comparable systems.

The article presents DeepCoNN, a dual neural network model that incorporates natural language processing (NLP) techniques to enhance Yelp recommendations. DeepCoNN achieves higher accuracy compared to baseline models and sentiment-analysis approaches, as measured on Yelp's dataset using RMSE, MAE, and FCP measures. The identified limits include the usage of fixed-length vector representation, the absence of temporal data, and the necessity for measurement of user enjoyment and business impact. These constraints indicate areas where improvements can be made.

The authors present a pioneering algorithm that significantly advances recommender systems. By advocating for the exploration of multi-category fuzzy preferences, the model transcends binary preferences, offering a nuanced understanding of user preferences and enhancing similarity calculations. Additionally, the introduction of weighted label attributes provides a sophisticated means to represent item characteristics, promising more accurate rating predictions by considering the varying importance of different attributes.The paper demonstrates a balanced perspective by highlighting both the strengths and limitations of the proposed algorithm, fostering transparency. The identified future research directions, including the expansion of fuzzy preferences and refined attribute weighting, showcase a forward-thinking approach. The paper emerges as a commendable contribution, combining innovation, transparency, and strategic foresight to advance the field of recommender systems.

III. DataSet

The Mi(Xiaomi) mobiles dataset, sourced from Kaggle.com, provides a comprehensive collection of mobile devices manufactured by MI. The dataset, which includes 549 rows and 8 columns, includes model names, price, ratings, image URLs, storage capacity, RAM configuration, os processor, network technologies, and battery specifications. The data is analyzed using textual data, which includes user ratings and reviews, price, storage capacity, RAM configuration, os processor, network technologies, and battery specifications. The dataset file in CSV format contains textual data pertaining to mobile devices, along with ratings for price and overall ratings. The text classification analysis classifies the dataset into three distinct classes: neutral, positive, and negative. The data is essential for conducting performance analysis and understanding the diverse range of mobile devices manufactured by MI.

A. Data Classfication

The data set consists of product information like the name,rate,price,image.Along with hardware specifications such as internal storage,RAM,memory card slot type.And other categories such as operating system.network connectivity and batteries information.

IV. Methodology

A. Data Preprocessing

To achieve efficient data preprocessing of our dataset containing 549 rows and 8 columns, prioritize addressing missing values, standardizing numerical features such as price and ratings, encoding categorical variables like model names and OS processors, and normalizing text data like picture URLs. Rectify anomalies, guarantee standardized formats for storage capacity and RAM arrangement, and verify coherence in network technologies and battery specifications. It is advisable to apply feature scaling to numerical features and utilize techniques such as one-hot encoding for categorical attributes. By preparing the dataset, it ensures that the analysis and modeling process is robust, leading to more accurate and reliable outcomes in machine learning.

B. Training Models

1) **Naive Bayes Classifier:** Naive Bayes is a classification technique that uses Bayes' theorem and assumes that the features describing an observation are independent of each other, given the class label. Text categorization is frequently employed for spam filtering and sentiment analysis.The formula of it:

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)} \tag{1}$$

In Naive Bayes, the "naive" assumption is that the features are conditionally independent given the class. This means that the presence or lack of one feature does not have any influence on the presence or absence of any other feature.

2) **BERT (Bidirectional Encoder Representations from Transformers):** BERT, which stands for Bidirectional Encoder Representations from Transformers, is a pre-trained natural language processing paradigm that Google introduced in 2018. The model employs the Transformer architecture and has been purposefully designed to understand bidirectional context in text. This capability allows it to discern the meaning of individual words in the context of the entire sentence. BERT is subjected to a comprehensive pre-training process using vast quantities of textual data. Following this, the model can be refined to execute various natural language processing tasks, including but not limited to text classification, named entity recognition, and question answering. The attention score α_{ij} in BERT is computed using the following formula:

$$\alpha_{ij} = \frac{e^{(W_Q q_i)^T W_K k_j}}{\sqrt{d_k}} \tag{2}$$

Here:

- q_i and k_j are the query and key vectors, respectively, for the i-th and j-th words in the input sequence.
- W_Q and W_K are the weight matrices for the query and key projections.
- d_k is the dimensionality of the key vectors.

The attention scores are subsequently used to calculate a weighted summation of the values V (value vectors) in order to obtain the output of the attention mechanism. By using the self-attention method, BERT is better able to understand the contextual meaning of words because it can take into account the entire context of a word when making predictions. The pre-training of BERT involves masked language modelling, where random phrases are hidden and the model is taught to predict these hidden words by using the contextual information from the surrounding words.

V. Result

A. Performance

*1) **BERT**:* Accuracy 56.36%: This is the proportion of successfully categorised cases out of the total occurrences in your dataset. The model accurately predicted the class labels for about 56.36% of the cases in your example.

Recall 56.36%: Recall, which is sometimes referred to as sensitivity or true positive rate, quantifies the model's capacity to accurately detect occurrences belonging to a specific class. Within this particular framework, it signifies that the model accurately classified around 56.36% of the occurrences across all categories.

The F1 51.08%: The F1 score is calculated as the reciprocal of the arithmetic mean of the reciprocals of precision and recall. It offers a trade-off.between precision (the capacity to accurately identify positive instances) and recall. The F1 score in your situation is around 51.08%, indicating a balance between precision and recall.

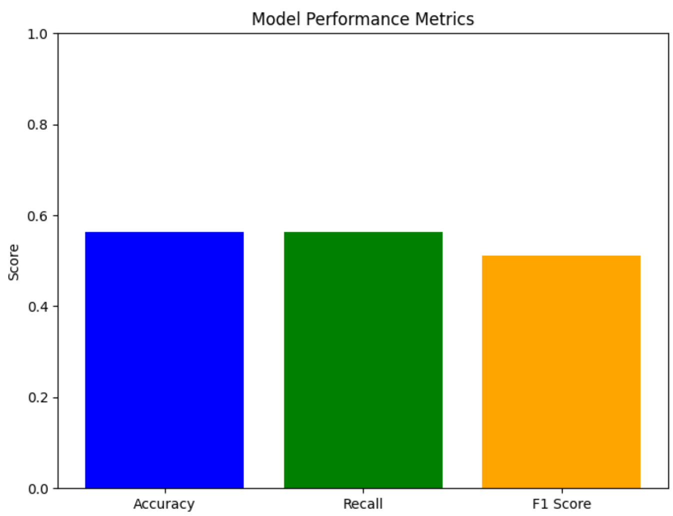

Fig. 1. BERT performance chart

*2) **Naive Bayes**:* Accuracy 52.73%: The accuracy is a measure that indicates the percentage of instances that are properly classified out of the total number of instances in the dataset. The Naive Bayes model achieved an accuracy of roughly 52.73%, implying that around 52.73% of the predictions were accurate.

The recall 52.73%: It represents the model's ability to correctly identify and capture all relevant instances of a specific class, sometimes referred to as sensitivity or true positive rate. Within this particular framework, the recall rate of 52.73% indicates that the model accurately recognised roughly 52.73% of cases that are part of the positive class.

F1 Score 49.66%: The F1 score is the balanced average of precision and recall, calculated using the harmonic mean. It offers an equitable assessment that takes into account both incorrect positive results and incorrect negative results. The F1 score of 49.66% indicates that the model has achieved a satisfactory trade-off between precision and recall, considering both false positives and false negatives.

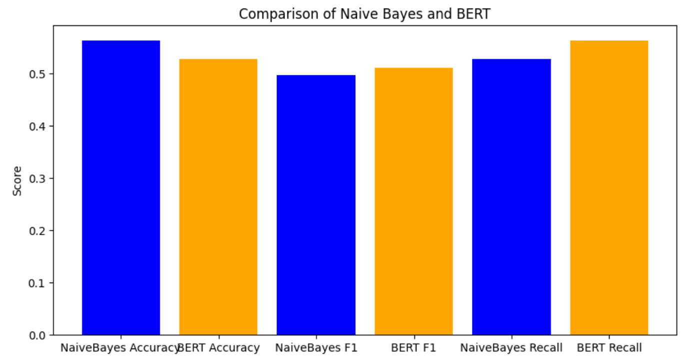

Fig. 2. BERT performance chart

B. Comparison

In the evaluation of our text classification models for enhancing product recommendation systems through customer reviews, both Naive Bayes and BERT exhibit notable performance, albeit with varying degrees of success. Naive Bayes achieves an accuracy, recall, and F1 score of 52.73%, demonstrating a moderate ability to classify customer sentiments. On the other hand, BERT, a more sophisticated transformer-based model, outperforms Naive Bayes with an accuracy of 56.36%, a recall of 56.36%, and an F1 score of 51.08%. While BERT shows a slight improvement across all metrics, the gains may not be considered substantial. Further exploration and model refinement are recommended, including hyperparameter tuning, investigating model complexity, addressing class imbalance, and conducting a thorough error analysis to uncover specific challenges in the data. The choice between Naive Bayes and BERT should also consider factors such as computational resources and deployment feasibility. Despite the observed differences, both models offer valuable insights into the enhancement of product recommendation systems through the text classification of customer reviews.

VI. Limitations

Although our study offers useful insights, it is not exempt from its limits. Firstly, the applicability of our findings may be limited by the characteristics of the particular dataset employed, therefore resulting in potential variations in the performance of the models across different scenarios. Moreover, the issue of interpretability continues to be a significant worry, particularly in intricate models such as BERT, which restrict the clarity of decision-making procedures. The problem of class imbalance may adversely affect the models' capacity to generalise effectively, and it is advisable to explore more advanced approaches to rectify this imbalance in future studies. Moreover, the performance measures, although helpful, may not encompass the complete range of user preferences and sentiments indicated in the reviews, therefore requiring a more nuanced evaluation methodology. Finally, the high computing requirements of transformer models might be problematic in contexts with limited resources, highlighting the importance of using efficient model topologies in real-world deployment scenarios.

VII. Future Work

There are several possibilities for further research in the field of enhancing product recommendation systems using text classification of customer ratings. Firstly, it is worth investigating the application of more advanced transformer architectures, such as GPT or XLNet, to ascertain if they offer further improvements in capturing complex language patterns. Moreover, including domain-specific pre-training or transfer learning methods tailored for the intricacies of mobile-related information is expected to enhance the models' understanding of context. Furthermore, the analysis of ensemble methodologies, which combine the benefits of several models, and the exploration of the impact of different data augmentation techniques should improve the dependability and accuracy of predictions. Staying abreast of the newest advancements and techniques in natural language processing is crucial for maintaining pace with the dynamic nature of the field and achieving consistent advancement.

VIII. Conclusion

Our study on the categorization of customer reviews for enhancing product recommendation systems offers valuable perspectives on the implementation of Naive Bayes and BERT models. While both models are capable of classifying customer feedback sentiments, BERT surpasses Naive Bayes in terms of accuracy, recall, and F1 score. The results suggest that transformer-based models such as BERT can enhance text categorization on mobile datasets by offering contextual comprehension. Nevertheless, the reported enhancements may lack statistical significance, hence requiring further meticulous investigation. Future research should focus on addressing hyperparameter tuning for both models, the complexity of the transformer model, and the issue of class imbalance. It is advisable to conduct a thorough error analysis to find any unique characteristics in the data that may impact targeted enhancements. When deciding between Naive Bayes and BERT, it is important to consider performance metrics, computational efficiency, and implementation feasibility. Our research enables us to analyse the advantages and disadvantages of these two methodologies within the framework of mobile datasets and customer evaluations as we seek to identify efficient text categorization models. This endeavour enhances the discourse on enhancing recommendation systems by extracting valuable information from written comments to provide a more effective service to consumers.

References

[1] Rahul Kumar Chaurasiya,Improving Performance of Product Recommendations Using User Reviews,09 May 2019

[2] Peter Kouvaris,Ekaterina Pirogova,Hari Sanadhya ,Albert Asuncion ,Arun Rajagopal ,Text Enhanced Recommendation System Model Based on Yelp Text Enhanced Recommendation System Model Based on Yelp Reviews,2018

[3] Dr. N. Srinivasan, Anandaraj Selvaraj,Mobile Based Data Retrieval using RDF and NLP in an Efficient Approach ,January 2018

[4] Zhan Su, Haochuan Yang, Jun Ai, Enhancing recommender systems with fuzzy preference, vector similarity, and user community for rating prediction, August 28, 2023.