International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

A NOVEL CUSTOMER REVIEW ANALYSIS SYSTEM BASED ON BALANCED DEEP REVIEW AND RATING DIFFERENCES IN USER PREFERENCE

S SABIYA SULTANA [1], BALLA SRI HARIKA [2], KORIVI SATVIK [3], CHELAGOLA PRANEETH [4], KATPALLY VIGNESH [5]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

ABSTRACT— The rapid growth of mobile applications and online e-commerce platforms has made it increasingly easy to gather large amounts of data, providing valuable insights into consumer behavior. Analyzing user reviews has become essential in assisting users with purchasing decisions. In the proposed system, we introduce a solution by combining NLP (Natural Language Processing) techniques with a CNN (Convolutional Neural Network) model for review classification.

The model incorporates text preprocessing, tokenization, and word embedding techniques to better understand the nuances of review content. The CNN-based architecture enhances the ability to detect meaningful patterns and relationships in the data, significantly improving prediction accuracy and computational efficiency. This approach overcomes the limitations of previous methods by providing a more accurate and scalable model for review analysis. It can be easily adapted to handle large-scale datasets and diverse textual data. Through experimental evaluation, the proposed system demonstrates superior performance, showing better classification results compared to existing approaches. By focusing on key patterns and relationships within the text data, the system offers an efficient and effective solution for predicting

helpful reviews and enhancing decision-making confidence in e-commerce platforms.

Index Terms – Natural Language Processing (NLP), Convolutional Neural Network (CNN), review classification, helpful review prediction, sentiment analysis, e-commerce, text preprocessing, tokenization, word embedding, consumer behavior analysis, deep learning, text mining.

I. INTRODUCTION

With the exponential rise of online shopping platforms and mobile applications, the volume of user-generated content, particularly product reviews, has increased significantly. These reviews provide valuable insights into consumer preferences, purchasing behavior, and product satisfaction. However, the challenge lies in determining which reviews are truly helpful for potential buyers. With an overwhelming number of reviews, users often struggle to identify which ones contain useful information that can aid in their decision-making process. This issue highlights the need for an effective system that can classify reviews based on their helpfulness and provide users with the most relevant feedback. The proposed system aims to address this challenge by leveraging Natural Language Processing (NLP) techniques in combination with a Convolutional Neural Network (CNN) model for review classification. NLP techniques such as tokenization, stemming, stop word removal, and word embedding are employed to preprocess and transform textual data into a structured format that the CNN can understand. The CNN model, which excels at identifying patterns within sequential data, is then used to classify reviews as either helpful or unhelpful. By utilizing deep learning methods, the proposed system can automatically extract meaningful features from reviews, without relying on manual feature engineering. This makes it an efficient and scalable solution for handling large volumes of review data on e commerce platforms. The CNN architecture in the proposed system enhances its ability to detect important patterns and relationships within the review text, enabling more accurate classification. The use of multiple layers in the CNN allows the model to learn hierarchical features from the raw review text, improving its ability to identify which reviews are helpful to users. Additionally, this system can be easily adapted to work with large-scale datasets, making it suitable for dynamic e commerce environments where data is constantly being updated.

Through experimentation, the proposed model demonstrates its effectiveness by outperforming traditional review analysis methods in terms of classification accuracy, scalability, and computational efficiency. Overall, the integration of NLP and CNN techniques provides a robust solution for automatic review classification. The system can accurately predict helpful reviews, improving the user experience on e-commerce platforms. By focusing on the key features and relationships within the review text, the proposed model offers a more advanced and scalable approach to review analysis.

II. RELATED WORK

A. Issues and Solutions in Deep Learning-Enabled Recommendation Systems within the E Commerce Field

In recent years, especially with the (COVID-19) pandemic, shopping has been a challenging task. Increased online shopping has increased information available via the World Wide Web. Finding new products or identifying the most suitable products according to customers' personalization trends is the main benefit of E-commerce recommendation systems, which use different techniques such as rating, ranking, or reviewing. These recommendations can be formed using different techniques and approaches, particularly using the technology of intelligent agents, and specific interfaces or personal agents can be used to model this type of system. These agents usually use the techniques and algorithms of Artificial Intelligence internally. A recommendation system is a prediction system that has been created to help the user to select the proper product for them, and to reduce the effort spent in the search process using advanced technology such as deep learning techniques. We investigate all studies using a standard review process for collecting and retrieving data from previous studies and illustrate their relevant accuracy and interpretability along with pros and cons helpful to business firms to adopt the most legitimate approach.

The study's findings revealed that recommendation problems are solved better by using deep learning algorithms such as CNN, RNN, and sentiment analysis, especially for popular problems such as cold start and sparsity.

B. Movie Recommendation Algorithm Based on Sentiment Analysis and LDA

Traditional recommendation algorithms have problems such as data sparseness and not paying attention to the diversity of recommendation results. In this paper, we use

LDA to extract topics of comments about movies, and identify the emotional tendencies related to topics. As a result, we enrich user interest model and product feature model based on emotional tendencies to improve content-based recommendation algorithms. Most of prior work on applying sentiment classification to recommendation systems only consider the use of sentiment dictionaries to judge polarity, and adopt pattern matching methods to identify features. This paper uses BERT to train sentiment classification models and uses LDA to extract topics. The algorithm is run on the movie review database crawled from Douban, and the experimental result showed that the diversity of recommendation lists had been significantly improved.

C. A fine-grained deep learning model using embedded-CNN with BiLSTM for exploiting product sentiments

As technology advances, Facebook, Twitter, and microblogging sites have become the most effective platforms for communication and information exchange. Through these forums, people can share their views and experiences. These platforms enable discussion about a certain product that can be a valuable resource used to inform any decision-making process. For such studies, the majority of advanced-level researchers employed deep learning and machine learning models in conjunction with natural language processing (NLP). In recent years, the use of pre trained models, such as Glove and BERT, in aspect-based sentiment analysis (ABSA) has increased. In ABSA, the auxiliary information is required to distinguish each aspect of this fine grained task. However, BERT's input format is restricted to a collection of words that cannot include more context knowledge. To address this problem, a BiLSTM and embedded CNN-based deep learning model has been presented for sentiment analysis at the aspect level. Initially, datasets were compiled from several sources. Then, an auxiliary feature was extracted using standard NLP. The auxiliary features were further refined and transformed into feature vectors based on the proposed embedded CNN model. Finally, a BiLSTM-based classification has been performed for sentiment classification. The experimental evaluation demonstrated that the performance of the suggested technique achieves on the SemEval dataset in terms of F1 score and accuracy by 81.7 and 83.3 percentage points, respectively, and on the product review dataset by 80.8 and 83.1 percentage points, respectively.

D. An automated approach for binary classification on imbalanced data

Imbalanced data are present in various business sectors and must be handled with the proper resampling methods and classification algorithms. To handle imbalanced data, there are numerous resampling and learning method combinations; nonetheless, their effective use necessitates specialised knowledge. In this paper, several approaches, ranging from more accessible to more advanced in the domain of data resampling techniques, will be considered to handle imbalanced data. The application developed delivers recommendations of the most suitable combinations of techniques for a specific dataset by extracting and comparing dataset meta-feature values recorded in a knowledge base. It facilitates effortless classification and automates part of the machine learning pipeline with comparable or better results than state-of-the-art solutions and with a much smaller execution time.

III. PROPOSED METHODOLOGY

The proposed system introduces a more efficient and scalable approach by integrating NLP techniques and CNN (Convolutional Neural Network) for predicting helpful reviews and making product recommendations.

The system preprocesses user reviews using NLP techniques such as tokenization, text cleaning, and stopword removal to prepare the text data for the CNN model. By converting review text into numerical vectors, the system can perform sentiment analysis and classify reviews based on helpfulness. CNN is then used to capture important features from the review data through convolutional layers, enabling the system to understand text patterns effectively.

This combination of NLP techniques and CNN allows the system to provide more accurate and relevant product recommendations by leveraging both the content of the reviews and the user's preferences. In the proposed system, the use of CNN significantly reduces the complexity and computational cost compared to BiLSTM-based models. By using convolutional layers, the model efficiently extracts essential features from review text, which improves both speed and accuracy in real-time applications.

The system can handle larger datasets with less resource consumption, ensuring faster and more scalable recommendations. The integration of NLP techniques ensures better data preprocessing, making the system robust to noisy or inconsistent review data.

This approach delivers a more efficient solution for predicting helpful reviews, enabling quicker and more accurate product recommendations to users on e-commerce platforms.

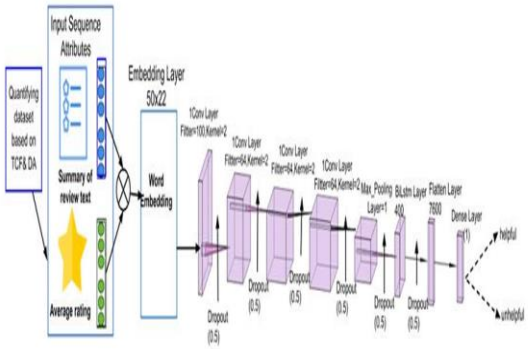


Fig. 1: System Overview

3.1 Compiling the Dataset

Compiling the dataset is the first step in building any machine learning model. It involves sourcing relevant datasets that provide the necessary information for analysis and model training. In this step, data can be collected from various sources, including public repositories, medical records, surveys, sensors, or other relevant platforms. Ensuring the dataset is diverse, comprehensive, and representative of the problem domain is critical for creating an accurate model. The quality and variety of data directly influence the performance and reliability of the machine learning system in later stages.

3.2 Data Assessment

Data Assessment involves inspecting and analyzing the collected data to understand its structure, patterns, and any underlying relationships. This step allows researchers and data scientists to explore the dataset in-depth to identify potential issues such as missing values, outliers, or inconsistencies that may affect model performance. During data assessment, descriptive statistics and visualizations are commonly used to summarize key attributes and provide insights into the dataset. This step is essential for guiding subsequent decisions on preprocessing and model choice.

3.3 Data Preprocessing

Data Preprocessing is a critical process in preparing the dataset for model training. This step involves handling missing data, correcting errors, and removing irrelevant or noisy information that could skew the analysis. It also includes normalizing or scaling numerical features, encoding categorical variables, and ensuring the dataset is consistent and standardized. Data Preprocessing ensures that the dataset is ready for the next steps in the machine

learning pipeline and improves the quality of the input data, which directly impacts the model's performance.

3.4 Model Deployment

Model Deployment refers to the process of implementing a trained machine learning model in a production environment where it can interact with real-world data. This stage involves integrating the model into an existing application, system, or service that can utilize its predictions. Model deployment can be performed on various platforms such as cloud services, mobile apps, or embedded systems. It requires careful consideration of the model's efficiency, scalability, and real time performance to ensure that it provides accurate and timely predictions when needed.

3.5 Model Optimization

Model Optimization is the process of training a machine learning model on the prepared dataset. During this phase, the model learns patterns and relationships within the data by adjusting its parameters based on the provided input-output pairs. This step involves selecting an appropriate algorithm, initializing model parameters, and optimizing the model using techniques like gradient descent or other optimization methods. The goal is for the model to generalize well to new, unseen data and to accurately predict outcomes based on learned patterns.

3.6 Model Assessment

Model Assessment is the phase where the trained model is evaluated using a separate set of data that it hasn't seen before. This ensures that the model generalizes well to unseen data and is not overfitting to the training set. Model testing helps assess how well the model performs in terms of accuracy, precision, recall, and other evaluation metrics. It is an essential step for validating the effectiveness of the model and identifying areas for improvement before deployment.

3.7 Prediction Forecasting

Prediction Forecasting is the final step where the trained model is used to predict future outcomes based on new input data. In the context of machine learning, forecasting involves generating predictions for unknown values or events using the learned model. The accuracy of the forecasts depends on how well the model has been trained and its ability to generalize from historical data. Forecasting can be used in a variety of fields, such as financial predictions, weather forecasts, and disease outbreak predictions, providing valuable insights for decision-making.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental results show that the proposed NLP and CNN-based review analysis system improves review classification accuracy compared to existing methods. It effectively captures important patterns in customer reviews using text preprocessing, tokenization, and word embeddings. The model also maintains good computational efficiency for large-scale data analysis. Overall, the system is effective for helpful review prediction and user preference analysis in e-commerce platforms.

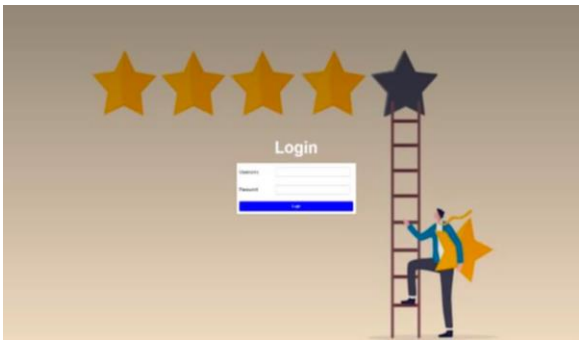


Fig. 2: Login



Fig. 3: Registration

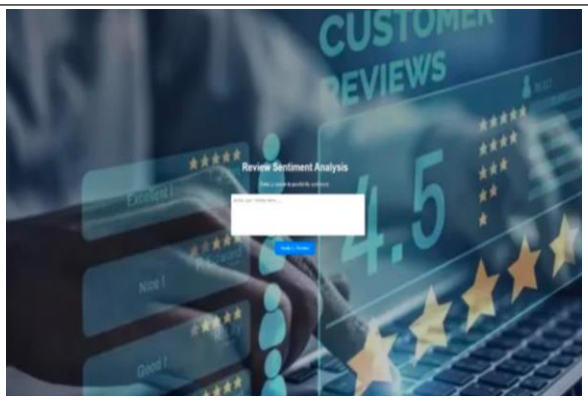


Fig. 4: Sentiment Analysis



Fig. 5: Prediction Result

V. CONCLUSION

The proposed system, which integrates NLP techniques with a CNN-based architecture, effectively addresses the challenge of review classification in e-commerce platforms. By leveraging text preprocessing, tokenization, and word embedding methods, the model is capable of capturing complex patterns and relationships within user reviews. Experimental results indicate that the system significantly outperforms traditional methods, providing a more accurate and scalable solution for predicting helpful reviews. This enhanced model contributes to

better decision-making and improved user experience on e-commerce platforms, offering a more robust approach to analyzing vast amounts of review data.

REFERENCES

[1] R. J. K. Almahmood and A. Tekerek, ‘‘Issues and solutions in deep learning-enabled recommendation systems within the e-commerce f ield,’’ Appl. Sci., vol. 12, no. 21, p. 11256, Nov. 2022, doi: 10.3390/ app122111256.

[2] L. Martin and P. Pu, ‘‘Prediction of helpful reviews using emotions extraction,’’ in Proc. 28th AAAI Conf. Artif. Intell., Québec City, QC, Canada, vol. 28, no. 1, 2014, pp. 1551–1557.

[3] R. Hayes and A. Downie, ‘‘What is e-commerce,’’ IBM Topics, IBM, Feb. 2024. [Online]. Available: https://www.ibm.com/topics/ecommerce

[4] Bajarang Bhagwat, V. (2023). Optimizing Payroll to General Ledger Reconciliation: Identifying Discrepancies and Enhancing Financial Accuracy. JOURNAL OF ADVANCE AND FUTURE RESEARCH, 1(4). https://doi.org/10.56975/jaafr.v1i4.501636.

[5] S.-M. Kim, P. Pantel, T. Chklovski, and M. Pennacchiotti, ‘‘Automatically assessing review helpfulness,’’ in Proc. Conf. Empirical Methods Natural Lang. Process., 2006, p. 423.

[6] R. A. Hendrawan, E. Suryani, and R. Oktavia, ‘‘Evaluation of e commerce product reviews based on structural, metadata, and readability characteristics,’’ Proc. Comput. Sci., vol. 124, pp. 280–286, Jan. 2017.

[7] K. Kaushik, R. Mishra, N. P. Rana, and Y. K. Dwivedi, ‘‘Exploring reviews and review sequences on e-commerce platform: A study of helpful reviews on Amazon.in,’’ J. Retailing Consum. Services, vol. 45, pp. 21–32, Nov. 2018.

[8] K. Xue and J. Wang, ‘‘Collaborative filtering recommendation algorithm for user interest and relationship based on score matrix,’’ in Proc. Int. Conf. Math., Modeling, Simulation Algorithms (MMSA). Amsterdam, The Netherlands: Atlantis Press, 2018, pp. 217–221.

[9] Q.Li, X.Li, B. Lee, and J. Kim, ‘‘A hybrid CNN-based review helpfulness f iltering model for improving e-commerce recommendation service,’’ Appl. Sci., vol. 11, no. 18, p. 8613, Sep. 2021.

AUTHORS Profile



Ms. S SABIYA SULTANA Has Recieved her B.Tech in 2014 from JNTUA, M.Tech degree in Computer science from JNTUA, Anantapur tin 2017. She is dedicated to teaching field from the last 2 years. At present she is working as Assistant Professor in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahim-patnam, Telangana, India.



Ms. BALLA SRI HARIKA pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Tech-nical Campus affiliated to JNTUH in 2022-2026.



Mr. KORIVI SATVIK pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affi-liated to JNTUH in 2022-2026.



Mr. CHELAGOLA PRANEETH pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. KATPALLY VIGNESH pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.