IJNRD.ORG ISSN : 2456-4184

INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG

An International Open Access, Peer-reviewed, Refereed Journal

EMPLOYING MACHINE LEARNING FOR BOOK REVIEW CLASSIFICATION

Suma Latha I
Associate Professor, Department of AI&ML Institute of Aeronautical Engineering College
Dundigal, Hyderabad,India

JAGADEESH THOTAPOLA
Department of Computer Science and Engineering
Institute of Aeronautical Engineering
Hyderabad, India

VAMSI CHATLA
Department of Computer Science and Engineering
Institute of Aeronautical Engineering
Hyderabad, India

INDHU PRIYA JARUPULA
Department of Computer Science and Engineering
Institute of Aeronautical Engineering
Hyderabad, India

Abstract : **Sentiment analysis, which is also known as opinion mining, is an essential task in Natural Language Processing (NLP). Over the past few years, there has been a growing interest in sentiment analysis, which involves classifying text to determine the intended meaning for the end user. One of the most valuable sources of consumer opinion is online book reviews, which are crucial in evaluating the quality of the book's content. To assist users in making informed decisions about which books to read, online review tools are now available. This paper explores various preprocessing techniques, such as removing HTML tags, URLs, punctuation, whitespace, special characters, and stemming, to eliminate noise. Additionally, machine learning algorithms are used for sentiment analysis to categorize book reviews and make recommendations based on user interests. By classifying user reviews as either positive or negative, clustering algorithms can be used to group people based on their interests, and a collaborative approach can be used to recommend books. The study aims to categorize book reviews using sentiment analysis and make book recommendations based on user interest variables. To achieve the most accurate results in the least amount of time, book feature sentiment must be extracted. This paper compares various levels of sentiment analysis and different approaches currently used to develop book recommendation systems.**

Keywords: Sentiment Analysis, Natural Language Processing, Machine Learning, Book Reviews, Recommendation Systems, and Clustering.

I. INTRODUCTION

In today's competitive market and ever-changing user demands, sentiment analysis (SA) plays a crucial role. Various entities such as government regulations, product status management, recommendation systems, and business intelligence applications rely on sentiment analysis. SA is a technique that analyzes people's feelings or perceptions of different things. The process of text classification mainly revolves around identifying whether a review expresses a positive or negative sentiment to the reader. With the growth of the internet, sentiment analysis has become increasingly important in understanding user interests. It provides valuable insights that enable users and businesses to make quick decisions. Several websites offer Application Programming Interfaces (APIs) that gather reviews for different books and utilize data analysis processes.

II . LITERATURE SURVEY

Sentiment analysis and text classification have both been the subject of substantial research, according to K. S. Sranan et al. One of the important difficulties in sentiment analysis is how to categorise sentiment polarity. At several levels, including entity or aspect level, phrase level, and document level, the categorization of sentiment polarity is possible. Finding out what individuals like and dislike in their opinions is a concern at the entity level. While evaluating the sentiment categorization of each individual sentence, the polarity of the entire document is taken into account at the sentence level. Machine learning approaches such as supervised, semi-supervised, and unsupervised have been used in studies by researchers all around the world. Noise-removing preprocessing methods include data cleansing, stemming, and html tag removal. To get rid of pointless features, utilise the chi-square feature selection method. The Support Vector Machine (SVM) is used to categorise evaluations into positive and negative groups. On the basis of accuracy for the movie review dataset, Tripathy et al. compared multiple classifiers. To decrease noisy data, preprocessing techniques were utilised, such as the removal of whitespace, numbers, stop words, and ambiguous information.

The features are extracted and represented using a TF-IDF and count vectorizer. Data are categorised as positive or negative using Naive Bayes (NB) and SVM, with SVM obtaining an accuracy of 94%. For categorising reviews as positive or negative, Turney et al. introduced an unsupervised learning technique. The system employs POS The pointwise mutual information and information retrieval (PMI-IR) algorithm are used to determine the semantic orientation of each phrase and categorize the review based on the average semantic orientation phrases.

Balasubramanian et al. assert that with the rise of e-commerce, customers have more options when shopping online, and more products are being launched in the market. To assist customers in making informed decisions, e-commerce websites such as Amazon, Flipkart, and Snap Deal encourage customers to rate products and leave reviews, which are useful for deciding whether to use a particular service or not. To help customers sort through the vast volume of information and choose products more effectively, big e-businesses create their recommender system, serving as a win-win strategy in e-commerce. Personalized user-recommendations are based on attributes clustering and score matrix, Using empirical metrics like as precision and recall, the effectiveness of this recommendation system was assessed for two different product categories: desktop computers and home theatre systems.using empirical measures such as precision and recall. The customizable product recommender system selects one further example for each feature, such that the total cost of all feature examples does not exceed the cost threshold. As the e-commerce sector is growing exponentially, the use of recommendation systems is increasing, enabling users to choose online content that is most relevant to them.

Edwin and colleagues conducted a study on improving service quality by comparing different sources of user feedback, such as expert evaluations, user-generated feedback, and internal sources. They used both quantitative surveys and qualitative interviews to gather information for their investigation. The study found that there is a strong relationship between customer feedback and satisfaction levels, and that both customer and expert reviews are essential in improving service quality. The sample for the analysis consisted of 140 hoteliers, which may limit the generalizability of their findings.

By examining the tone of customer evaluations, the algorithm assesses and suggests products depending on user preferences.
The greedy technique picks one item from each class and converts the multiple-choice knapsack issue into the 0-1 knapsack problem, as presented by Aravind and colleagues, to create a tailored product recommender system. The configurable product recommender system then selects a second example for each feature, making sure that the sum of all feature examples' costs does not go above the cost threshold. Recall and precision measures for two separate product categories, namely desktop computers and home theatre systems, were used to assess the performance of this recommendation system.

Silvana and colleagues introduced a prioritization mechanism for a recommendation system that utilizes consumer product reviews. In this study, the authors discuss the challenge of obtaining consumer reviews, which are often unstructured and in text format. To address this issue, ontologies are utilized to standardize the format of reviews, and product ratings are generated based on opinions expressed in these reviews. The quality of different product features is evaluated based on consumer comments, and products are recommended to consumers based on both opinion ratings and feature rankings. The proposed system is applied to a case study on digital camera reviews to demonstrate its effectiveness.

Tobias et al. proposed a customer satisfaction model for online product ratings that combines pre-purchase expectations with product performance ratings. This model is more effective in explaining the rating score compared to the conventional quality-centered explanations. The study tested this concept by examining internet ratings without considering the textual reviews and could not verify the review's authenticity before extracting it.

Smriti Rekha et al. introduced a recommendation system that prioritizes products based on several criteria such as star rating, helpfulness score, age of the review, and review polarity. The product score is calculated based on the sum of subjective and objective criteria. The study focused only on mobile phone reviews from the Flipkart website and used relatively small data sets. The system presents graphs to compare two products based on the criteria mentioned above.

To determine the sentiment of reviews, natural language processing is used. The sum of both objective and subjective criteria determines the product score. The research discussed in the article solely focuses on mobile phone reviews on the Flipkart website, making modest datasets more suitable.

Cheung et al. suggested a personalised marketing approach for product recommendations utilising machine learning algorithms like Support Vector Machines and Latent Class Model. Because it solves the problem of feature selection and outperforms more traditional content-based methods, the Support Vector Machine algorithm is used for content-based recommendations. The system's performance can be further enhanced by integrating consumer preference ratings for suggested products with collaborative recommendations produced using the latent class model.

III . METHODOLOGY

This project was developed using Python and required the installation of several packages, including pandas, numpy, matplot, seaborn, and sklearn. The necessary libraries for input and output processing were loaded at the beginning of the project. OpenCV and PyCharm libraries were also imported for this project. The inputs are received from the user terminal as strings and are classified by the algorithm based on these inputs.

At the start of this project, a large number of reviews were saved in an Excel sheet, and these were used to train the algorithm. The reviews were labeled as positive, negative, or neutral using specific words designated for each category, with positive reviews labeled as 1, and negative reviews labeled as 0. As new reviews were submitted by customers, they were added to the testing dataset, which was also stored in the Excel sheet. To classify the book reviews, packages such as NumPy, Seaborn, scikit-learn, and Support Vector Machine (SVM) were used, and algorithms like Naïve Bayes, SVM, and Random Forest were employed. When the "predict" option is selected, the output is displayed as positive, negative, or neutral using Naïve Bayes, SVM, and Random Forest algorithms. The Random Forest algorithm, in particular, utilizes multiple decision trees as base learning models, randomly performing row and feature sampling from the dataset to form sample datasets for each model. This technique is called Bootstrap, and it is a form of ensemble learning, which combines the predictions from multiple machine learning algorithms to improve accuracy. The final model made up of many models is known as an ensemble model.

The below diagram shows the architecture of the book review classification using the sentimental analysis.

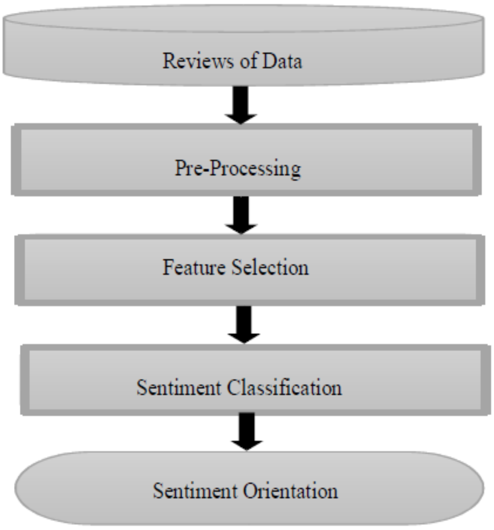

Fig.1 Block diagram for the proposed work.

B. IMPLEMENTATION

The following steps have been followed in a chronology in the proposed methodology.

INPUT DATA DESCRIPTION

a. *Load Data Set* : Load data set using pandas reacts() method. Here we will read the excel sheet data and store into a variable.

b. Split Data Set : Split the data set to two types. One is train data test and another one is testing data set. Here we will remove missing values from the dataset.

c. Train Data Set : Train data set will train our data set using fit method. 80% of data from dataset we use for training the algorithm.

d. Test Data Set : Test data set will test the data set using algorithm. 20% of data from dataset we use for testing the algorithm.

e. *Predict Data Set* : Predict() method will predict the results. In this step we will predict the ranking of the google play store app.

IV . RESULTS AND DISCUSSIONS

The testing has focused on the classification of the reviews submitted by the all the users of the e-commerce site. In this the reviews submitted by the user is stored in the excel sheet and it will evaluate the reviews and classify those based on their reviews count like positive reviews or negative reviews .And the training reviews also stored in the form of text in excel sheets .The results of the test has been given below.

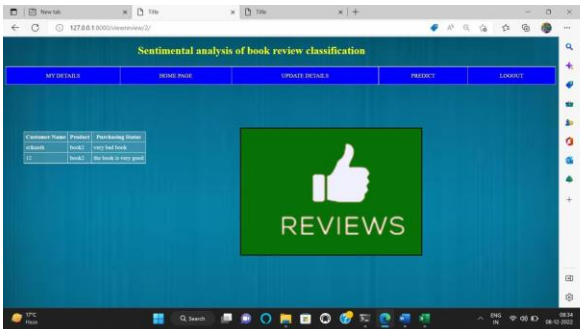

Fig 2 . Submitted reviews

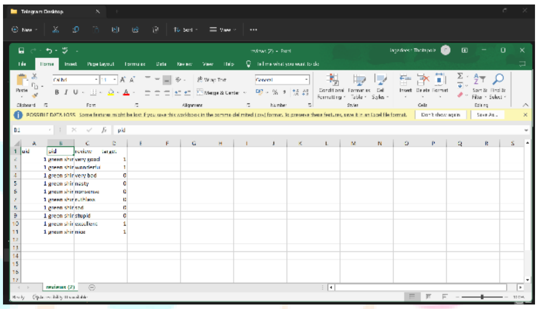

Fig 3.Reviews

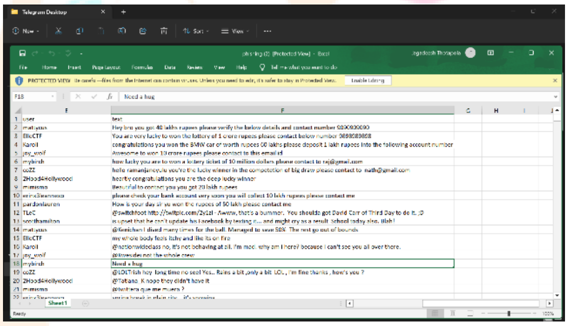

Fig .4 Reviews stored in Excel sheet

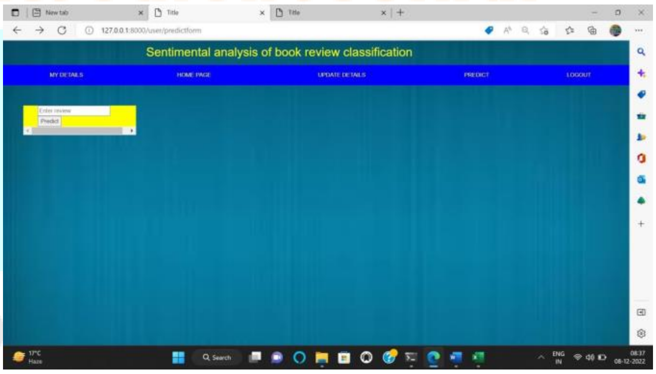

Fig.5 Review Classification

After clicking on the Predict option we will get the sentimental analysis of a reviews .

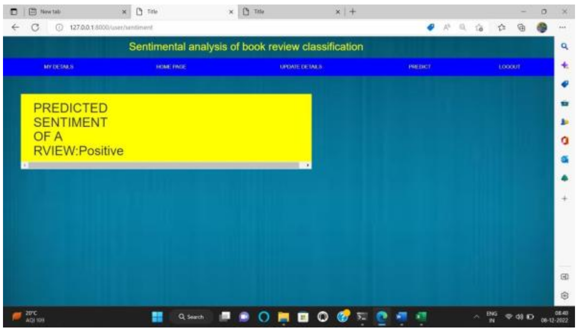

Fig 6. classification of reviews of book1 and the output is positive reviews so that the book is good to buy

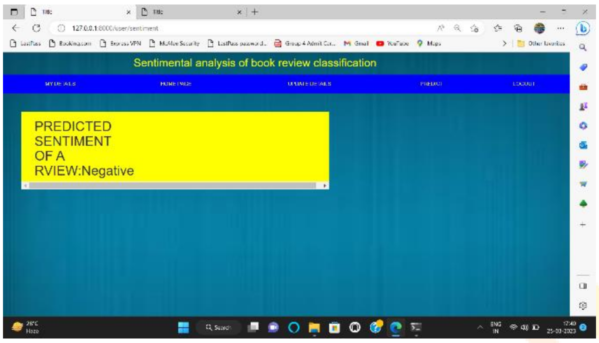

Fig 7. classification of reviews of book2 and the output is neagative reviews so that the book is good to buy

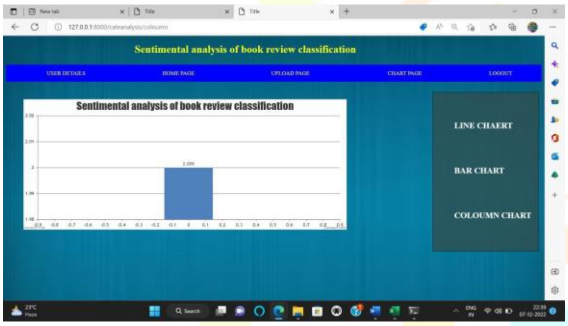

Fig 8 .Line Chart

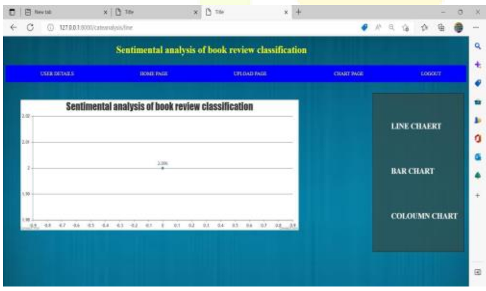

Fig 9. Bar Chart

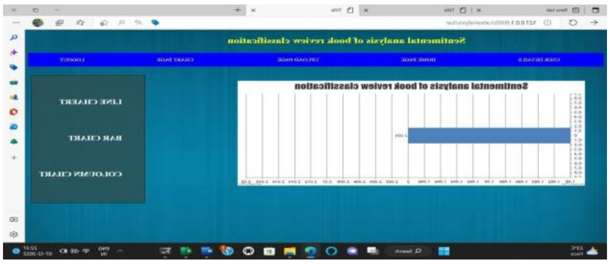

Fig .10.Column Chart

V . CONCLUSION

Online book reviews are a crucial means for customers to provide feedback on books. Currently, online review tools can be utilized by users to gain insights into the books before making a purchase decision. In order to suggest specific books based on user interest variables, this study offers sentiment analysis for categorizing book reviews. The user reviews can be classified as either positive or negative. To achieve maximum accuracy within a limited timeframe, it is necessary to extract the sentiment of book features. This paper examines different levels of sentiment analysis and compares various approaches used to make book recommendations.

FUTURE ENHANCEMENT

Companies can use sentiment analysis to determine the positive, negative, or neutral views that customers hold about their brand. Monitoring brand mentions and sentiment is crucial for customer engagement and interest. Brand recognition, trust, loyalty, and the effectiveness of advertising can be improved by branding. Brand mentions and discussions occur not only on social media, but also on blogs, news websites, forums, and product evaluations. While it is important to track the volume of brand mentions, it is equally important to analyze how customers are referring to the brand. Companies can gain valuable real-time and trend insights by examining the sentiment of customer comments.

VI . REFERENCES

[1] W. Medhat, A. Hassan and H. Korashy, "***Sentiment Analysis Algorithms and Applications: A Survey***", Ain Shams Engineering Journal, Elsevier, vol. 5, no. 4, pp. 1093-1113, Emerging Sources Citation Index (ESCI), 2014.
[2] K. S. Srujan, S. S. Nikhil, H. RaghavRao, K. Karthik, B. S. Harish and H. M. Keerthi Kumar, "***Classification of Book Reviews Based on Sentiment Analysis***", Information Systems Design and Intelligent Applications, Springer, vol. 672, pp. 401-411, 2018.
[3] C. Balasubramanian, J. Raja Sekar and M. Shenbaga Devi, "***A Personalized User Recommendation Based on Attributes Clustering and Score Matrix***", International Journal of Pure and Applied Mathematics (IJPAM), vol. 119, no.12, pp. 13751-13757, 2018.
[4] Santosh Kumar and Varsha, "***Survey on Personalized Web Recommender System***", I. J. Information Engineering and Electronic Business (IJIEEB), vol. 4, pp. 33-40, 2018.
[5] V. K. Kiran, R. E. Vinodhini, R. Archanaa and K. Vimalkumar, "***User Specific Product Recommendation and Rating System by Performing Sentiment Analysis on Product Reviews***", in International Conference on Advanced Computing and Communication Systems (ICACCS *-2017)*, pp. 978-983, Coimbatore, 2017.
[6] AyseCufoglu, "***User Profiling-A Short Review***", International Journal of Computer Applications (IJCA), vol. 108, no.3, December 2014.

[7] Aashutosh Bhatt, Ankit Patel, Harsh Chheda and KiranGawande***"Amazon Review Classification and Sentiment Analysis"***, International Journal of Computer Science and Information Technologies (IJCSIT), vol. 6, no. 6,pp.5107-5110,2015.

[8] AbinashTripathy, AnkitAgrawal and Santanu Kumar Rath, "***Classification of Sentiment Reviews using Machine Learning Techniques***", in 3rd International Conference on Recent Trends in Computing (ICRTC-2015), Elsevier, vol. 57, pp. 821-829, 2015.

[9] X. Fang and J. Zhan, "***Sentiment Analysis using Product Review Data***", Journal of Big Data, vol. 2, no. 5, pp. 529–551, Scopus, 2015.

[10] E. UkoOkon, B. O. Eke and P. O. Asagba, ***"An Improved Online Book Recommender System using Collaborative Filtering Algorithm"***, International Journal of Computer Applications(IJOCA), vol. 179 , no.46, pp. 0975 – 8887, June 2018.

[11] https://archive.ics.uci.edu/ml/datasets/Amazon+book+reviews

[12] Noia, T.D., Ostuni, V.C., Tomeo, P., & Sciascio, E.D., ***"SPrank: Semantic path-based ranking for top n recommendations using linked open data"***, ACM Transactions on Intelligent Systems and Technology, vol. 8, no. 1, pp. 9:1–9:34, 2016.

[13] McAuley, J., & Leskovec, J., ***"Hidden factors and hidden topics: understanding rating dimensions with review text"***, In Proceedings of the 7th ACM conference on recommender systems, RecSys '13, pp. 165– 172, New York: ACM, 2013.