E-COMMERCE PRODUCT REVIEW SYSTEM USING SENTIMENT ANALYSIS

Nandhana S, Nivedha Shankar, Rajeswari S
Department of Computer Science and Engineering
Sri Venkateswara College of Engineering
Sriperumbudur, India
2020cs0020@svce.ac.in

Rajasekaran T, Vinothiyalakshmi P
Department of Computer Science and Engineering
Sri Venkateswara College of Engineering
Sriperumbudur, India
rajasekaran@svce.ac.in

Abstract--**In the dynamic landscape of global commerce, the role of digital reviews, facilitated by e-commerce giants like Amazon and Flipkart, has become instrumental in shaping consumer behaviors and influencing market trends. These platforms serve as virtual hubs where consumers from diverse corners of the world converge to share their experiential insights, creating a vast repository of information. Within this milieu, sentiment analysis emerges as a crucial facet, involving the intricate task of deciphering and categorizing opinions embedded in reviews, posts, and comments. At the heart of this evolving terrain lies a notable project focusing on supervised binary sentiment classification through the implementation of advanced text mining algorithms. The project undertakes a meticulous collection and analysis of diverse customer review data, seeking not only to discern sentiments but also to distinguish between positive and negative expressions. The utilization of varied data preprocessing techniques enriches the analytical process, contributing to heightened accuracy and reliability in sentiment classification. Beyond serving as a tool for consumer decision-making, this project establishes a symbiotic relationship with manufacturers. By providing a nuanced understanding of consumer sentiments, it equips manufacturers with invaluable insights, creating a feedback loop that fosters continuous product improvement.**

I. INTRODUCTION

Sentiment analysis holds significant importance in comprehending customer attitudes toward products and services, particularly within the e-commerce sector, where reviews are plentiful and influential. Utilizing sophisticated text mining algorithms, companies can extract valuable insights into customer preferences, satisfaction levels, and areas requiring enhancement. The project aims to not only classify reviews but also to extract meaningful patterns and trends from the data, enabling businesses to make data-driven decisions. Additionally, the project explores the use of sentiment analysis in personalized marketing strategies, where understanding individual customer sentiments can lead to tailored marketing campaigns and enhanced customer engagement.

One of the key challenges in sentiment analysis is handling the nuances of human language, such as sarcasm, irony, and ambiguity. The project addresses these challenges by employing sophisticated natural language processing (NLP) techniques, including word embeddings and deep learning models, to capture the subtleties of language and improve the accuracy of sentiment classification. Furthermore, the project considers the ethical implications of sentiment

analysis, such as privacy concerns and bias in algorithms, and seeks to mitigate these issues through transparent and responsible data practices.

By empowering businesses with actionable insights derived from sentiment analysis, the project aims to drive innovation, improve customer satisfaction, and ultimately, contribute to the growth and success of businesses in the e-commerce industry. With the rapid evolution of technology and the increasing importance of customer feedback, sentiment analysis has become an indispensable tool for businesses looking to stay competitive in the digital age.

II. LITERATURE SURVEY

Yanrong Zhang, Jiayuan Sun, Lingyue Meng [1] proposed a method which focuses on analyzing consumer feedback within the context of e-commerce, with a particular emphasis on three key elements: keywords, evaluation objects, and emotional resources. The text comments are categorized using a constructed emotional dictionary. The methodology involves employing the tf-idf algorithm to identify keywords, extracting evaluation objects through a combination of part-of-speech methods and keyword similarity evaluation, and determining emotional resources based on part-of-speech characteristics and location rules.In the final step, an emotional dictionary is established using the identified emotional resources. Additionally, a reverse dictionary is created to account for the variation in polarity for the same emotional words across different evaluation objects. The experimental results confirm the effectiveness of the emotional dictionary in accurately classifying the emotions expressed in e-commerce text reviews.

Saju, Binju and Jose, Siji and Antony, Amal [2] says that the effectiveness of science and technology is realized when applied to solve real-world problems. Sentiment analysis emerges as a valuable solution for numerous business challenges, representing a convergence of various technologies and accumulated knowledge. It is evident that sentiment analysis is a multidisciplinary field, drawing insights from computer science, psychology, statistics, machine learning, data mining, and deep learning.

The field of sentiment analysis is continuously evolving and holds great potential. Technological advancements have propelled sentiment analysis from its traditional rule-based approaches to a more dynamic state, enhanced by advanced data mining and machine learning methods.

Manal Loukili, Fayçal Messaoudi, and Mohammed El Ghazi [3] explore the application of machine learning techniques in the context of e-commerce product review analysis. The study focuses on sentiment analysis, a crucial aspect of understanding customer feedback, and utilizes machine learning algorithms to categorize reviews as positive or negative. The authors aim to harness this sentiment analysis to enhance e-commerce recommendation systems. By mining and interpreting customer sentiments from reviews, the paper seeks to provide personalized product recommendations, thereby improving the overall shopping experience for customers and assisting e-commerce businesses in increasing sales and customer satisfaction. The paper likely delves into the 19 methodologies and results of their sentiment analysis-based recommendation system, offering valuable insights into the practical application of machine learning in the ecommerce domain.

Liu, B[4] has proposed a method focusing on sentiment analysis and opinion mining concentrating on discerning positive or negative sentiments within text. This book presents an abstract formulation of these issues, offering a structured comprehension and unveiling interconnected sub-problems. This abstraction assists researchers in understanding the problem and facilitates the development of more effective solutions. It also provides practitioners with insight into the requisite sub-problems and their relationships, guiding the design of practical systems.

A. Pak and P. Paroubek [5] proposed the 2010 paper titled "Twitter as a Corpus for Sentiment Analysis and Opinion Mining" in the LREc journal, in which the authors investigate the utility of Twitter as a valuable corpus for sentiment analysis and opinion mining. They highlight the immense potential of Twitter's expansive and ever-changing dataset as a real-time source of user-generated text, providing insights into public sentiment and opinions. The

paper delves into diverse techniques, incorporating natural language processing and machine learning, to analyze Twitter data and categorize text into positive, negative, or neutral sentiments. This paper provides insights into the challenges of using social media data for sentiment analysis, such as dealing with noise and context, and it highlights the benefits of Twitter as a practical resource for
understanding public sentiment.

M. Wahyudi and D. A. Kristiyanti [6], proposed a study on "Analyzing Sentiments in Smartphone Product Reviews Using Support Vector Machine Algorithm Enhanced by Particle Swarm Optimization" in the Journal of Theoretical & Applied Information Technology in 2016, which explores the use of support vector machines (SVM) in conjunction with a particle swarm optimization (PSO) algorithm for sentiment analysis in smartphone product reviews.The study demonstrates the effectiveness of this approach in classifying reviews into positive or negative categories, with the SVM-PSO model outperforming other machine learning methods. The research contributes to the field of product review
analysis by offering an innovative approach.

III. PROPOSED WORK

The proposed methodology includes several stages:
1- Data pre-processing
2- Polarity analysis
3-Text preprocessing
4-Word Cloud Analysis
The study begins by gathering experimental data from multiple e-commerce platforms. Each dataset is provided in Comma Separated Values (CSV) file format and can be accessed as a supplementary resource. The next stage involves preprocessing the data to eliminate stop words, punctuation, spaces, digits, and special characters. A text mining package is utilized for this purpose.

A. DATA PREPROCESSING

This module is primarily focused on handling a sizable dataset comprising more than 25,000 product reviews, each associated with a corresponding rating. To facilitate sentiment analysis, these customer ratings are categorized as positive (numerical value=1) when they exceed 3 and negative (numerical value=0) when they fall below that threshold. This dataset serves as the training data for the model. The process relies on popular Python libraries, such as WordCloud for text visualization, Pandas for dataset import, Scikit-learn for model integration, accuracy module for performance evaluation, TfidfVectorizer for feature extraction, and Matplotlib for data visualization. In the context of e-commerce product review systems, this approach allows for the systematic analysis of customer feedback and ratings, helping businesses understand sentiment trends and enabling the development of predictive models for more informed decision-making and improved customer experiences

B. POLARITY ANALYSIS

Manufacturers and sellers can use polarity analysis to assess the quality and performance of their products based on customer feedback. They can identify areas that need improvement and those that are performing well. Potential buyers can make more informed decisions by considering the polarity of reviews. They can weigh the pros and cons of a product based on the sentiments expressed by previous customers. Aggregating polarity scores across a range of reviews can reveal trends in customer sentiment over time. This can help in tracking product performance and understanding evolving consumer preferences

C. TEXT PREPROCESSING

Text preprocessing is crucial for preparing textual data in ecommerce product review systems. It involves tokenization, lowercasing, removing stop words and punctuation, and implementing stemming or lemmatization. This process transforms raw customer feedback into a structured format, ensuring more accurate sentiment analysis, keyword extraction, and data mining. Through the process of cleaning and organizing data, businesses can acquire dependable insights into customer opinions, pinpoint the strengths and weaknesses of their products, and make well-informed decisions to improve customer satisfaction and product quality.In e-commerce text preprocessing, removing stop words and applying stemming and lemmatization are essential. This

eliminates common words, focuses on meaningful content, and reduces words to their root forms, ensuring consistency and improving the accuracy and efficiency of sentiment analysis and text mining, ultimately aiding businesses in understanding customer feedback and making data-driven decisions.

D. WORD CLOUD ANALYSIS

This data preprocessing module focuses on refining a dataset of over 25,000 product reviews for sentiment analysis, categorizing ratings into positive and negative values. Leveraging Python libraries like WordCloud, Pandas, Scikit-learn, TfidfVectorizer, and Matplotlib, the process streamlines data handling, model integration, accuracy assessment, feature extraction, and visualization.The dataset obtained facilitates a methodical examination of customer feedback, assisting businesses in grasping trends in sentiment and constructing predictive models to support informed decision-making and improve customer experiences. Utilizing WordClouds visually displays word frequencies in 41 reviews, highlighting positive and negative sentiments. WordCloud analysis proves compelling in e-commerce, providing an instant overview of prevalent keywords. It condenses customer reviews, highlights critical product aspects, and enables businesses to swiftly discern trends and strengths/weaknesses, fostering data-driven decisions. This visual technique accelerates review analysis, helping e-commerce platforms refine offerings, boost customer satisfaction, and improve overall product quality based on prevailing sentiments expressed in reviews.

E. USE OF TEXT MINING ALGORITHM:

Text mining algorithms are specialized data mining algorithms used on unstructured text data that has been converted into a structured numerical format. These algorithms aim to offer insights into the processing of text without requiring human reading. This technology transforms the analysis of e-commerce product reviews by utilizing natural language processing and machine learning.It classifies sentiments, distinguishing between positive, negative, and neutral reviews for a nuanced understanding. Utilizing feature extraction, it identifies key aspects influencing customer

sentiments, aiding businesses in targeted improvements. The algorithm ensures real-time analysis, crucial in the dynamic e-commerce landscape, enabling swift adaptation to changing customer preferences. Moreover, it enhances recommendation systems by uncovering patterns in reviews, guiding personalized product suggestions. This data-driven approach promises to elevate product quality, customer satisfaction, and competitive advantage in online retail.

Figure 1 shows the proposed workflow diagram.
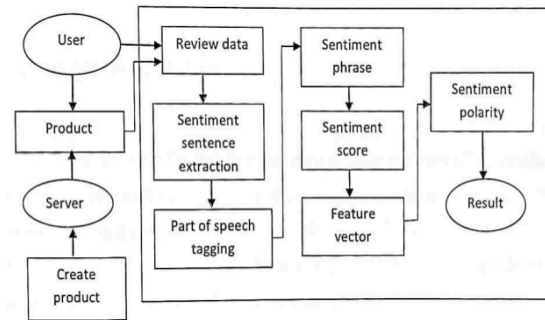
Figure 1: Proposed Workflow

IV. IMPLEMENTATION

The process begins with the importation of necessary libraries and datasets, setting the foundation for subsequent tasks. Following this, a crucial step involves preprocessing and cleaning the reviews to enhance the quality of the data. Subsequent to data refinement, an in-depth analysis of the dataset is conducted, shedding light on its key characteristics. The next phase revolves around converting textual information into vectors, a pivotal step in preparing the data for model training. The subsequent stages encompass model training, thorough evaluation, and prediction, completing the comprehensive workflow for leveraging and analyzing the provided datasets.

Figure 2 shows the dataset utilized in the process.

Figure 2: Dataset

Figure 3 shows the classification of words used in reviews into positive and negative word clouds.

Figure 3: WordCloud

Figure 4 shows the confusion matrix for performance evaluation.


```
[24] from sklearn import metrics
     from sklearn.metrics import confusion_matrix
     cm = confusion_matrix(y_test,pred)

     cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = cm,
                             display_labels = [False, True])

     cm_display.plot()
     plt.show()
```

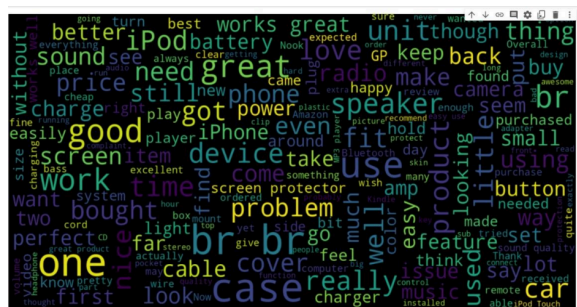**Figure 4: Confusion Matrix for performance evaluation**

The confusion matrix serves as a comprehensive summary, delineating the correct and incorrect

predictions made by a model and categorizing them into distinct classes. The visual depiction of this matrix distinctly segregates the outcomes based on their alignment with real values. The terminology within the confusion matrix includes True Positives (TP) and True Negatives (TN) for correct predictions, and False Positives (FP) and False Negatives (FN) for incorrect predictions.The conclusions drawn from the analysis reveal that the model outperforms not only individual models but also alternative ensemble approaches in terms of accuracy, sensitivity, and specificity.

V. CONCLUSION

The assessment of the suggested sentiment analysis model for e-commerce product reviews demonstrates its superior performance compared to both individual models and ensemble methods, particularly in terms of accuracy, sensitivity, and specificity. Particularly, Logistic Regression stands out for its superior performance, showcasing higher accuracy scores and better confusion matrix parameters compared to other models. This suggests its robustness and suitability for capturing nuanced sentiments across various product categories. The model's capability to enhance customer relationship management through improved sentiment analysis results is emphasized, underscoring its value in offering dependable insights for well-informed decision-making. Prospective avenues for research involve refining performance metrics and continually improving the model based on user feedback, with the goal of further optimizing its efficacy and influence in e-commerce environments.

VI. REFERENCES

[1] Yanrong Zhang, Jiayuan Sun, Lingyue Meng(2020), "SENTIMENT ANALYSIS OF E-COMMERCE TEXT REVIEWS BASED ON SENTIMENT DICTIONARY" in IEEE International Conference on Artificial Intelligence and Computer Applications.

[2]Saju, Binju and Jose, Siji and Antony, Amal (2020), "COMPREHENSIVE STUDY ON SENTIMENT ANALYSIS: TYPES, APPROACHES, RECENT, APPLICATIONS, TOOLS AND APIs" in

Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA)

[3]Manal Loukili, Fayçal Messaoudi, and Mohammed El Ghazi(2023), "SENTIMENT ANALYSIS OF PRODUCT REVIEWS FOR ECOMMERCE RECOMMENDATION BASED ON MACHINE LEARNING" in International Journal of Advances in Soft computing and its Application

[4]Liu, B.(2012), "SENTIMENT ANALYSIS AND OPINION MINING" in Synthesis Lectures on Human Language Technologies

[5] Pak and P. Paroubek, (2010) "Twitter as a corpus for sentiment analysis and opinion mining." in LREc, vol. 10, no. 2010, 2010.

[6] M. WAHYUDI and D. A. KRISTIYANTI (2016), "Sentiment analysis of smartphone product review using support vector machine algorithm-based particle swarm optimization." Journal of Theoretical & Applied Information Technology, vol. 91, no. 1,2016.

[7] V. Narayanan, I. Arora, and A. Bhatia (2013), "Fast and accurate sen-timent classification using an enhanced naive bayes model," in International Conference on Intelligent Data Engineering and Automated Learning. Springer, pp. 194–201

[8] Manal Loukili, F. Messaoudi, Mohammed El Ghazi (2023), "Sentiment Analysis of Product Reviews for E-Commerce Recommendation based on Machine Learning", in International Journal of Advances in Soft computing and its Application.

[9] Zad, S., Heidari, M., Jones, J. H., & Uzuner, O. (2021). A survey on concept level sentiment analysis techniques of textual data. In IEEE World AI IoT Congress (AIIoT) (pp. 0285-0291). IEEE. 64

[10] Ainin, S., Feizollah, A., Anuar, N. B., & Abdullah, N. A. (2020). Sentiment analyses of multilingual tweets on halal tourism. Tourism Management Perspectives, 34, 100658.

[11] Keikhosrokiani, P., & Pourya Asl, M. (Eds.). (2022). Handbook of Research on Opinion Mining and Text Analytics on Literary Works and social media. IGI Global.

[12] Chen, X., Xue, Y., Zhao, H., Lu, X., Hu, X., & Ma, Z. (2019). A novel feature extraction methodology for sentiment analysis of product reviews. Neural Computing and Applications, 31(10), 6625-6642