Comparative Analysis of Sentiment Classification Techniques on Flipkart Product Reviews: A Study Using Logistic Regression, SVC, Random Forest, and Gradient Boosting

Henderi[1,*], Quba Siddique[2]

[1]Informatics Engineering, University of Raharja, Tangerang 15117, Indonesia

[2]Institute of Banking and Finance, Bahauddin Zakariya University Multan, Pakistan

ABSTRACT

Sentiment analysis plays a crucial role in e-commerce, providing valuable insights from customer reviews on platforms like Flipkart. This study aims to compare the effectiveness of various sentiment classification techniques, specifically Logistic Regression, Support Vector Classifier (SVC), Random Forest, and Gradient Boosting. The dataset, collected from Flipkart, consists of 205,052 product reviews spanning various categories. Key data preprocessing steps included handling missing values, removing duplicates, normalizing text, and applying TF-IDF vectorization for feature extraction. We implemented and tuned the hyperparameters for each algorithm using grid search and randomized search. The data was divided into training and testing sets with an 80-20 split, and cross-validation techniques ensured robust model evaluation. The performance of each model was assessed using several metrics: accuracy, precision, recall, F1-score, and ROC-AUC. The results revealed that Logistic Regression achieved an accuracy of 0.8995, precision of 0.8773, recall of 0.8995, an F1 score of 0.8736, and a ROC AUC score of 0.9105. The SVC model showed slightly higher accuracy at 0.8997, precision of 0.8619, recall of 0.8997, and an F1 score of 0.8738. The Random Forest model, while robust, had lower accuracy (0.7953) and struggled with precision (0.6326), recall (0.7953), and an F1 score of 0.7047, but achieved a ROC AUC score of 0.9037. Gradient Boosting performed comparably to Logistic Regression with an accuracy of 0.8993, precision of 0.8512, recall of 0.8993, an F1-score of 0.8735, and a ROC AUC score of 0.9098. Comparative analysis identified SVC and Logistic Regression as top performers, balancing accuracy and computational efficiency. These findings suggest that implementing these models can significantly enhance sentiment analysis in e-commerce, improving customer insights and business strategies. Future research should explore advanced deep learning techniques and address class imbalances to further refine sentiment analysis capabilities.

Keywords Sentiment Classification, Logistic Regression, SVC, Random Forest, Gradient Boosting

Submitted 10 January 2024
Accepted 20 April 2024
Published 1 June 2024

Corresponding author
Henderi, henderi@raharja.info

Additional Information and Declarations can be found on page 41

DOI: 10.47738/jdmdc.v1i1.4

INTRODUCTION

Sentiment analysis is crucial in decoding the vast data generated by customer interactions and opinions on e-commerce platforms. Sentiment analysis is a valuable tool for businesses across various domains. It allows them to monitor online conversations about their brand, products, or services [1]. By analyzing sentiments, businesses can gain insights to make informed decisions,

How to cite this article: Henderi, Q. Siddique, "Comparative Analysis of Sentiment Classification Techniques on Flipkart Product Reviews: A Study Using Logistic Regression, SVC, Random Forest, and Gradient Boosting," *J. Digit. Mark. Digit. Curr.*, vol. 1, no. 1, pp. 21-42, 2024.

understand customer feedback, and track market trends [2]. Sentiment analysis helps identify the emotions or attitudes expressed towards a company, its offerings, personnel, or events [3]. It can also help pinpoint strong and weak performance points of businesses and services, facilitating the formulation of action steps to enhance performance [4].

The application of sentiment analysis extends to different areas within business intelligence, supporting decision-making processes and mining consumer insights to enhance business performance [5]. Sentiment analysis is crucial in business strategy planning, providing valuable insights from blogs to aid decision-making [6]. Furthermore, sentiment analysis is a critical sub-area in Natural Language Processing research, offering valuable insights to customers, business owners, and stakeholders [7].

Sentiment analysis is essential for maintaining a positive image in branding. Technology, particularly sentiment analysis tools, is crucial in managing negative opinions to uphold a favorable brand perception [8]. Additionally, sentiment analysis aids in understanding public opinions about products or services offered, enabling businesses to conduct research and make informed decisions [9].

Businesses can align their strategic direction by analyzing sentiments from online user reviews to enhance customer satisfaction and improve service quality [10]. This understanding is pivotal as it directly influences purchasing decisions. Consumers frequently rely on the sentiments expressed in product reviews to make informed purchasing choices, trusting peer evaluations to guide their shopping behavior. Sentiment analysis of these reviews is crucial in e-commerce, where customers cannot physically inspect products before buying [11]. By analyzing sentiments, businesses can understand consumer preferences, predict needs, and improve products and services [12], [13].

Product review sentiment analysis helps extract emotional attitudes towards specific product features, aiding decision-making processes [14]. This analysis also predicts consumer satisfaction, leading to product and service enhancements [15]. Businesses use sentiment analysis to accurately grasp consumer perceptions, identify improvement areas, and meet customer expectations [16]. This sentiment mining is vital in decision-making, especially in e-commerce, where online reviews significantly impact consumer behavior [17].

The application of sentiment analysis tools automates the classification of customer reviews into categories such as positive, negative, or neutral. This automation facilitates businesses in swiftly identifying customer sentiments and trends, allowing them to tailor their marketing strategies effectively. For example, a surge in positive reviews on a particular product can trigger targeted promotions. At the same time, an aggregation of negative feedback can prompt quick remedial measures, ensuring customer satisfaction and loyalty.

The global e-commerce sector has experienced significant growth over the past decade, reshaping the retail landscape. This surge in e-commerce has transformed how businesses operate and how consumers shop. According to eMarketer, global e-commerce sales were projected to reach 2.356 trillion U.S. dollars in 2018, nearly doubling in 2013 [18]. This growth has encouraged brands to adopt direct-to-consumer (DTC) models, enhancing online marketing, selling, and direct consumer transactions [19].

The expansion of e-commerce has not only influenced sales volumes. Still, it has also necessitated the development of specialized warehouses tailored to meet the demands of online retailers serving end-customers in the business-to-consumer (B2C) segment [20]. Retailers across various sectors, including food, fashion, and electronics, have broadened their online presence, increasingly relying on e-commerce for business operations [21]. This shift has disrupted traditional retail environments, leading more consumers to choose online shopping [22].

E-commerce growth has extended beyond discretionary items to essential goods like household items and groceries, indicating a broader adoption of e-commerce as a consumption paradigm [23]. The emergence of e-commerce has significantly impacted all retail industries, providing digital alternatives to traditional brick-and-mortar retail operations [24]. This transition has been particularly evident during the COVID-19 pandemic, with major retail companies in various industries embracing e-commerce as a crucial retailing channel [25].

Flipkart is a leading e-commerce company based in India that has shown significant progress in the online retail sector. Flipkart, established in 2007, is a testament to significant success in e-commerce within India. As one of the leaders in the Indian e-commerce market, Flipkart offers a diverse array of products ranging from electronics to home furnishings. With over 300 million registered users and a robust seller base, Flipkart substantially impacts India's digital economy. The platform's extensive reach and deep market penetration make it an ideal subject for studying the impacts of sentiment analysis on e-commerce practices. The company's growth from its establishment to its acquisition by Walmart has been driven by strategic decisions [26]. Flipkart has had a substantial impact on the Indian retail sector, adapting to changes in foreign direct investment policies and compliance strategies alongside other major e-commerce competitors like Amazon [27].

Customer reviews on e-commerce platforms are a crucial source of information that significantly influences others' purchasing behavior. These reviews attract visitor traffic to companies and serve as a valuable source of information accessed by buyers [28]. These reviews significantly impact a business's reputation and its sales volumes. Positive reviews often translate into increased sales as they enhance the perceived reliability and quality of the product or service. Conversely, negative reviews can deter potential customers and substantially harm a business's reputation if not managed correctly.

Positive reviews can reinforce a customer's decision to purchase, creating a psychological reward mechanism by affirming the decision's correctness. Negative reviews, however, can invoke doubt and hesitation, often leading to abandoned purchases. Economically, the influence of reviews extends to affecting the overall brand perception and long-term business revenue. Companies that actively engage with customer reviews and adapt their offerings accordingly tend to experience better customer retention rates and enhanced brand loyalty.

Sentiment analysis, while transformative, presents several inherent challenges that can hinder its effectiveness, especially in the diverse and dynamic environment of e-commerce. One primary challenge is detecting sarcasm. Sarcasm or irony in text can often be subtle and context-dependent, making it difficult for algorithms, which typically rely on straightforward keyword cues, to interpret accurately. This can lead to misclassifications, where positive sentiments are erroneously tagged as negative and vice versa. Another significant hurdle is handling mixed sentiments within a single review. Customers often list both pros and cons in their reviews, or their sentiments may shift from positive to negative from one sentence to another. Current tools may need help to recognize and correctly classify these shifts, leading to a diluted or incorrect understanding of the customer's overall sentiment.

Moreover, processing multilingual reviews adds a layer of complexity. E-commerce platforms cater to a global audience, resulting in reviews in multiple languages. Each language has its idioms, colloquialisms, and syntax, which can pose significant challenges for sentiment analysis tools not specifically tuned for multilingual capabilities. The limitations highlighted pose a significant barrier to leveraging sentiment analysis effectively. There is a specific need for more accurate sentiment analysis models that can adeptly handle the nuances of human language. These models must understand context, detect subtle cues like sarcasm, and accurately represent mixed sentiments within texts.

To address these issues, research must focus on developing and comparing various machine learning algorithms that can improve the accuracy and reliability of sentiment analysis. By doing so, businesses can better harness the power of customer feedback to refine their products and services, ensuring a better market fit and customer experience. The main goal of this research is to critically evaluate and compare the effectiveness of various machine learning algorithms—namely Logistic Regression, SVC, Random Forest, and Gradient Boosting—in accurately classifying the sentiments expressed in product reviews on Flipkart. This comparison aims to identify which algorithms perform best in accuracy, reliability, and efficiency, providing actionable insights that can be applied to improve sentiment analysis practices in e-commerce settings. Specific Aims is:

1) Determine which algorithm provides the highest accuracy in sentiment classification. This involves comparing each algorithm's precision, recall, and F1-score when applied to the same dataset of product reviews. Accuracy is critical as it directly impacts the reliability of sentiment analysis results used by businesses to make informed decisions.
2) Investigate the capability of each algorithm to handle mixed sentiments effectively. Reviews often contain mixed emotions, and the ability of an algorithm to discern and accurately classify these sentiments is crucial for a nuanced understanding of customer feedback. This study aims to identify the algorithm that best recognizes and processes complex sentiment structures within text data.
3) Evaluate the processing speed of each algorithm in analyzing and classifying sentiments. In the fast-paced world of e-commerce, where real-time data processing can significantly influence business strategies, the speed at which sentiment analysis can be performed is as important as its accuracy. Faster processing times can enhance the agility of business responses to customer feedback.

This research will significantly enhance the existing body of knowledge in sentiment analysis by systematically comparing the effectiveness of various machine learning algorithms, including Logistic Regression, SVC, Random Forest, and Gradient Boosting. By conducting a detailed evaluation of these models in classifying sentiments within e-commerce product reviews, the study offers comprehensive insights into how different algorithms perform under identical conditions, highlighting the strengths and weaknesses of each. This comparative analysis contributes a deeper understanding of the models most suitable for specific sentiment analysis tasks. Furthermore, the findings could lead to potential improvements in sentiment analysis techniques, such as the development of hybrid models or new methodologies that combine the high accuracy of one algorithm with the speed of another or better adaptability to handle mixed or complex sentiments effectively.

The practical implications of this research are substantial, especially for e-commerce businesses looking to enhance their customer relationship management (CRM) and marketing strategies. By identifying the most effective sentiment analysis tools, companies can better understand customer feelings and preferences, leading to more tailored marketing efforts and improved customer engagement. Additionally, the results will be invaluable for software developers and data scientists working to create more sophisticated sentiment analysis tools. The insights gained from this study can guide the development of algorithms more adept at processing nuanced human language, making these tools more robust and applicable to real-world e-commerce platforms like Flipkart.

Literature Review

Overview of Sentiment Analysis

Sentiment analysis, a crucial subfield of natural language processing (NLP), involves the computational study of opinions, sentiments, and emotions expressed in text. It aims to determine the attitude of a speaker or a writer concerning some topic or the overall contextual polarity of a document. The scope of sentiment analysis extends across various data sources such as social media posts, reviews, news, and blogs, making it a versatile tool for understanding human emotions in written language. In various industries, sentiment analysis is pivotal in gauging consumer reactions, monitoring brand reputation, and gathering insights from customer feedback. Its application is particularly significant in e-commerce, where understanding customer sentiments can directly influence purchasing behavior and brand loyalty. By analyzing sentiments, companies can swiftly address customer concerns, tailor their marketing approaches, and adapt their product offerings to meet consumer demands better.

In e-commerce, sentiment analysis is primarily utilized to analyze customer reviews and feedback on various platforms. These reviews are rich sources of insight, providing direct feedback on customer satisfaction and preferences. Sentiment analysis tools help automate categorizing these reviews into positive, negative, or neutral sentiments, thus allowing businesses to handle large volumes of data efficiently. The impact of sentiment analysis in e-commerce extends to several key areas, such as:

1) By understanding customer sentiments, marketing teams can create more targeted campaigns that resonate better with the audience. Positive sentiments can be leveraged to promote products more aggressively, while negative sentiments can help identify areas for improvement.

2) Sentiment analysis provides direct feedback on what features or aspects of a product are well-received and which are not. This information is invaluable

in guiding product development and innovation.

3) Analyzing sentiments in customer feedback helps prioritize customer service actions.

For example, a surge in negative reviews on a particular issue can trigger immediate customer service responses to address the concerns, potentially turning a negative customer experience into a positive one. Overall, applying sentiment analysis in e-commerce enhances understanding of customer preferences and satisfaction and contributes to more strategic decision-making, ultimately leading to improved business outcomes.

Several significant milestones and technological advancements have marked the evolution of sentiment analysis. Initially focused on simple polarity detection (positive, negative, neutral), the field has expanded with the advent of machine learning and deep learning techniques. Early methods relied heavily on manually curated lexicons and basic statistical techniques. Still, the introduction of machine learning in the early 2000s represented a paradigm shift, enabling more nuanced analysis based on large datasets.

Foundational studies in sentiment analysis often deal with smaller, more controlled datasets, such as movie reviews or product reviews from specific websites. Developing datasets like the Movie Review Data and, later, more complex ones like the Multi-Domain Sentiment Dataset provided researchers with resources to train and test more sophisticated models. These studies laid the groundwork for current methodologies, introducing concepts like sentiment lexicons, n-gram models, and, eventually, sentiment-specific word embeddings.

Research focusing on sentiment analysis in e-commerce settings has grown substantially with the increase in online shopping and the availability of consumer-generated data. Studies have specifically examined how sentiment analysis can be applied to product reviews to predict consumer purchasing patterns or assess overall customer satisfaction. For example, research has demonstrated that reviews containing mixed sentiments provide deeper insights into consumer preferences than uniformly positive or negative reviews.

These settings have explored various approaches, ranging from traditional machine learning models like Naive Bayes and SVM to more recent deep learning models like LSTM and convolutional neural networks. Key findings from these studies often highlight the importance of context and the ability of certain models to capture subtle linguistic cues better than others. The effectiveness of hybrid models, which combine several machine learning techniques, has also been a notable area of focus, showing promise in handling the complexities of customer reviews more efficiently.

One of the common challenges in sentiment analysis within e-commerce is handling large datasets containing diverse expressions, slang, and sometimes multiple languages. Detecting nuanced expressions such as sarcasm or irony remains a significant hurdle, as these often require understanding context beyond the text. Previous studies have proposed various solutions to these challenges. Techniques such as transfer learning, where a model trained on one type of data is adapted to another, have effectively improved handling large, diverse datasets. Enhancements in natural language processing, such as developing context-aware sentiment analysis models that utilize textual and contextual information, have also been pivotal.

Moreover, integrating user-generated metadata (like user ratings) with textual reviews has been explored to enhance the accuracy of sentiment analysis tools. The success of these approaches varies, with improvements often balanced by increased computational complexity or the need for more sophisticated data preprocessing techniques.

Logistic Regression

Logistic Regression is a statistical model that, despite its name, is commonly used for classification tasks, particularly binary classification. In the context of sentiment analysis, Logistic Regression estimates the probabilities of a particular piece of text (like a product review) being positive or negative by using a logistic function. It is particularly useful due to its simplicity and efficiency in handling linear relationships between the dependent and independent variables.

One of the main strengths of Logistic Regression in sentiment analysis is its interpretability. The model provides coefficients directly related to the importance of input features, making it easier to understand which words or phrases are most influential in determining sentiment. However, its weaknesses lie in its assumption of linearity between the dependent and independent variables and its struggle with the complexities of natural language, such as sarcasm, idioms, or context-dependent expressions.

Support Vector Classifier

The SVC is part of the Support Vector Machines (SVMs) family, which are more complex and powerful models designed for classification and regression tasks. SVC works by finding the hyperplane that best divides a dataset into classes. In sentiment analysis, SVC helps classify positive, negative, or neutral text by maximizing the margin between the dataset's classes.

SVC is highly effective in high-dimensional spaces, making it suitable for sentiment analysis, where text data is often transformed into high-dimensional vectors. In e-commerce contexts, SVC can robustly handle varied and complex datasets, such as customer reviews. However, the performance of SVC can sometimes be impacted by the choice of the kernel function, and it generally requires more computational resources, making it less efficient than simpler models when dealing with very large datasets.

Random Forest

Random Forest is an ensemble learning method that operates by constructing many decision trees at training time and outputting the class, which is the mode of the classes (classification) of the individual trees. It is known for its high accuracy, robustness, and ease of use. In sentiment analysis, Random Forest aggregates the decisions of multiple trees to improve predictive accuracy and control over-fitting.

The primary advantage of using Random Forest in sentiment analysis is its ability to effectively handle large datasets with a mixture of numerical and categorical data. It is also relatively immune to overfitting and can model complex interactions between words in text data. However, challenges include its model complexity, which can lead to extensive computational demands, especially with very large datasets, and difficulties interpreting the model compared to simpler models like Logistic Regression.

Gradient Boosting

Gradient Boosting is an advanced ensemble technique known for its predictive power and accuracy, which builds models in a stage-wise fashion and generalizes them by allowing optimization of an arbitrary differentiable loss function. In sentiment analysis, Gradient Boosting can incrementally build an ensemble of models to improve prediction accuracy over iterations.

Gradient Boosting is particularly effective in scenarios where precision is critical, as it fine-tunes predictions with each iteration. This feature is beneficial in sentiment analysis for adjusting to subtle nuances in textual data. However, its drawbacks include being prone to overfitting if not tuned properly and requiring careful configuration along with high computational resources. Its complex nature makes the model more challenging to interpret than straightforward algorithms.

Data Mining in Marketing

Data mining involves using statistical, mathematical, and computational techniques to discover patterns and insights from large datasets. In marketing analytics, data mining is crucial for extracting actionable information to inform and optimize marketing strategies.

The importance of data mining in marketing analytics lies in its ability to process vast amounts of data and uncover hidden patterns and trends. By leveraging these techniques, marketers can gain deeper insights into customer behavior, preferences, and interactions, enabling them to design more effective and targeted campaigns. Numerous studies have explored the use of machine learning models to predict key marketing outcomes. These studies have demonstrated the potential of machine learning to enhance campaign effectiveness, predict customer behavior, and optimize marketing spend.

Research papers and case studies have shown that machine learning models can accurately predict marketing KPIs such as conversion rates, customer lifetime value, and churn rates. For instance, predictive models based on Decision Trees and Random Forests have been used to identify factors influencing customer engagement and conversion, providing valuable insights for campaign optimization. SVM and regression models have also been employed to forecast sales and revenue, enabling marketers to make data-driven decisions about resource allocation.

The findings from these studies highlight the strengths and limitations of different machine learning approaches in marketing analytics. For example, while Decision Trees and Random Forests offer high interpretability and accuracy, they may require extensive computational resources for large datasets. Regression models provide straightforward predictions but may need help with complex, non-linear relationships in the data. Clustering techniques are effective for segmentation but may require careful tuning of parameters to achieve optimal results.

Overall, the application of machine learning in marketing analytics has proven to be highly effective in enhancing the precision and efficiency of marketing strategies. By reviewing these studies, marketers can better understand the capabilities of various machine learning techniques and select the most appropriate methods for their specific needs. This knowledge enables them to leverage data-driven insights to drive better marketing outcomes and achieve higher ROI.

Decision Trees

Decision Trees are supervised learning algorithms used for classification and regression tasks. They are structured as a tree with nodes, branches, and leaves. The root node represents the entire dataset, split into branches based on specific criteria. Each branch represents a decision rule, leading to further splits until reaching the leaf nodes, representing the outcome or classification.

The process of building a Decision Tree involves several key steps. Initially, the dataset is split based on the best attribute that maximizes the separation of the classes. This splitting criterion can be based on measures such as Gini impurity, entropy, or variance reduction, depending on whether the task is classification or regression. As the tree grows, it may become overly complex and prone to overfitting. Pruning methods are employed to cut back the tree, removing branches that have little importance in predicting the target variable, thus enhancing the tree's generalizability. Decision Trees can handle categorical and continuous variables, making them versatile for various data types. The advantages of Decision Trees include their simplicity and interpretability. As the tree structure represents the decision-making process, they are easy to understand and visualize. This makes them particularly useful for explaining model predictions to non-technical stakeholders. Decision Trees do not require extensive data preprocessing, such as normalization or scaling, making them straightforward to implement.

Decision Trees are widely used in marketing analytics for various applications. One common use case is customer segmentation, where Decision Trees classify customers into different segments based on their behaviors and attributes. This allows marketers to tailor their strategies and offers to specific customer groups, enhancing the relevance and effectiveness of marketing efforts.

Another significant application is churn prediction. By analyzing historical customer data, Decision Trees can identify patterns and factors that lead to customer attrition, enabling businesses to address the reasons behind churn and implement retention strategies proactively. Additionally, Decision Trees are used to target personalized marketing offers. By understanding which attributes influence customer decisions, marketers can design personalized campaigns that resonate more with individual customers, increasing engagement and conversion rates.

Several studies have demonstrated the effectiveness of Decision Trees in improving campaign targeting and ROI. For instance, research has shown that Decision Trees can accurately predict which marketing messages will be most effective for different customer segments, leading to higher conversion rates and better resource allocation. These studies underscore the practical benefits of using Decision Trees for data-driven marketing strategies.

However, Decision Trees also have challenges and limitations. One major issue is overfitting, where the model becomes too complex and captures noise in the data rather than the underlying pattern. This can lead to poor performance on new, unseen data. Additionally, Decision Trees can be sensitive to variations in the data; small changes in the dataset can result in significantly different trees. To mitigate these issues, techniques such as pruning and ensemble methods (e.g., Random Forests) are often used to enhance the robustness and accuracy of Decision Trees. Despite these challenges, Decision Trees remain a valuable tool in marketing analytics, offering clear insights and actionable recommendations for optimizing marketing strategies.

Random Forests

Random Forests are an ensemble learning method that combines multiple Decision Trees to improve predictive performance. As an ensemble method, Random Forests build a collection of Decision Trees during training and merge their outputs to make the final prediction. The core idea is that by averaging the results of multiple trees, the model can achieve better accuracy and robustness than any individual tree alone.

The process of creating a Random Forest involves several steps. First, bootstrap sampling is used to generate multiple subsets of the training data. Each subset is created by randomly selecting samples from the original dataset with replacement, ensuring diversity among the subsets. For each subset, a Decision Tree is constructed. During the construction of each tree, a random selection of features is used at each split, rather than considering all features. This process, known as random feature selection, helps to decorrelate the trees and reduce overfitting. Once all the trees are built, the final prediction is made by aggregating the results of all the trees, typically through majority voting for classification tasks or averaging for regression tasks.

The advantages of Random Forests are numerous. They offer higher accuracy and robustness than single Decision Trees, as the ensemble approach reduces the likelihood of overfitting and captures more complex patterns in the data. Random Forests are also well-suited for handling large datasets with numerous features, as the random feature selection process helps manage the computational complexity. Additionally, Random Forests can handle missing values and maintain good performance even when some features are uninformative or noisy.

Random Forests provide several advantages over single Decision Trees. One of the most significant benefits is improved performance through ensemble averaging. By combining the predictions of multiple trees, Random Forests mitigate the risk of overfitting that individual Decision Trees often face. This ensemble approach leads to more stable and accurate predictions. Random Forests also enhance predictive power and generalization to unseen data. The random sampling of both data and features ensures that the individual trees are diverse and less likely to replicate each other's errors. This diversity enables the Random Forest model to generalize new data better, providing more reliable predictions.

Regarding computational efficiency and scalability, Random Forests can be more efficient than other complex models, especially when implemented with parallel processing techniques. Each tree in the forest can be built independently, allowing for efficient use of computational resources. Despite their complexity, Random Forests scale well with the size of the dataset and the number of features, making them suitable for large-scale applications in marketing analytics. Random Forests have been successfully applied in various marketing analytics scenarios. One notable application is in predicting customer lifetime value (CLV). By analyzing historical customer data, Random Forests can identify patterns that indicate high-value customers, enabling businesses to tailor their marketing efforts to maximize long-term revenue.

Another application is in optimizing ad spend. Random Forests can analyze the performance of past advertising campaigns across different channels and determine the factors that lead to higher returns on ad spend. This insight allows marketers to allocate their budgets more effectively, focusing on the most profitable strategies.

Random Forests are also used to forecast sales. By incorporating various predictors such as past sales data, market trends, and customer behavior, Random Forests can provide accurate sales forecasts, helping businesses manage inventory and plan marketing activities more effectively. Case studies have demonstrated the effectiveness of Random Forests in improving campaign performance metrics and ROI. For example, a company might use Random Forests to analyze customer engagement data and identify the most effective marketing messages for different segments. This targeted approach can lead to higher conversion rates and overall campaign performance.

Random forests' flexibility and adaptability make them valuable in various marketing scenarios. They can handle various data types and structures, making them suitable for diverse marketing tasks. Whether used for segmentation, prediction, or optimization, Random Forests give marketers powerful tools to enhance their decision-making processes and achieve better outcomes.

Comparative Studies

Numerous studies have compared the performance of Decision Trees and Random Forests across various domains, providing insights into their relative strengths and weaknesses. These comparisons typically focus on key metrics such as accuracy, precision, and computational efficiency.

Studies in finance have shown that Random Forests generally outperform Decision Trees in predicting credit risk and stock market trends. Random Forests' ability to handle high-dimensional data and their robustness against overfitting contribute to their accuracy and reliability in financial predictions. Research comparing these models for disease diagnosis and patient outcome predictions in the healthcare domain has demonstrated that Random Forests provide higher accuracy and better generalization to new patient data. The ensemble nature of Random Forests helps in capturing complex interactions between medical variables, leading to more precise predictions.

In e-commerce, comparisons of Decision Trees and Random Forests for tasks such as customer segmentation, purchase prediction, and personalized recommendations have revealed that Random Forests often achieve higher predictive performance. Their capability to manage diverse and large datasets makes them well-suited for the dynamic and data-rich e-commerce environment. Studies in customer relationship management (CRM) have highlighted the advantages of Random Forests in predicting customer churn and lifetime value. The model's ability to incorporate a wide range of customer behavior indicators and its resilience to noisy data result in more accurate and actionable insights for CRM strategies. These comparative studies consistently find that while Decision Trees offer simplicity and interpretability, Random Forests provide enhanced predictive power and robustness, particularly in complex and high-dimensional data scenarios.

Despite the extensive research comparing decision trees and random forests in various fields, more comprehensive studies need to focusing specifically on their application in predicting digital marketing ROI. Most existing research in marketing analytics has concentrated on general predictive tasks such as customer segmentation, churn prediction, and sales forecasting, rather than the direct prediction of ROI. This gap highlights the need for more research in digital marketing ROI prediction. Understanding how these models perform in this context is crucial for marketers seeking to optimize their campaign strategies and budget allocations. There is a pressing need to evaluate the accuracy, precision, and practical applicability of Decision Trees and Random Forests for predicting ROI in digital marketing campaigns, considering this field's unique challenges and data characteristics.

Future research should fill this gap by conducting detailed comparative studies focusing on digital marketing datasets and ROI as the primary outcome. Potential areas for future research include exploring the impact of different feature selection techniques on model performance, assessing the scalability of these models in real-time marketing environments, and developing hybrid approaches that combine the strengths of both Decision Trees and Random Forests. Addressing these gaps will contribute significantly to marketing analytics, providing marketers with deeper insights and practical recommendations. By leveraging advanced predictive models to forecast ROI accurately, businesses can enhance their marketing effectiveness, improve resource allocation, and achieve better overall performance in their digital marketing efforts.

Method

The methodology of this study is visually represented in a flowchart, covering each major step from data collection and preprocessing to model implementation and evaluation, as shown in figure 1 below:

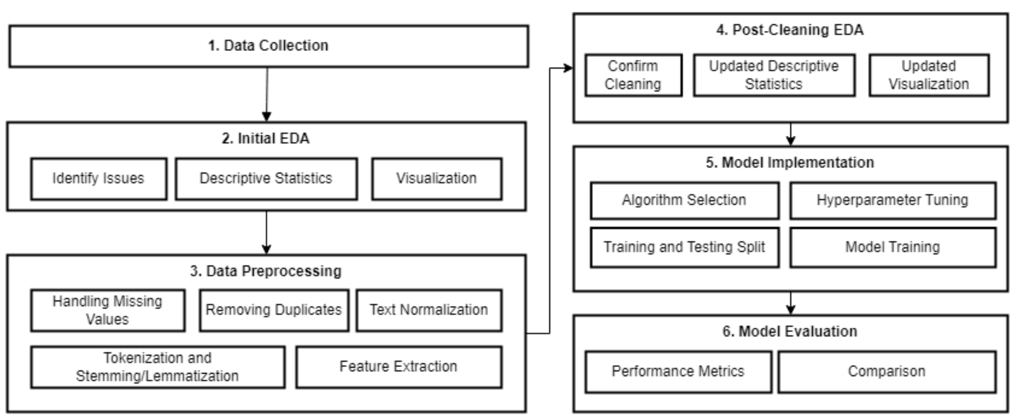

Figure 1 Research Methodology Flowchart

In the first stage, Data Collection, product reviews are collected from Flipkart using web scraping techniques with BeautifulSoup. This initial step is crucial for gathering the raw data necessary for the subsequent analysis. The second stage begins with an initial EDA to understand the raw data and identify potential issues, such as missing values or duplicates. The third stage follows the Data Preprocessing as data cleaning process, which involves removing duplicates, handling missing values, and normalizing the text. The text data is then tokenized and stemmed/lemmatized to split into tokens and reduce words to their base forms. Feature extraction uses techniques like TF-IDF vectorization to convert the text into numerical features. The fourth stage, a post-cleaning EDA, is conducted to confirm the cleaning steps' effectiveness and further explore the cleaned data.

In the fifth stage, Model Implementation, several machine learning algorithms are selected for comparison, including Logistic Regression, SVC, Random Forest, and Gradient Boosting. Hyperparameter tuning is performed to optimize the performance of each algorithm. The data is then divided into training and testing sets, and cross-validation is employed to ensure robust model evaluation. Each model is trained using the prepared data, setting the stage for performance evaluation. The final stage, Model Evaluation, involves assessing the performance of each model using various metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. A comparative analysis is conducted to identify which model performs best overall and in specific areas. This structured approach ensures a comprehensive analysis of the sentiment classification techniques applied to Flipkart product reviews.

Data Collection

The dataset for this study was collected from Flipkart, a leading e-commerce platform, through web scraping using the BeautifulSoup library in December 2022. This dataset comprises 205,053 rows and 6 columns, encompassing a wide range of product categories, including electronics, clothing, home decor, and automated systems, totaling 104 different types of products. Each row in the dataset represents a customer review and includes the following features: Product Name, Product Price, Rate, Reviews, Summary, and Sentiment. The Product Name refers to the name of the product being reviewed, providing a reference to the specific item discussed in the review. The Product Price indicates the cost of the product at the time of the review, offering context on the product's market segment. The Rate is the rating given to the product, usually on a scale (1 to 5 stars), reflecting the customer's overall satisfaction.

The Reviews feature contains the full text of the customer review, capturing detailed feedback and opinions about the product. The Summary provides a review summary, offering a condensed version of the customer’s feedback. Lastly, the Sentiment feature categorizes the review's sentiment as positive, neutral, or negative, indicating the overall tone. This sentiment classification is essential for the sentiment analysis performed in this study.

Initial Exploratory Data Analysis (EDA)

The dataset consists of 205,052 entries with 6 columns: product_name, product_price, Rate, Review, Summary, and Sentiment. An initial exploration of the dataset reveals various characteristics. First, the dataset contains significant missing values, with the Review column having 24,664 missing entries and the Summary column having 11 missing entries. Additionally, 34,375 duplicate rows need to be addressed to ensure data integrity. Some anomalies are also present in the Rate column, where unexpected entries such as product names indicate potential data entry errors.

Descriptive statistics provide a summary of the dataset's numerical and categorical features. The product_price column includes 525 unique prices for numerical features, with the most frequent price being 1299, appearing 9,150 times. The Rate column predominantly features the rating 5, which appears 118,765 times, followed by ratings of 4, 3, 2, and 1. In terms of categorical features, the product_name column contains 958 unique product names, with the most common product being "cello Pack of 18 Opalware Cello Dazzle Lush Fiesta Opalware Dinner Set, 18 Pieces Dinner Set" appearing 6,005 times. The Sentiment column shows a significant imbalance, with 166,581 positive entries, 28,232 negative entries, and 10,239 neutral entries. Figure 2 below show the sentiment distribution clearly.

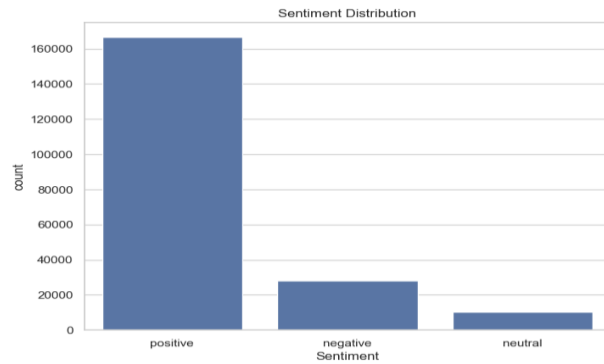

Figure 2 Sentiment Distribution Bar Plot

Various visualizations are employed to gain deeper insights into the data distribution and relationships. Histograms for numerical features, such as product_price and Rate, illustrate their distribution. Bar plots for categorical features, including product_name and Sentiment, visually represent their frequency distribution. Scatter plots are used to explore the relationship between product_price and Rate, helping to identify any potential correlations or patterns.

Data Preprocessing

In the data cleaning process, several steps were undertaken to prepare the dataset for analysis. Initially, missing values were handled by removing rows that contained missing reviews or sentiments to ensure that the data used for analysis was complete and reliable. Duplicate entries were then removed to maintain the integrity of the dataset, preventing any bias or skewing of results due to repeated data points. Text normalization was performed by converting all text to lowercase and removing punctuation, which helped standardize the text data for further processing.

Tokenization and stemming/lemmatization were crucial steps in preparing the textual data for analysis. The tokenization process involved splitting the text into individual words or tokens, making it easier to analyze the structure and content of the text. Stemming or lemmatization was applied to reduce words to their base or root forms. This step helped minimize the variations of words with the same meaning, thus simplifying the text data and enhancing the performance of the machine learning models.

Feature extraction was done using TF-IDF vectorization, which converts text into numerical features based on term frequency and inverse document frequency. TF-IDF vectorization transformed the textual data into a numerical form that could be used as input for machine learning algorithms. This method assigned weights to each term in the text, reflecting its importance in the document relative to the entire dataset, thereby enabling the models to understand better and analyze the text data.

Post-Cleaning EDA

The post-cleaning data exploration involved conducting an EDA on the cleaned dataset to confirm the effectiveness of the cleaning steps. Initially, the dataset contained 205,052 entries with missing values, duplicate entries, and anomalies in certain columns. After cleaning, the dataset was reduced to 154,475 entries, with all missing values in the `Review` and `Sentiment` columns addressed and duplicates removed. Text normalization was performed, converting text to lowercase and removing punctuation, resulting in a standardized and clean dataset.

The updated descriptive statistics reflect the cleaned data's characteristics more accurately. The `product_price` column now includes 479 unique prices, with the most frequent price being 1299, appearing 6,859 times. The `Rate` column shows a similar distribution but fewer anomalies. The `product_name` column contains 841 unique product names, with the most common product being "cello Pack of 18 Opalware Cello Dazzle Lush Fiesta Opalware Dinner Set, 18 Pieces Dinner Set" appearing 4,362 times. The `Sentiment` column shows a significant imbalance, with 122,813 positive entries, 23,353 negative entries, and 8,309 neutral entries.

Visualizations were employed to illustrate changes in data distributions and relationships after cleaning. Updated histograms for numerical features such as `product_price` and `Rate` were generated, showing a more consistent distribution after removing anomalies. Bar plots for categorical features, including `product_name` and `Sentiment`, provided a clearer visual representation of their frequency distribution post-cleaning. Scatter plots explored the relationship between `product_price` and `Rate`, confirming that the cleaning process had effectively addressed any initial inconsistencies and anomalies.

The cleaned dataset information showed a significant improvement in data quality, with no missing values in critical columns and a substantial reduction in duplicate entries. The `Processed_Review` column was added, reflecting the normalized and tokenized text, and it is ready for further analysis. This thorough cleaning ensured the dataset was robust and reliable for subsequent machine learning tasks, such as sentiment analysis and model training.

Model Implementation

For this study, four algorithms were chosen for comparison: Logistic Regression, SVC, Random Forest, and Gradient Boosting. Each algorithm was carefully selected based on its robustness and effectiveness in handling sentiment classification tasks. Hyperparameter tuning was conducted to optimize the performance of these algorithms. Techniques such as grid search and randomized search were employed to identify the best parameters for each model. For instance, the best parameters for Logistic Regression were found to be `{'C': 1, 'penalty': 'l1', 'solver': 'liblinear'}`, achieving a score of 0.8998. Similarly, SVC performed best with `{'C': 10, 'gamma': 0.1, 'kernel': 'rbf'}`, Random Forest with `{'n_estimators': 200, 'max_features': 'sqrt', 'max_depth': 8, 'criterion': 'entropy'}`, and Gradient Boosting with `{'learning_rate': 0.1, 'max_depth': 4, 'n_estimators': 300}`.

To evaluate the models effectively, the data was divided into training and testing sets. An 80-20 split was used, ensuring that 80% of the data was used for training and 20% for testing. Cross-validation techniques were employed to ensure robust model evaluation, which helped mitigate overfitting and provide a more accurate estimate of the model's performance on unseen data.

Each model was trained using the scikit-learn library, a popular machine learning toolkit in Python. The Logistic Regression model was trained and evaluated, achieving an accuracy of 0.8995, precision of 0.8773, recall of 0.8995, and an F1 score of 0.8736. The SVC model followed, with an accuracy of 0.8997, precision of 0.8619, recall of 0.8997, and an F1 score of 0.8738. While achieving a lower accuracy of 0.8011, the Random Forest model faced challenges in precision, particularly with the neutral sentiment class, which resulted in warnings about undefined metrics. Finally, the Gradient Boosting model achieved an accuracy of 0.8993, precision of 0.8512, recall of 0.8993, and an F1 score of 0.8734.

The confusion matrices and classification reports provided detailed insights into each model's performance. Logistic Regression and SVC showed strong performance across most metrics, while Random Forest struggled with precision in the neutral class. Gradient Boosting offered a balanced performance but highlighted the difficulty of accurately classifying neutral sentiments across all models. This thorough evaluation and comparison of algorithms provided a clear understanding of their strengths and limitations in sentiment classification tasks on Flipkart product reviews.

Model Evaluation

To evaluate the performance of the models, several metrics were employed, including accuracy, precision, recall, F1-score, and ROC-AUC. Accuracy measures the overall correctness of the model by calculating the ratio of correctly predicted instances to the total instances. Precision assesses the model’s ability to correctly identify positive instances from the total instances it predicted as positive, indicating the quality of positive predictions. Recall, or sensitivity, evaluates the model’s capability to identify all positive instances, reflecting how well the model captured positive cases. The F1-score provides a harmonic mean of precision and recall, offering a balanced measure that accounts for false positives and false negatives. Lastly, the ROC-AUC score quantifies the model's performance across all classification thresholds, providing a comprehensive measure of class separability.

The Logistic Regression model achieved an accuracy of 0.8995, precision of 0.8773, recall of 0.8995, an F1-score of 0.8736, and an ROC-AUC score of 0.9105. However, the confusion matrix showed that the model struggled to classify neutral sentiments, resulting in undefined metrics for precision due to the lack of positive samples in certain true classes. Similarly, the Gradient Boosting model attained an accuracy of 0.8993, precision of 0.8512, recall of 0.8993, an F1-score of 0.8735, and a ROC-AUC score of 0.9098. Still, it also faced challenges with the neutral sentiment class.

The evaluation criteria for comparing the algorithms included performance metrics and computational efficiency. Performance metrics provided a quantitative assessment of each model's effectiveness in classifying sentiments correctly, while computational efficiency involved evaluating the time taken for training and prediction. This dual approach ensured a thorough assessment of each model's strengths and weaknesses. For example, despite its high accuracy of 0.8997 and precision of 0.8619, the SVC model had a significantly longer training time compared to Logistic Regression, which had a much shorter training time of 5.24 seconds but similar accuracy and precision. The Random Forest model, with an accuracy of 0.7953 and precision of 0.6326, demonstrated the need for balanced consideration of both predictive performance and computational demands.

The evaluation results, sorted by accuracy, showed that SVC was the most accurate, followed closely by Logistic Regression and Gradient Boosting, with Random Forest performing less well in comparison. These insights guided the selection of the most suitable model for sentiment analysis tasks, balancing both performance and practical deployment considerations.

Result and Discussion

Result

The performance of each algorithm was thoroughly evaluated using several key metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. These metrics provide a comprehensive overview of how well each model classified the sentiment of product reviews. The evaluation results are presented in the following tables and visualizations to facilitate a clear comparison of the models. The table 1 below summarizes the performance metrics for each algorithm.

Table 1 **Performance Metrics for Each Algorithm**

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC	Training Time	Prediction Time
Logistic Regression	0.8995	0.8773	0.8995	0.8736	0.9105	5.24 s	0.002 s
SVC	0.8997	0.8619	0.8997	0.8738	N/A	190.34 s	4.40 s
Random Forest	0.7953	0.6326	0.7953	0.7047	0.9037	24.78 s	0.32 s
Gradient Boosting	0.8993	0.8512	0.8993	0.8735	0.9098	281.60 s	0.72 s

The Logistic Regression model achieved an accuracy of 0.8995, a precision of 0.8773, a recall of 0.8995, an F1-score of 0.8736, and an ROC-AUC score of 0.9105. However, the confusion matrix highlighted the model's difficulty in classifying neutral sentiments, as evidenced by the undefined precision metrics for this category. The SVC model exhibited a slightly higher accuracy of 0.8997 and a recall of 0.8997 but a lower precision of 0.8619 compared to Logistic Regression. Its F1-score was 0.8738, indicating a balanced performance. The confusion matrix showed similar challenges in accurately classifying neutral sentiments.

The Random Forest model, with an accuracy of 0.7953, faced significant issues in precision, particularly with the neutral class, resulting in numerous undefined metrics for precision. The F1-score for Random Forest was 0.7047, reflecting its lower overall performance. The Gradient Boosting model, achieving an accuracy of 0.8993, a precision of 0.8512, and a recall of 0.8993, demonstrated a robust performance similar to that of Logistic Regression and SVC. The F1-score was 0.8735, with a ROC-AUC score of 0.9098, indicating strong overall performance despite the challenges in classifying neutral sentiments.

When comparing the algorithms' performance, the SVC model emerged as the most accurate, closely followed by Logistic Regression and Gradient Boosting. The Random Forest model, although generally effective, underperformed in comparison, particularly in its precision and ability to handle the neutral sentiment class. Logistic Regression and Gradient Boosting demonstrated balanced performance across all metrics, with Logistic Regression slightly outperforming Gradient Boosting in precision and recall. While highly accurate, the SVC model required significantly more computational resources, as indicated by its longer training and prediction times. Figure 3 below show the model comparison based on accuracy.

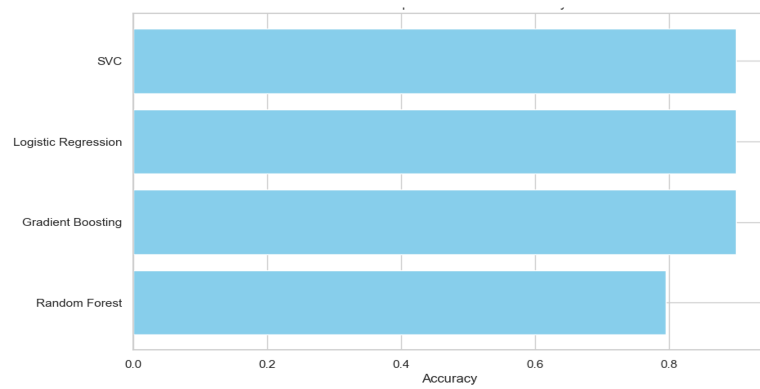

Figure 3 Model Comparison Based on Accuracy

Logistic Regression and Gradient Boosting are recommended for scenarios that balance performance and computational efficiency. SVC may be preferred for applications where the highest accuracy is critical and computational resources are less of a concern. Despite its lower performance in this study, Random Forest remains a viable option, particularly when its strengths in handling large and complex datasets are required. These detailed evaluations and comparisons provide valuable insights into the strengths and limitations of each model, guiding the selection of the most appropriate algorithm for sentiment classification tasks on Flipkart product reviews.

Discussion

The comparative analysis of sentiment classification techniques revealed that SVC, Logistic Regression, and Gradient Boosting consistently outperformed Random Forest regarding accuracy, precision, recall, and F1-score. The superior performance of SVC can be attributed to its ability to handle high-dimensional data and find a decision boundary that maximizes the margin between classes. Logistic Regression performed well due to its simplicity and effectiveness in binary classification tasks, making it a robust choice for sentiment analysis. Gradient Boosting's performance is largely due to its ability to build an ensemble of weak learners, improving the accuracy with each iteration. The lower performance of Random Forest might be linked to its inability to handle the imbalanced nature of the dataset effectively, particularly in distinguishing neutral sentiments, which resulted in lower precision and recall scores.

The findings of this study have significant practical implications for e-commerce platforms like Flipkart. Implementing the best-performing models, such as SVC and Logistic Regression, can greatly enhance the ability to analyze customer reviews accurately, providing deeper insights into customer sentiments. These insights can help businesses tailor their marketing strategies, improve product recommendations, and enhance customer service by understanding the underlying sentiment trends. For instance, accurately classified sentiments can aid in identifying common pain points and positive aspects of products, enabling businesses to address issues promptly and promote features that customers appreciate. With its strong performance, Gradient Boosting can also be used for complex sentiment analysis tasks where higher computational resources are available.

Despite the robust findings, this study has several limitations. The data quality plays a crucial role in model performance; any inaccuracies or inconsistencies in the dataset could affect the results. The study also relies on model assumptions, such as the linearity in Logistic Regression, which might only hold for some data types. Additionally, the scope of the analysis is limited to product reviews from Flipkart, which may not generalize to other e-commerce platforms or types of reviews. The imbalance in the sentiment classes, particularly the underrepresentation of neutral sentiments, also posed challenges in model training and evaluation.

Future research could focus on several areas to enhance sentiment analysis in e-commerce. Improving data quality by incorporating more comprehensive and diverse datasets from multiple e-commerce platforms could provide more generalized insights. Exploring advanced techniques like deep learning models, including LSTM and BERT, could improve sentiment classification accuracy by capturing more complex patterns in the data. Additionally, addressing class imbalance through techniques like Synthetic Minority Over-sampling Technique (SMOTE) or cost-sensitive learning could further enhance model performance. Finally, integrating additional features such as user metadata, product categories, and temporal aspects of reviews could provide richer context and improve the robustness of sentiment analysis models.

Discussion

This study compared various sentiment classification techniques applied to Flipkart product reviews. Logistic Regression, SVC, Random Forest, and Gradient Boosting were evaluated. SVC, Logistic Regression, and Gradient Boosting outperformed Random Forest in accuracy, precision, recall, and F1-score. SVC was the most accurate model, followed by Logistic Regression and Gradient Boosting. Limitations included dataset quality, model assumptions, platform-specific findings, and class imbalance. Future research could explore more diverse datasets, advanced techniques, and additional features to enhance sentiment analysis. Effective sentiment analysis can provide valuable insights for businesses, allowing them to tailor their strategies and improve customer service.

Conclusion

This study compared various sentiment classification techniques applied to

Flipkart product reviews. Logistic Regression, SVC, Random Forest, and Gradient Boosting were evaluated. SVC, Logistic Regression, and Gradient Boosting outperformed Random Forest in accuracy, precision, recall, and F1-score. SVC was the most accurate model, followed by Logistic Regression and Gradient Boosting. Limitations included dataset quality, model assumptions, platform-specific findings, and class imbalance. Future research could explore more diverse datasets, advanced techniques, and additional features to enhance sentiment analysis. Effective sentiment analysis can provide valuable insights for businesses, allowing them to tailor their strategies and improve customer service.

Declarations

Author Contributions

Conceptualization: H. and Q.S.; Methodology: Q.S.; Software: H.; Validation: H. and Q.S.; Formal Analysis: H. and Q.S.; Investigation: H.; Resources: Q.S.; Data Curation: Q.S.; Writing Original Draft Preparation: H. and Q.S.; Writing Review and Editing: Q.S. and H.; Visualization: H.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. Y. B. Yin, N. H. M. Saad, and Z. Yaacob, “Exploring Sentiment Analysis on E-Commerce Business: Lazada and Shopee,” Tem J., pp. 1508–1519, 2022, doi: 10.18421/tem114-11.

[2] S. Shayaa, A. Sulaiman, A. Z. Piprani, M. A. Al-Garadi, and M. Ashraf, “Big Data Social Media Analytics for Purchasing Behaviour,” Int. J. Eng. Technol., vol. 7, no. 4.36, p. 463, 2018, doi: 10.14419/ijet.v7i4.36.23917.

[3] M.-H. Yang, B. Jiang, Y. Wang, T. Hao, and Y. Liu, “News Text Mining-Based Business Sentiment Analysis and Its Significance in Economy,” Front. Psychol., vol. 13, 2022, doi: 10.3389/fpsyg.2022.918447.

[4] A. Bandhakavi, N. Wiratunga, S. Massie, and Rushi. Luhar, “Opinion Context

Extraction for Aspect Sentiment Analysis," Proc. Int. Aaai Conf. Web Soc. Media, vol. 12, no. 1, 2018, doi: 10.1609/icwsm.v12i1.15069.

[5] H. R. Sankar, V. Subramaniyaswamy, V. Vijayakumar, A. K. Sangaiah, L. Ravi, and A. Umamakeswari, "Intelligent Sentiment Analysis Approach Using Edge Computing-based Deep Learning Technique," Softw. Pract. Exp., vol. 50, no. 5, pp. 645–657, 2019, doi: 10.1002/spe.2687.

[6] P. Rokade and A. K. D, "Business Recommendation Based on Collaborative Filtering and Feature Engineering – Aproposed Approach," Int. J. Electr. Comput. Eng. Ijece, vol. 9, no. 4, p. 2614, 2019, doi: 10.11591/ijece.v9i4.pp2614-2619.

[7] J. Tao and X. Fang, "Toward Multi-Label Sentiment Analysis: A Transfer Learning Based Approach," J. Big Data, vol. 7, no. 1, 2020, doi: 10.1186/s40537-019-0278-0.

[8] M. Abdussalam, D. Richasdy, and M. A. Bijaksana, "BERT Implementation on News Sentiment Analysis and Analysis Benefits on Branding," J. Media Inform. Budidarma, vol. 6, no. 4, p. 2064, 2022, doi: 10.30865/mib.v6i4.4579.

[9] C. Steven and W. Wella, "The Right Sentiment Analysis Method of Indonesian Tourism in Social Media Twitter," Ijnmt Int. J. New Media Technol., vol. 7, no. 2, pp. 102–110, 2020, doi: 10.31937/ijnmt.v7i2.1732.

[10] M. N. Win, S. D. R. Ravana, and L. Shuib, "Sentiment Attribution Analysis With Hierarchical Classification and Automatic Aspect Categorization on Online User Reviews," Malays. J. Comput. Sci., vol. 35, no. 2, pp. 89–110, 2022, doi: 10.22452/mjcs.vol35no2.1.

[11] F. Alzami, E. D. Udayanti, D. P. Prabowo, and R. A. Megantara, "Document Preprocessing With TF-IDF to Improve the Polarity Classification Performance of Unstructured Sentiment Analysis," Kinet. Game Technol. Inf. Syst. Comput. Netw. Comput. Electron. Control, pp. 235–242, 2020, doi: 10.22219/kinetik.v5i3.1066.

[12] B. Berlilana, A. M. Wahid, D. Fortuna, A. N. A. Saputra, and G. Bagaskoro, "Exploring the Impact of Discount Strategies on Consumer Ratings: An Analytical Study of Amazon Product Reviews," J. Appl. Data Sci., vol. 5, no. 1, Art. no. 1, Jan. 2024, doi: 10.47738/jads.v5i1.163.

[13] R. Mu, Y. Zheng, and K. Zhang, "Research on Customer Satisfaction Based on Multidimensional Analysis," Int. J. Comput. Intell. Syst., 2021, doi: 10.2991/ijcis.d.210114.001.

[14] H. Chen, S. Li, P. Wu, Y. Nian, S. Li, and X. Huang, "Fine-Grained Sentiment Analysis of Chinese Reviews Using LSTM Network," J. Eng. Sci. Technol. Rev., vol. 11, no. 1, pp. 174–179, 2018, doi: 10.25103/jestr.111.21.

[15] T. Astuti and I. Pratika, "Product Review Sentiment Analysis by Artificial Neural Network Algorithm," Ijiis Int. J. Inform. Inf. Syst., vol. 2, no. 2, pp. 61–66, 2019, doi: 10.47738/ijiis.v2i2.15.

[16] B. Hemalatha*, "Impact of Customer Feedback System Using Machine Learning Algorithms for Sentiment Mining," Int. J. Innov. Technol. Explor. Eng., vol. 9, no. 4, pp. 1475–1483, 2020, doi: 10.35940/ijitee.d1537.029420.

[17] "Sentiment Mining of Product Opinion Data," Int. J. Innov. Technol. Explor. Eng., vol. 9, no. 3, pp. 1218–1222, 2020, doi: 10.35940/ijitee.c8641.019320.

[18] J. O. Haryanto and F. I. Chang, "Analysis of E-Commerce Providers' Role in

Solving the Issues of Retail E-Commerce Logistics in Jakarta (Study Case of Pt. Accommerce Solusi Lestari)," J. Manaj., vol. 22, no. 1, 2018, doi: 10.24912/jm.v22i1.310.

[19] S. McKee, S. Sands, J. Pallant, and E. Cohen, "The Evolving Direct-to-consumer Retail Model: A Review and Research Agenda," Int. J. Consum. Stud., vol. 47, no. 6, pp. 2816–2842, 2023, doi: 10.1111/ijcs.12972.

[20] D. D. Damayanti, N. Novitasari, E. B. Setyawan, and P. S. Muttaqin, "Intelligent Warehouse Picking Improvement Model for E-Logistics Warehouse Using Single Picker Routing Problem and Wave Picking," Joiv Int. J. Inform. Vis., vol. 6, no. 2, p. 418, 2022, doi: 10.30630/joiv.6.2.1006.

[21] L. L. Mao, "Retail Quality, Market Environment and Business Survival in the Retail Apocalypse: An Investigation of the Sporting Goods Retail Industry," Int. J. Sports Mark. Spons., vol. 23, no. 3, pp. 547–571, 2021, doi: 10.1108/ijsms-06-2021-0115.

[22] S. T. Peinkofer, T. L. Esper, R. J. Smith, and B. D. Williams, "Assessing the Impact of Price Promotions on Consumer Response to Online Stockouts," J. Bus. Logist., vol. 36, no. 3, pp. 260–272, 2015, doi: 10.1111/jbl.12095.

[23] J.-P. Rodrigue, "The Distribution Network of Amazon and the Footprint of Freight Digitalization," J. Transp. Geogr., vol. 88, p. 102825, 2020, doi: 10.1016/j.jtrangeo.2020.102825.

[24] A. Y. Liu and S.-M. Lin, "Service for Specialized and Digitalized: Consumer Switching Perceptions of E-commerce in Specialty Retail," Manag. Decis. Econ., vol. 44, no. 3, pp. 1521–1533, 2022, doi: 10.1002/mde.3762.

[25] S. Liu, "Imitation or Development? Physical Retailers in China," BCP Bus. Manag., vol. 42, pp. 94–107, 2023, doi: 10.54691/bcpbm.v42i.4567.

[26] Z. Riaz, "Editorial," Asian J. Manag. Cases, vol. 18, no. 2, pp. 107–108, 2021, doi: 10.1177/09728201211044857.

[27] "Changes in the Fdi Policy With Special Reference to E-Commerce in Indian Retail SECTOR AND IT's COMPLIANCE STRATEGIES As Adopted by Flipkart and Amazon: An Observational Review," Gap Gyan - Glob. J. Soc. Sci., vol. 3, no. 2, pp. 29–44, 2020, doi: 10.47968/gapgyan.32006.

[28] R. Regina, E. S. Rini, and B. K. F. Sembiring, "The Effect of Online Customer Review and Promotion Through E-Trust on the Purchase Decision of Bukalapakin Medan City," Int. J. Res. Rev., vol. 8, no. 8, pp. 236–243, 2021, doi: 10.52403/ijrr.20210833.