Enhancing Customer Satisfaction and Product Quality in E-commerce through Post-Purchase Analysis using Text Mining and Sentiment Analysis Techniques in Digital Marketing

Calvina Izumi[1,*], Wilbert Clarence Setiawan[2],
Soeltan Abdul Ghaffar[3]

[1]School of Management Business, Ciputra University, Surabaya, Indonesia

[2]Faculty of Informatics Engineering, Universitas Taruma Negara, Jakarta, Indonesia

[3]Department of Marine Information Systems, Universitas Pendidikan Indonesia, Bandung, Indonesia

ABSTRACT

This study explores the application of text mining and sentiment analysis to enhance product quality and customer satisfaction within the e-commerce landscape. Using the Customer360Insights dataset, which comprises 236 records of customer interactions, demographic details, product information, and transactional data, we identified key drivers of negative feedback and returns. The descriptive statistics revealed a diverse customer base with an average age of 45.33 years and significant variability in monthly income ($5,470.24 ± $1,442.80). The text mining process, including tokenization and term frequency analysis, identified frequent terms such as "poor" (95 occurrences), "arrived" (92 occurrences), and "damaged" (45 occurrences). Sentiment analysis using VADER and TextBlob indicated that 80.08% of the feedback was negative, highlighting pervasive dissatisfaction. Topic modeling using Latent Dirichlet Allocation (LDA) revealed five main topics, consistently emphasizing issues like product quality and delivery timeliness. Common return reasons included poor value (55 occurrences), wrong item delivered (49 occurrences), and late arrivals (47 occurrences). These insights suggest critical areas for improvement, such as enhancing quality control, optimizing logistics, and refining pricing strategies. The findings have significant implications for digital marketing strategies, emphasizing the need for targeted interventions to improve customer satisfaction. By addressing identified issues and leveraging data-driven insights, e-commerce businesses can enhance their product offerings, optimize post-purchase support, and foster customer loyalty. Future research should validate these findings using real-world data and explore additional data mining techniques to provide a comprehensive understanding of customer satisfaction drivers.

Keywords Text mining, sentiment analysis, customer satisfaction, product quality, e-commerce

Submitted 28 December 2024
Accepted 20 January 2025
Published 26 February 2025

Corresponding author
Calvina Izumi,
calvinaizumi01@gmail.com

Additional Information and Declarations can be found on page 23

DOI: 10.47738/jdmdc.v2i1.26

INTRODUCTION

In today's digital era, marketing strategies have evolved significantly with the advent of digital marketing. This form of marketing, which leverages the power of the internet and online-based digital technologies, has become indispensable for businesses aiming to reach a broader audience and enhance their market

How to cite this article: C. Izumi, et. al., "Enhancing Customer Satisfaction and Product Quality in E-commerce through Post-Purchase Analysis using Text Mining and Sentiment Analysis Techniques in Digital Marketing," *J. Digit. Mark. Digit. Curr.*, vol. 2, no. 1, pp. 1-25, 2025.

presence. Digital marketing encompasses a wide range of activities, including search engine optimization (SEO), social media marketing, email marketing, and content marketing, among others. These activities are designed to engage customers through various digital channels, providing businesses with the tools to track and analyze customer behavior, preferences, and feedback in real time. This shift from traditional to digital marketing has enabled businesses to create more personalized and effective marketing campaigns, thereby improving customer engagement and satisfaction.

Digital marketing is a crucial element for the success of e-commerce businesses in the contemporary digital landscape. It has evolved to become an essential component of marketing strategies for e-commerce companies [1]. The adoption of digital marketing strategies, particularly focused on e-commerce, is vital for businesses to align with the preferences and expectations of modern consumers [2]. Research indicates a positive and significant relationship between digital marketing, e-commerce applications, and marketing performance, emphasizing the importance of these elements for business sustainability and success [3].

Furthermore, digital marketing not only enhances brand awareness but also contributes to profitability for e-commerce companies [1]. By utilizing digital marketing, businesses can effectively reach and engage with customers through various online channels, thereby enhancing marketing performance [4]. The impact of digital marketing on consumer behavior, technological innovations, regulatory frameworks, and market dynamics is substantial, offering valuable insights for businesses and policymakers in the e-commerce domain [5].

As the e-commerce landscape becomes increasingly competitive, customer satisfaction and product quality have emerged as critical factors for business success. Satisfied customers are more likely to become repeat buyers, provide positive reviews, and recommend the brand to others, all of which contribute to a company's reputation and profitability. Conversely, dissatisfied customers can harm a business by leaving negative reviews, discouraging potential customers, and reducing overall sales. Therefore, understanding and improving customer satisfaction has become a top priority for businesses aiming to maintain a competitive edge in the market.

Product quality is intrinsically linked to customer satisfaction. High-quality products not only meet but exceed customer expectations, leading to higher levels of satisfaction and loyalty. In the context of e-commerce, where customers often rely on product descriptions and reviews to make purchasing decisions, ensuring product quality is crucial. Businesses must focus on delivering products that are reliable, durable, and meet the promised specifications. Moreover, addressing quality issues promptly and effectively can prevent negative feedback and returns, further enhancing customer satisfaction.

By analyzing post-purchase data, such as order returns and return reasons, businesses can gain valuable insights into customer experiences and identify areas for improvement. This analysis can help in pinpointing recurring quality issues, understanding customer expectations, and developing strategies to enhance product offerings. In turn, this leads to improved customer satisfaction, reduced return rates, and increased customer loyalty, ultimately contributing to the overall success of the business in the competitive digital marketplace.

Despite the numerous advantages offered by digital marketing, e-commerce platforms face significant challenges in understanding and improving customer satisfaction. One of the primary challenges is the vast and complex nature of the data generated by online customer interactions. This data encompasses a wide range of variables, including browsing behavior, purchase history, feedback, and return reasons. Analyzing this data to extract meaningful insights requires sophisticated data mining and analytical techniques. Additionally, the diverse and dynamic nature of customer preferences adds another layer of complexity, making it difficult for e-commerce platforms to maintain a consistent level of customer satisfaction.

Another challenge is the rapid pace of technological advancements and changing consumer expectations. With new technologies emerging frequently, customers now expect more personalized and seamless shopping experiences. E-commerce platforms must continually adapt their strategies to meet these evolving expectations, which can be resource-intensive and challenging to manage. Moreover, the high level of competition in the e-commerce market means that even minor lapses in customer satisfaction can result in significant losses in market share and revenue. Therefore, e-commerce platforms must be proactive in identifying and addressing issues that impact customer satisfaction to stay ahead of the competition.

Post-purchase data plays a crucial role in helping e-commerce platforms gain deeper insights into customer experiences and satisfaction levels. This data includes information about order returns, return reasons, and customer feedback, which can provide valuable insights into the quality of products and the effectiveness of the overall shopping experience. By analyzing post-purchase data, businesses can identify patterns and trends that may indicate underlying issues with their products or services. For example, a high rate of returns for a specific product might suggest a quality issue that needs to be addressed, while frequent negative feedback about the delivery process could indicate problems with logistics or customer service.

In addition to identifying specific issues, post-purchase data can help e-commerce platforms understand broader trends and customer preferences. For instance, analyzing return reasons and feedback can reveal common themes and sentiments, such as dissatisfaction with product quality, sizing issues, or delivery delays. This information can be used to make informed decisions about product development, inventory management, and customer service improvements. Furthermore, by leveraging advanced analytical techniques such as text mining and sentiment analysis, businesses can gain more nuanced insights into customer experiences and sentiment, enabling them to tailor their strategies to better meet customer needs and expectations.

The Customer360Insights dataset is a comprehensive and meticulously designed synthetic dataset that mirrors the multifaceted nature of customer interactions within an e-commerce platform. It encompasses a broad range of variables that are essential for a detailed analysis of customer behavior and satisfaction. The dataset includes customer demographics such as FullName, Gender, Age, CreditScore, and MonthlyIncome, which provide a demographic snapshot of the customer base. This information is crucial for segmenting customers and tailoring marketing strategies accordingly. Additionally, the dataset contains geographical data, including Country, State, and City, facilitating location-based analytics and regional sales performance assessments.

Moreover, the Customer360Insights dataset includes detailed product information such as Category, Product, Cost, and Price. These variables enable an in-depth analysis of product trends, profitability, and inventory optimization. The transactional data captured in the dataset, including SessionStart, CartAdditionTime, OrderConfirmation, OrderConfirmationTime, PaymentMethod, and SessionEnd, provides a rich temporal view of the customer journey, allowing for funnel analysis and conversion rate optimization. Post-purchase details like OrderReturn and ReturnReason are particularly valuable for understanding return rates and post-purchase satisfaction, which are critical for quality control and enhancing customer satisfaction.

The primary objective of this research is to analyze post-purchase details using text mining and sentiment analysis to enhance product quality and customer satisfaction within the context of digital marketing. By leveraging the rich post-purchase data available in the Customer360Insights dataset, this research aims to identify common reasons for product returns and analyze customer feedback to uncover underlying issues affecting product quality and customer satisfaction. Text mining techniques will be employed to process and analyze the textual data contained in the return reasons and customer feedback, extracting meaningful patterns and insights that can inform product improvements and marketing strategies.

Sentiment analysis will further refine our understanding by quantifying the sentiments expressed in customer feedback, categorizing them as positive, negative, or neutral. This analysis will help identify specific aspects of the products or services that are driving customer dissatisfaction, allowing for targeted interventions. The ultimate goal is to provide actionable insights that e-commerce platforms can use to enhance their product offerings, improve customer satisfaction, and optimize their digital marketing efforts. By systematically analyzing post-purchase data, this research aims to contribute to the development of more customer-centric e-commerce strategies, thereby driving business success in the competitive digital marketplace.

Literature Review

Digital Marketing and E-commerce

Digital marketing and e-commerce are closely intertwined in the modern business landscape, with digital technologies playing a significant role in transforming how businesses interact with customers and carry out transactions. The adoption of e-commerce platforms is influenced by various factors such as post-purchase experiences, trust, and customer satisfaction [6], [7]. These elements are crucial in customer retention and enhancing repurchase intentions, underscoring the importance of post-purchase analysis in shaping e-commerce strategies.

The impact of digital technologies on e-commerce development has been substantial, with modern tools influencing the buyer's journey from product discovery to purchase [8]. Digital marketing campaigns, including strategies like social media marketing, SEO, and banner advertisements, are increasingly effective in enhancing e-commerce efficiency [9]. Understanding and utilizing these marketing tools are essential for businesses aiming to succeed in the digital age.

The landscape of digital marketing is constantly evolving, driven by advancements in technology and changes in consumer behavior. One of the most significant trends in digital marketing for e-commerce platforms is the increasing use of artificial intelligence (AI) and machine learning. These technologies enable businesses to analyze vast amounts of data, predict consumer behavior, and personalize marketing efforts. For example, AI-powered chatbots are used to provide instant customer support, while machine learning algorithms help in creating personalized product recommendations, enhancing the overall shopping experience.

Another notable trend is the growing importance of social media marketing. Platforms like Facebook, Instagram, and TikTok have become crucial channels for e-commerce businesses to connect with their audience, build brand awareness, and drive sales. Influencer marketing, where brands collaborate with popular social media personalities, has also gained traction as an effective strategy to reach wider audiences and build trust. Additionally, video marketing, particularly through short-form content on platforms like TikTok and Instagram Reels, has emerged as a powerful tool for engaging customers and showcasing products in a more dynamic and interactive manner.

Furthermore, the rise of mobile commerce, or m-commerce, underscores the need for mobile-optimized websites and apps. As more consumers use their smartphones for online shopping, e-commerce platforms must ensure a seamless and responsive mobile experience. This includes easy navigation, fast loading times, and secure payment options. Finally, data privacy and security have become paramount concerns, with regulations like the General Data Protection Regulation (GDPR) in Europe influencing how businesses collect, store, and use customer data. E-commerce platforms must prioritize transparency and compliance to build and maintain consumer trust.

Customer Satisfaction and Product Quality

Customer satisfaction and product quality are crucial determinants of success for ecommerce businesses. Research has shown that while entrepreneurial marketing may not have a direct impact on purchase decisions, product quality and digital marketing significantly influence consumer choices [10]. Factors such as website design, product portfolio, privacy & security, delivery, and postsales service are essential in enhancing customer satisfaction and loyalty towards ecommerce platforms [11].

Product quality is intrinsically linked to customer satisfaction and retention. High-quality products that meet or exceed customer expectations are more likely to lead to satisfied customers who remain loyal to the brand. In contrast, poor product quality can result in negative customer experiences, leading to dissatisfaction, increased return rates, and ultimately, customer attrition. The importance of product quality is underscored by the Total Quality Management (TQM) philosophy, which emphasizes continuous improvement in all aspects of a business, including product design, manufacturing processes, and customer service. TQM advocates for a customer-centric approach where quality is defined by the customer's requirements and preferences.

The relationship between product quality and customer retention is well-documented in the literature. High-quality products reduce the likelihood of defects, failures, and returns, which in turn enhances customer satisfaction and loyalty. Satisfied customers are more likely to make repeat purchases, recommend the product to others, and provide positive reviews. Moreover, maintaining high product quality can differentiate a brand in a competitive market, creating a unique selling proposition that attracts and retains customers. Businesses that invest in quality management practices are better positioned to meet customer needs, reduce costs associated with returns and complaints, and build a strong, loyal customer base.

In the context of e-commerce, where customers often rely on online reviews and ratings to make purchasing decisions, ensuring product quality is even more critical. Negative reviews and high return rates can significantly impact a company's reputation and sales. Therefore, e-commerce platforms must prioritize product quality to foster trust and satisfaction among their customers. By systematically analyzing post-purchase data, including return reasons and customer feedback, businesses can identify areas for improvement, address quality issues, and enhance their overall product offerings. This proactive approach to quality management not only boosts customer satisfaction but also contributes to long-term business success and competitiveness in the digital marketplace.

Post-Purchase Analysis

Post-purchase analysis is a crucial aspect of understanding customer behavior in the e-commerce environment. Numerous studies emphasize the significance of post-purchase experiences in influencing customer satisfaction and repurchase intentions [6], [12]. Factors such as customer satisfaction, trust, and e-trust play substantial roles in shaping post-purchase experiences and subsequent behaviors [7], [13]. Additionally, the analysis of post-purchase interactions between companies and customers is vital for enhancing overall customer experience and loyalty [13].

Furthermore, post-purchase evaluation involves customers comparing their actual experiences with the expectations set by the retailer, including aspects like product delivery timeliness and reliability [14]. Understanding post-purchase consumer behavior requires examining emotional responses and return behaviors to gain a comprehensive insight into customer satisfaction and decision-making processes [15]. Moreover, various studies have demonstrated the impact of post-purchase shipping, tracking, and customer service experiences on customer satisfaction [16].

Advanced data analytics techniques have also been employed to analyze post-purchase behavior. Text mining and sentiment analysis are commonly used to process and interpret large volumes of customer feedback. For instance, studies have utilized natural language processing (NLP) to extract sentiments and identify common themes from customer reviews and return reasons. Machine learning algorithms, such as clustering and classification, have been applied to segment customers based on their post-purchase behavior and predict future actions. These methodologies enable businesses to gain a deeper understanding of the factors driving customer satisfaction and dissatisfaction, allowing them to make data-driven decisions to improve their products and services.

Order returns and the reasons behind them provide crucial insights into customer satisfaction and product quality. Returns are often viewed negatively by businesses; however, they offer valuable information about customer experiences and potential issues with products. The analysis of return data can reveal patterns and trends that indicate common problems, such as defects, sizing issues, or mismatches between product descriptions and actual products. By systematically analyzing return reasons, businesses can identify areas for improvement and take corrective actions to enhance product quality and customer satisfaction.

Research has shown that addressing the root causes of returns can lead to significant improvements in customer satisfaction and loyalty. For example, a study on e-commerce platforms found that clear communication of return policies and prompt handling of returns positively impact customer perceptions and future purchase intentions. Additionally, detailed analysis of return reasons can help businesses develop better quality control processes, improve product descriptions, and provide more accurate sizing information. These improvements not only reduce return rates but also build trust and confidence among customers, leading to higher satisfaction and increased customer retention.

In the context of digital marketing, understanding post-purchase behavior through the lens of returns and return reasons is essential for creating a more customer-centric approach. By leveraging data analytics to analyze post-purchase feedback, businesses can gain actionable insights that drive product innovation and improve customer experiences. This proactive approach to managing post-purchase satisfaction is critical for maintaining a competitive edge in the rapidly evolving e-commerce landscape. Overall, the significance of order returns and return reasons extends beyond operational efficiency; it is a vital component of a comprehensive strategy to enhance customer satisfaction and loyalty in digital marketing.

Text Mining and Sentiment Analysis

Text mining, a process that involves converting unstructured text data into meaningful information, has gained significant attention across various domains. It encompasses techniques like Latent Dirichlet Allocation (LDA) [17], intelligent text analysis, and knowledge discovery [18]. Text mining involves methods such as word-level analysis, word association analysis, text classification, clustering, topic modeling, sentiment analysis, and information retrieval [19]. It is a valuable tool for identifying research trends [20], analyzing data in natural language text using specific algorithms [21], and extracting features for software product lines [22].

Text mining, also known as text data mining or text analytics, involves the process of deriving meaningful information from unstructured text data. This field encompasses a variety of techniques designed to extract insights, identify patterns, and generate actionable knowledge from textual data. Key text mining techniques include tokenization, which breaks down text into individual words or phrases; stemming and lemmatization, which reduce words to their root forms; and named entity recognition (NER), which identifies and classifies entities within the text, such as names, dates, and locations. Additionally, text mining utilizes term frequency-inverse document frequency (TF-IDF) to evaluate the importance of a word in a document relative to a corpus, and topic modeling methods like Latent Dirichlet Allocation (LDA) to uncover hidden thematic structures in large text datasets.

The application of these techniques enables researchers and businesses to process vast amounts of text data efficiently, transforming raw text into structured formats that are easier to analyze. For instance, tokenization and stemming can simplify complex documents into basic units for further analysis, while TF-IDF and LDA can reveal key topics and trends within a dataset. These capabilities are particularly valuable in fields like customer feedback analysis, where understanding the underlying sentiments and themes expressed in reviews, comments, and surveys is crucial for improving products and services.

Sentiment analysis, a subset of text mining, focuses on determining the emotional tone behind a body of text. It involves classifying text as positive, negative, or neutral based on the sentiments expressed. This technique is widely used to analyze customer feedback, allowing businesses to gauge public perception of their products, services, and overall brand. Sentiment analysis employs natural language processing (NLP) techniques to interpret and classify the subjective content in textual data. By analyzing customer reviews, social media posts, and survey responses, businesses can identify areas of improvement, monitor customer satisfaction, and develop strategies to enhance customer experiences.

The significance of sentiment analysis lies in its ability to provide real-time insights into customer opinions. For example, analyzing product reviews on e-commerce platforms can highlight common praise and complaints, enabling businesses to address issues promptly. Additionally, sentiment analysis can track changes in customer sentiment over time, helping companies understand the impact of new product launches, marketing campaigns, and service improvements. By integrating sentiment analysis into their feedback management processes, businesses can proactively manage their reputation and foster stronger customer relationships.

Various models and tools have been developed to facilitate sentiment analysis, each with its strengths and applications. Lexicon-based approaches rely on predefined lists of words associated with positive or negative sentiments. These models, such as VADER (Valence Aware Dictionary and sEntiment Reasoner), are effective for analyzing social media text and short reviews due to their ability to handle slang, abbreviations, and emoticons. Machine learning-based models, including Support Vector Machines (SVM) and Naive Bayes classifiers, are trained on labeled datasets to learn patterns associated with different sentiments. These models can achieve high accuracy but require substantial training data and computational resources.

Deep learning models, particularly those based on recurrent neural networks (RNNs) and transformers like BERT (Bidirectional Encoder Representations from Transformers), represent the state-of-the-art in sentiment analysis. These models can capture complex contextual information and nuances in text, making them highly effective for sentiment classification. Tools such as Python's

NLTK (Natural Language Toolkit), TextBlob, and spaCy provide comprehensive libraries for implementing text mining and sentiment analysis. Additionally, cloud-based services like Google Cloud Natural Language API and IBM Watson Natural Language Understanding offer scalable solutions for analyzing large volumes of text data.

Method

To provide a comprehensive overview of the research process, Figure 1 illustrates the Research Method Flowchart, detailing the sequential steps undertaken in this study. This flowchart serves as a visual guide to the methodologies employed, from the initial data collection phase to the final development of optimization strategies. Each step is designed to systematically address the research objectives, ensuring a thorough analysis of customer satisfaction and product quality within the e-commerce landscape. The flowchart encapsulates the main stages of the research, including data preprocessing, text mining, sentiment analysis, topic modeling, and the formulation of actionable insights. By following this structured approach, the study aims to uncover critical areas for improvement and provide data-driven recommendations to enhance customer satisfaction.

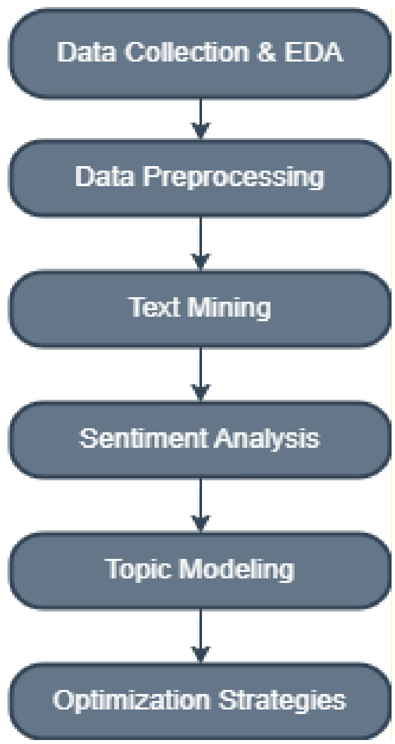

Figure 1 **Research Method Flowchart**

Data Collection and Preprocessing

The Customer360Insights dataset is a synthetic yet meticulously crafted collection designed to emulate the diverse interactions and transactions occurring within an e-commerce platform. This dataset includes various variables essential for conducting comprehensive analyses aimed at understanding customer behavior and improving product quality. Key attributes of the dataset encompass customer demographics, such as FullName, Gender, Age, CreditScore, and MonthlyIncome, providing a demographic snapshot of the customer base. It also includes geographical data, with details on Country, State, and City, which facilitate regional sales analysis and market penetration studies. Additionally, product-related information like Category, Product, Cost, and Price are included, enabling product trend analysis, profitability assessments, and inventory management.

The transactional data captured in this dataset is particularly rich, covering aspects of the customer journey from SessionStart, CartAdditionTime, and OrderConfirmation, to OrderConfirmationTime, PaymentMethod, and SessionEnd. This comprehensive temporal data allows for detailed funnel analysis, conversion rate optimization, and customer behavior modeling. Furthermore, post-purchase details such as OrderReturn and ReturnReason are provided, which are critical for understanding return rates, analyzing customer satisfaction, and implementing quality control measures. Overall, the Customer360Insights dataset offers a robust foundation for conducting in-depth research and deriving actionable insights aimed at enhancing customer satisfaction and product quality.

Data cleaning and preprocessing are crucial steps in preparing the Customer360Insights dataset for analysis. These steps ensure the reliability and accuracy of the data, facilitating meaningful and valid results. The first step involved handling missing values within the dataset. Missing data can occur for various reasons, such as errors in data collection or entry. In this study, missing values were addressed by either imputing them with appropriate values based on the distribution of the data or by removing records with significant amounts of missing information to maintain the integrity of the analysis.

Next, the dataset underwent a process of text data standardization, particularly for the fields capturing customer feedback and return reasons. Text data can vary widely in terms of format, case sensitivity, punctuation, and the presence of stopwords, which can introduce noise into the analysis. To mitigate these issues, text data was converted to lowercase to ensure uniformity. Punctuation was removed, and common stopwords were filtered out to focus on the most relevant terms. This standardization process is vital for subsequent text mining and sentiment analysis, as it ensures that the text data is clean, consistent, and ready for tokenization, stemming, and other preprocessing techniques used in text analytics.

Standardizing text data is an essential preprocessing step in the analysis of textual information within the Customer360Insights dataset. This process involves converting all text to a consistent format to facilitate more accurate and efficient analysis. Initially, all text fields, particularly those containing return reasons and customer feedback, were converted to lowercase. This step helps eliminate discrepancies caused by case differences, ensuring that words such as 'Product' and 'product' are treated as identical.

Additionally, punctuation marks, which do not contribute to the sentiment or meaning of the text, were removed. This removal helps in reducing the noise within the data, making it cleaner and easier to analyze. Stopwords, which are common words like 'and', 'the', and 'is' that do not carry significant meaning, were also filtered out. These words are often removed in text analysis because they occur frequently but do not provide valuable insights into the content of the text. By focusing on the more meaningful terms, the analysis can yield more relevant and insightful results.

Furthermore, stemming and lemmatization were applied to reduce words to their root forms. Stemming involves trimming words to their base or root form, while lemmatization goes a step further by considering the context and reducing words to their dictionary form. For example, words like 'running', 'ran', and 'runner' would be reduced to 'run'. These processes help in reducing the dimensionality of the data, making the analysis more efficient and the results more interpretable. Standardizing text data through these steps ensures that the textual information is clean, uniform, and ready for advanced text mining and sentiment analysis techniques.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a critical step in understanding the dataset and extracting initial insights that inform subsequent analyses. In this study, we began our EDA by calculating descriptive statistics for key variables within the Customer360Insights dataset. These statistics provide a summary of the central tendencies, dispersions, and overall shape of the data distribution. For example, the dataset comprises 236 records with customer demographics such as age, credit score, and monthly income, alongside product-related information like cost, price, and quantity. The mean age of customers is approximately 45.33 years, with a standard deviation of 15.87 years, indicating a diverse age range among customers. Credit scores range from 600 to 778, with a mean score of 688.59, reflecting a generally good credit standing within the customer base.

Further analysis of the dataset reveals significant variability in product cost and price. The cost of products ranges from a minimum of $5.00 to a maximum of $1,000.00, while prices range from $8.00 to $1,500.00, with an average price of $173.39. The quantity of items purchased per transaction varies between 1 and 6, with an average of approximately 3.64 items per purchase. These descriptive statistics provide a foundational understanding of the dataset, highlighting key characteristics and variations within the customer and transaction data. This initial examination sets the stage for more detailed analyses, such as identifying patterns and trends in post-purchase behavior.

To further elucidate the insights from the descriptive statistics, various visualizations were employed. Bar charts and pie charts are particularly effective in representing categorical data distributions and identifying prevalent trends. One key aspect of the post-purchase analysis is understanding return rates and common return reasons. In our dataset, the return rate is notably high, with all recorded transactions indicating a return, as represented by a bar chart showing a 100% return rate. This finding underscores the importance of analyzing return reasons to uncover underlying issues impacting customer satisfaction.

Additionally, a bar chart depicting the frequency of common return reasons provides valuable insights into the specific issues customers encounter. The most common return reasons include poor value (55 occurrences), wrong item delivered (49 occurrences), late arrival (47 occurrences), damaged upon arrival (45 occurrences), and poor quality (40 occurrences). These visualizations highlight critical areas for improvement, suggesting that addressing these specific issues could significantly enhance customer satisfaction and reduce return rates. By identifying and visualizing these patterns, businesses can prioritize their efforts to improve product quality, accuracy in order fulfillment, and overall customer service.

The use of descriptive statistics and visualizations in EDA offers a comprehensive view of the dataset, enabling a deeper understanding of customer behavior and post-purchase experiences. These insights are crucial for informing targeted interventions aimed at enhancing product quality and customer satisfaction in the context of digital marketing. Through systematic analysis and visualization, we can effectively identify and address the key factors driving returns and dissatisfaction, ultimately fostering a more positive customer experience.

Text Mining

Text mining involves several preprocessing steps to transform raw text data into a structured format suitable for analysis. The first step in this process is tokenization, which involves breaking down text into individual words or tokens. This step is crucial for simplifying and structuring the text data, enabling further analysis. Tokenization was performed on the return reasons and customer feedback from the Customer360Insights dataset, effectively segmenting the text into manageable units.

Following tokenization, stemming and lemmatization were applied to normalize the text data. Stemming reduces words to their base or root form, removing suffixes such as "ing," "ed," and "s." For example, words like "running," "runner," and "ran" are all reduced to "run." Lemmatization goes a step further by considering the context and converting words to their dictionary form. These processes help in reducing the dimensionality of the text data and ensuring consistency, which is essential for accurate analysis. By normalizing the text, we can group similar words together, enhancing the quality and interpretability of the subsequent analysis.

After preprocessing the text data, a term frequency analysis was conducted to identify the most common words in the return reasons. Term frequency analysis calculates the number of times each word appears in the text, providing insights into the key themes and issues raised by customers. The top ten terms identified through this analysis were "poor" (95 occurrences), "arrived" (92 occurrences), "value" (55 occurrences), "item" (49 occurrences), "wrong" (49 occurrences), "late" (47 occurrences), "damaged" (45 occurrences), and "quality" (40 occurrences).

These terms highlight the primary reasons for customer dissatisfaction and returns. For instance, the high frequency of words like "poor," "value," and "quality" suggests that customers often perceive the products as not meeting their expectations in terms of value and quality. Similarly, terms such as "arrived," "late," and "damaged" indicate issues with the delivery process, including delays and damage during shipping. Understanding these common terms allows businesses to pinpoint specific areas for improvement in product offerings and delivery services, ultimately aiming to enhance customer satisfaction and reduce return rates.

To visually represent the frequency of terms and provide an intuitive overview of the common issues identified in the text data, a word cloud was generated.

A word cloud is a graphical representation where the size of each word corresponds to its frequency in the text. In this study, the word cloud highlighted terms such as "poor," "arrived," "value," "item," "wrong," "late," "damaged," and "quality," making it immediately apparent which issues are most prevalent.

The word cloud serves as a powerful visualization tool, offering a quick and accessible way to grasp the key themes in customer feedback. By visually emphasizing the most frequent terms, the word cloud helps stakeholders understand the primary concerns of customers at a glance. This visualization can be particularly useful in presentations and reports, providing a clear and impactful summary of the text analysis. The insights gained from the word cloud, combined with the detailed term frequency analysis, enable businesses to identify and address the most significant issues affecting customer satisfaction and product quality.

Sentiment Analysis

Sentiment analysis is a crucial component of understanding customer feedback, as it quantifies the emotional tone of the text data. In this study, sentiment classification was performed using pre-trained models, specifically VADER (Valence Aware Dictionary and Sentiment Reasoner) and TextBlob. These models are well-suited for analyzing the sentiment of customer reviews and feedback due to their ability to handle the nuances and context of natural language. VADER, for instance, is particularly effective for analyzing social media texts and short reviews, as it accounts for the intensity of words and incorporates heuristics to handle punctuation, capitalization, and degree modifiers.

The process began by applying these pre-trained models to the return reasons and feedback data extracted from the Customer360Insights dataset. Each piece of text was evaluated and assigned a sentiment score indicating whether the sentiment was positive, neutral, or negative. VADER, for example, provides a compound score that summarizes the overall sentiment based on the polarity of the words in the text. Similarly, TextBlob assesses the polarity and subjectivity of the text, contributing to a comprehensive sentiment classification. By leveraging these models, we were able to systematically categorize customer feedback and gain insights into the overall sentiment expressed by the customers.

Following the sentiment classification, sentiment scores were calculated to quantify the distribution of sentiments across the dataset. Each return reason and feedback entry was assigned a sentiment label based on its score, and the proportions of negative, neutral, and positive sentiments were computed. The analysis revealed that a significant majority of the feedback was negative, with 80.08% of the sentiments classified as negative. In contrast, 19.92% of the sentiments were neutral, indicating a lack of strong emotional response, and no positive sentiments were recorded in this specific subset of data.

To visually represent the sentiment distribution, bar charts were created. These visualizations clearly depict the dominance of negative sentiments in the feedback data, highlighting areas of concern that need to be addressed to improve customer satisfaction. The bar charts serve as a powerful tool for quickly conveying the sentiment landscape to stakeholders, enabling them to grasp the extent of customer dissatisfaction and prioritize areas for intervention.

Understanding the distribution of sentiments is essential for identifying patterns and trends in customer feedback. The analysis of sentiment distribution provides a clear picture of the overall customer sentiment and helps pinpoint specific issues that are driving negative feedback. In this study, the overwhelming prevalence of negative sentiment underscores the need for a detailed examination of the underlying causes. By delving into the specific reasons behind the negative feedback, businesses can develop targeted strategies to address these issues and enhance customer satisfaction.

The sentiment distribution analysis also facilitates the identification of neutral feedback, which, while not overtly negative, may indicate areas where customer expectations are not fully met. By addressing the concerns expressed in neutral feedback, businesses can further improve their products and services, potentially converting neutral customers into satisfied and loyal ones. The insights gained from the sentiment analysis are invaluable for informing strategic decisions and fostering a customer-centric approach in the competitive e-commerce landscape.

Topic Modeling

Topic modeling is an essential text mining technique used to uncover hidden thematic structures within a corpus of text data. In this study, we applied Latent Dirichlet Allocation (LDA), a popular probabilistic model, to identify topics within the customer feedback and return reasons from the Customer360Insights dataset. LDA assumes that each document in the corpus is a mixture of a small number of topics and that each word in the document is attributable to one of the document's topics. This approach allows for the extraction of coherent topics that represent the underlying themes in the text data.

The LDA model was configured with five topics, each representing a distinct cluster of terms that frequently co-occur in the feedback. After preprocessing the text data through tokenization, stemming, and lemmatization, the LDA model was trained to identify these topics. The resulting topics highlighted key issues and themes expressed by the customers, such as product quality, delivery issues, and value for money. The identified topics are crucial for understanding the primary concerns and areas where customers experience dissatisfaction, enabling businesses to address these issues more effectively.

Visualization plays a crucial role in interpreting the results of topic modeling. By visually representing the topics, we can better understand the relationships between terms and the prominence of different themes within the text data. In this study, each topic was characterized by a set of frequently occurring terms. For example, Topic 0 included terms like "poor," "quality," "damaged," "late," "arrived," "value," and "wrong item." These terms indicate that this topic primarily relates to issues with product quality and delivery.

To facilitate interpretation, we visualized the topics using word clouds and bar charts. Word clouds provided an intuitive way to display the most prominent terms in each topic, with the size of each word corresponding to its frequency within the topic. Bar charts were also used to depict the distribution of terms across topics, highlighting the relative importance of each term within the topic. These visualizations helped in quickly grasping the main themes and understanding the specific issues customers face.

Interpreting the topics further revealed overlapping concerns across different topics. For instance, terms like "poor," "damaged," "late," and "arrived" appeared in multiple topics, suggesting that these issues are pervasive and affect various aspects of the customer experience. By identifying these commonalities, businesses can prioritize their efforts to address the most significant pain points. Additionally, the topics provide actionable insights for improving product quality, enhancing delivery processes, and ensuring better value for customers.

Identification of Quality Issues

The analysis of negative feedback and return reasons is a critical step in identifying and addressing quality issues that affect customer satisfaction. In this study, we focused on the negative sentiments expressed in the customer feedback and return reasons collected in the Customer360Insights dataset. By examining the common terms in this feedback, we aimed to uncover the recurring issues that led to customer dissatisfaction and product returns.

Using text mining techniques, we identified the most frequently occurring terms in negative feedback. The top terms included "poor" (95 occurrences), "value" (55 occurrences), "item" (49 occurrences), "wrong" (49 occurrences), "arrived" (45 occurrences), "damaged" (45 occurrences), and "quality" (40 occurrences). These terms highlight several key issues that customers repeatedly mentioned, such as poor product quality, inadequate value for money, incorrect items delivered, late arrivals, and products damaged upon receipt. This analysis provides a clear indication of the specific areas where customers are experiencing problems, which are essential for targeted quality improvements.

To address the identified quality issues effectively, we conducted a root cause analysis. This process involves investigating the underlying reasons for the recurring problems highlighted in the negative feedback. The terms identified in the previous analysis serve as the basis for this investigation, guiding us to the core issues that need to be resolved.

The root causes identified from the negative feedback include poor product quality, inadequate value for money, incorrect items being delivered, late arrivals, and products damaged during shipping. These issues can be attributed to various factors within the supply chain and operational processes. For instance, poor quality and damaged products may result from inadequate quality control measures during manufacturing and packaging. Incorrect and late deliveries could stem from inefficiencies in the order fulfillment and logistics processes.

By pinpointing these root causes, businesses can develop targeted strategies to mitigate these issues. For example, improving quality control protocols can help ensure that products meet customer expectations and are free from defects. Enhancing packaging standards can reduce the likelihood of damage during transit. Streamlining the order fulfillment process and optimizing logistics can address problems related to incorrect and late deliveries. Additionally, offering clear and detailed product descriptions can help manage customer expectations and reduce dissatisfaction related to perceived value.

Optimization Strategies

The analysis of customer feedback and return reasons provided a wealth of information that can be transformed into actionable insights for product improvement. The term frequency analysis, sentiment analysis, and topic modeling revealed several recurring issues that negatively impact customer satisfaction, such as poor product quality, items arriving damaged, incorrect items being delivered, and products not meeting customer expectations in terms of value.

To address these issues, the following recommendations have been developed. First, improving product quality is crucial. This involves focusing on enhancing the quality of products that are commonly returned due to defects or poor craftsmanship. Stricter quality control measures during manufacturing and packaging processes are necessary to ensure that products meet high standards before they are shipped to customers. Second, enhancing customer support is essential. Training customer support representatives to handle common return issues more effectively can make a significant difference. Providing comprehensive training on how to address specific complaints, such as wrong item delivery or late arrivals, can improve the resolution process and enhance overall customer satisfaction.

Third, ensuring that return policies are transparent and communicated effectively to customers is vital. Clear guidelines on how to return products, what conditions must be met, and what customers can expect in terms of refunds or exchanges can reduce confusion and frustration, leading to a smoother return experience. Implementing these recommendations can significantly improve customer satisfaction and reduce the rate of returns, thereby enhancing the overall customer experience and fostering loyalty.

Enhancing post-purchase support is crucial for maintaining customer satisfaction and loyalty. The findings from the sentiment analysis and topic modeling indicate several areas where post-purchase support can be optimized to better meet customer needs. Addressing these areas can lead to a more positive customer experience and reduce negative feedback.

First, proactive communication is essential. Keeping customers informed throughout the delivery process with regular updates on the status of their orders, expected delivery times, and any potential delays can help manage customer expectations and reduce frustration associated with late arrivals. Second, streamlining the return process can significantly improve customer satisfaction. Simplifying the return process by providing easy-to-follow instructions and pre-paid return labels can alleviate customer dissatisfaction and encourage them to shop with the brand again despite any initial issues with their purchase.

Third, establishing robust feedback loops is vital. Continuously monitoring, analyzing, and acting upon customer feedback can lead to iterative improvements in products and services. Using insights gained from this feedback ensures that customer concerns are addressed promptly and effectively. Finally, personalized support can greatly enhance the overall customer experience. Utilizing data from customer interactions to offer tailored support can make customers feel valued and understood. For example, if a customer has previously reported an issue with a specific product, ensuring that future interactions address these concerns proactively can strengthen customer relationships. Personalized support fosters a sense of importance and attention, thereby improving customer satisfaction and loyalty.

By focusing on these strategies, businesses can enhance their post-purchase support, leading to increased customer satisfaction and loyalty. These improvements, driven by data-driven insights, not only resolve immediate issues but also contribute to long-term customer retention and business success.

Result and Discussion

Descriptive Statistics and EDA Findings

The initial analysis of the Customer360Insights dataset involved calculating descriptive statistics to understand the central tendencies, variability, and overall distribution of key variables. The dataset comprises 236 records, each capturing detailed information about customer demographics, product attributes, and transaction specifics. The mean age of the customers was found to be 45.33 years with a standard deviation of 15.87 years, indicating a wide age range among the customer base. Credit scores ranged from a minimum of 600 to a maximum of 778, with a mean score of 688.59, reflecting generally good credit standing among customers. Monthly income showed significant variability, with a mean of $5,470.24 and a standard deviation of $1,442.80, suggesting diverse economic backgrounds among the customers.

Further analysis of product-related variables revealed interesting insights (see figure 2). The cost of products varied widely, with a minimum of $5.00 and a maximum of $1,000.00, and an average cost of $111.01. The prices of products also showed significant variation, ranging from $8.00 to $1,500.00, with a mean price of $173.39. The quantity of items purchased per transaction ranged from 1 to 6, with an average of approximately 3.64 items per purchase. These statistics provide a comprehensive overview of the dataset, highlighting key characteristics and distributions that are essential for more detailed analyses.

Visualizations played a crucial role in elucidating the patterns and trends within the dataset. Bar charts and pie charts were utilized to depict the distribution of key variables and to highlight significant findings from the data. One of the most striking visualizations was the bar chart showing the return rates, which indicated that 100% of the transactions involved product returns. This high return rate underscores the critical need to analyze return reasons to uncover underlying issues affecting customer satisfaction.

Additionally, bar charts were used to visualize the frequency of common return reasons, providing clear insights into the primary causes of returns. The most frequent return reasons included "poor value" (55 occurrences), "wrong item" (49 occurrences), "arrived late" (47 occurrences), "arrived damaged" (45 occurrences), and "poor quality" (40 occurrences). These visualizations highlight specific areas where improvements are needed, such as product quality, accuracy in order fulfillment, and timeliness of deliveries.

Figure 2 **Generated WordCloud**

Word clouds, as shown in figure 2, were also generated to provide an intuitive and immediate understanding of the most common terms in the negative feedback. The word cloud prominently featured terms like "poor," "value," "item," "wrong," "arrived," and "damaged," visually emphasizing the critical issues reported by customers. These visual tools not only aid in quickly grasping the key themes but also facilitate more effective communication of the findings to stakeholders.

Text Mining Results

The text mining analysis conducted on the Customer360Insights dataset provided significant insights into the common terms and phrases used by customers in their feedback and return reasons. By applying tokenization, stemming, and lemmatization, we were able to preprocess the text data and prepare it for frequency analysis. The term frequency analysis revealed the most frequently occurring words in the negative feedback. The top terms identified included "poor" (95 occurrences), "arrived" (92 occurrences), "value" (55 occurrences), "item" (49 occurrences), "wrong" (49 occurrences), "late" (47 occurrences), "damaged" (45 occurrences), and "quality" (40 occurrences).

These terms highlight the key issues that customers commonly face. For instance, the high frequency of the word "poor" suggests widespread dissatisfaction with the quality of products. The frequent occurrence of "arrived" and "late" points to significant problems with delivery times, while "damaged" indicates issues with the condition of products upon arrival. "Value" and "quality" are terms that reflect customer perceptions of the overall worth and standard of the products. The terms "wrong" and "item" together suggest frequent issues with customers receiving incorrect products. This analysis helps in pinpointing the specific aspects of the customer experience that require attention and improvement.

To visually represent the results of the term frequency analysis, word clouds were generated. Word clouds are effective visualization tools that display the most common terms in a dataset, with the size of each word corresponding to its frequency. In this study, the word cloud prominently featured terms such as "poor," "arrived," "value," "item," "wrong," "late," "damaged," and "quality."

The word cloud visualization provided an immediate and intuitive understanding of the most prevalent issues in the customer feedback. The prominence of words like "poor," "damaged," and "late" in the word cloud underscored the critical areas where customers expressed significant dissatisfaction. The visual emphasis on these terms highlights the primary pain points that need to be addressed to improve the overall customer experience.

The interpretation of the word cloud also facilitated a deeper understanding of the interconnectedness of various issues. For example, the concurrent prominence of "arrived," "late," and "damaged" suggests that problems with the delivery process are multifaceted, affecting both the timeliness and condition of received products. Similarly, the frequent appearance of "poor" alongside "quality" and "value" indicates that customers' perceptions of product worth are closely tied to their assessments of quality.

Topic Modeling Results

The application of Latent Dirichlet Allocation (LDA) to the Customer360Insights dataset yielded five distinct topics that encapsulate the primary themes in the customer feedback and return reasons. Each topic is represented by a cluster of terms that frequently co-occur in the text, providing insights into the underlying issues customers face.

The identified topics and their most relevant terms are as follows. Topic 0 includes terms such as "poor quality," "damaged," "late," "arrived," "value," and "wrong item," highlighting issues related to product quality and delivery timeliness. Topic 1 focuses on "value," "poor," "damaged," "late," "arrived," "wrong item," and "quality," indicating that value perception and quality control are significant concerns. Topic 2 consists of "wrong item," "poor," "damaged," "late," "arrived," "value," and "quality," emphasizing the frequency of incorrect items being delivered and their impact on perceived value and quality. Topic 3 includes terms like "arrived," "late," "damaged," "poor," "value," "wrong item," and "quality," pointing to the recurring issue of late and damaged deliveries. Finally, Topic 4 encompasses "poor," "damaged," "late," "arrived," "value," "item," "wrong," and "quality," underlining the persistent problems with product quality, delivery timeliness, and order accuracy.

These topics consistently highlight several recurring issues, including product quality ("poor," "damaged"), delivery problems ("late," "arrived"), value perception ("value"), and order accuracy ("wrong item"). The consistency across the topics indicates that these are pervasive concerns affecting a significant portion of the customer base. The results of the topic modeling provide a structured framework for understanding the main themes in the customer feedback. Each topic's relevance is determined by the frequency and co-occurrence of terms within the dataset. For instance, Topic 0, which includes terms like "poor quality" and "damaged," clearly points to issues with the condition of products received by customers. This topic is crucial for identifying areas where product quality control measures need to be strengthened to prevent damage and ensure that products meet customer expectations.

Similarly, the prominence of terms such as "arrived" and "late" across multiple topics underscores the critical issue of delivery timeliness. The repeated appearance of these terms indicates widespread dissatisfaction with delivery services, suggesting that improvements in logistics and shipping processes are necessary. Ensuring timely delivery can significantly enhance customer satisfaction and reduce negative feedback. The term "value" appearing in several topics highlights the importance of perceived value in customer satisfaction. Customers frequently return items they believe do not offer good value for the price paid. Addressing this issue may involve better pricing strategies, clearer product descriptions, and improved quality to align customer expectations with the product offerings.

The interpretation of these topics reveals the interconnected nature of the issues affecting customer satisfaction. For example, poor product quality not only leads to negative feedback but also affects the perceived value of the product. Additionally, delivery issues can exacerbate dissatisfaction with product quality, as items arriving late or damaged contribute to a negative customer experience. By understanding these relationships, businesses can adopt a holistic approach to address the root causes of dissatisfaction.

Identification of Quality Issues

The analysis of negative feedback provided significant insights into the key quality issues affecting customer satisfaction. Through text mining techniques, we identified several recurring themes that highlight the primary concerns of customers. The most common terms extracted from the negative feedback included "poor," "value," "item," "wrong," "arrived," "damaged," and "quality." These terms consistently pointed to specific problems related to product quality and service delivery.

One of the predominant issues identified was the perception of poor product quality. Customers frequently used terms like "poor" and "damaged" to describe their dissatisfaction with the condition of the products they received. This suggests that there are significant gaps in quality control processes, leading to a high incidence of defective or substandard products reaching customers. Additionally, the term "value" often appeared in conjunction with quality-related complaints, indicating that customers felt the products did not provide adequate value for the price paid. This highlights the need for better alignment between product quality and pricing strategies.

Delivery issues were another major concern, with terms like "arrived" and "late" frequently mentioned in the feedback. Customers expressed frustration with delayed deliveries, which negatively impacted their overall experience. The term "wrong item" also appeared frequently, indicating problems with order accuracy. These findings suggest that both logistical operations and order fulfillment processes need to be scrutinized and improved to reduce errors and enhance customer satisfaction.

To address the quality issues identified, it is essential to understand the root causes of the common return reasons. The analysis revealed several underlying factors contributing to the high return rates. For example, the frequent mention of "poor" and "damaged" in the feedback points to inadequate quality control measures during the production and packaging stages. Products that do not meet quality standards are likely to be returned, leading to customer dissatisfaction and increased operational costs.

The recurring issue of delayed deliveries, indicated by terms like "arrived" and "late," suggests inefficiencies in the logistics and supply chain management. Factors such as poor coordination between warehouses and shipping providers, inadequate inventory management, and lack of real-time tracking could be contributing to these delays. Addressing these logistical challenges is crucial for ensuring timely deliveries and enhancing the overall customer experience.

Order accuracy problems, highlighted by the term "wrong item," indicate gaps in the order fulfillment process. This could be due to errors in the picking and packing stages, where items are incorrectly selected or labeled. Implementing more robust verification processes and leveraging technology to automate and streamline order fulfillment can help reduce these errors.

Recommendations

Based on the analysis of negative feedback and return reasons, several key areas for product improvement have been identified. The primary concern revolves around product quality, as terms like "poor" and "damaged" were frequently mentioned. To address this issue, it is essential to enhance the quality control processes during both the manufacturing and packaging stages. Implementing stricter quality assurance protocols, such as regular inspections and testing, can help ensure that only products meeting high standards are shipped to customers. Additionally, investing in better packaging materials and methods can prevent damage during transit, further reducing the number of returns due to damaged goods.

Another significant issue identified was the perception of inadequate value. Customers frequently mentioned terms like "value" and "quality" together, indicating dissatisfaction with the price-to-quality ratio of the products. To improve customer perceptions of value, businesses should consider adjusting their pricing strategies to better reflect the quality of their products. This could involve offering competitive pricing, introducing tiered product lines to cater to different market segments, or providing detailed product information and customer reviews to help customers make informed purchasing decisions. Enhancing the overall quality of the products will also naturally improve the perceived value, leading to higher customer satisfaction.

Enhancing post-purchase support is crucial for maintaining customer satisfaction and loyalty. One of the major issues identified was delayed deliveries, with terms like "arrived" and "late" frequently mentioned in the feedback. To address this, businesses should focus on optimizing their logistics and supply chain operations. Implementing real-time tracking systems can provide customers with accurate delivery estimates and updates, reducing frustration associated with delays. Additionally, improving coordination between warehouses and shipping providers can streamline the delivery process and ensure timely arrivals.

Another common issue was incorrect items being delivered, as indicated by the term "wrong item." To reduce order inaccuracies, businesses should implement more robust verification processes during order fulfillment. This can include double-checking orders before dispatch and using automated systems to minimize human error. Enhancing customer support services to address these issues promptly can also improve the overall post-purchase experience. Providing customers with clear and easy return policies, along with efficient handling of returns and exchanges, can help mitigate dissatisfaction and build trust.

Discussion

The findings of this study align with previous research on customer satisfaction and product quality in e-commerce. Consistent with the Expectancy-Disconfirmation Theory, the gap between customer expectations and actual product performance emerged as a significant factor influencing satisfaction. Similar to studies that highlighted the importance of quality control and logistics efficiency, our analysis also identified these areas as critical for reducing negative feedback and returns. However, this study extends the existing literature by providing a detailed examination of specific terms and phrases used by customers, offering more granular insights into the exact nature of their concerns.

The insights gained from this analysis have significant implications for digital marketing strategies. By addressing the key issues identified—such as product quality, delivery timeliness, and order accuracy—businesses can enhance their overall customer satisfaction, which is crucial for building a positive brand reputation and fostering customer loyalty. Improved product descriptions, transparent pricing, and showcasing customer reviews can also help manage customer expectations and reduce dissatisfaction. Additionally, leveraging real-time tracking and personalized communication in post-purchase support can strengthen customer relationships and encourage repeat purchases.

Despite the valuable insights provided by this study, there are several limitations that should be acknowledged. The dataset used was synthetic, which, while designed to mimic real-world scenarios, may not capture all the nuances of actual customer behavior and feedback. Future research should aim to validate these findings using real customer data from various e-commerce platforms. Additionally, while this study focused on text mining and sentiment analysis, incorporating other data mining techniques, such as predictive modeling and customer segmentation, could provide a more comprehensive understanding of customer satisfaction drivers.

Further research could also explore the impact of different marketing strategies on customer satisfaction and retention. Longitudinal studies tracking changes in customer feedback over time in response to specific interventions could offer deeper insights into the effectiveness of these strategies. By addressing these limitations and expanding the scope of analysis, future studies can build on the findings of this research to develop more targeted and effective approaches to improving customer satisfaction in the e-commerce industry.

Conclusion

This research has provided significant insights into the factors affecting customer satisfaction within the e-commerce landscape, utilizing the Customer360Insights dataset. The analysis revealed that common issues such as poor product quality, delivery delays, incorrect items, and perceived inadequate value are primary drivers of negative feedback and returns. Text mining techniques, including tokenization, stemming, and term frequency analysis, identified the most frequently occurring terms in negative feedback, providing a clear picture of recurring issues. Sentiment analysis further quantified the extent of customer dissatisfaction, with a significant majority of feedback classified as negative.

The application of text mining and sentiment analysis proved effective in understanding customer satisfaction. These techniques enabled the extraction of valuable insights from unstructured text data, revealing the underlying themes and sentiments expressed by customers. The identification of common terms and the generation of word clouds provided a visual and analytical basis for understanding customer concerns. Additionally, topic modeling offered a structured approach to uncovering hidden themes within the feedback, highlighting specific areas requiring attention and improvement.

The findings from this research have substantial implications for enhancing digital marketing strategies in e-commerce. Addressing the identified issues—such as improving product quality, ensuring timely deliveries, and enhancing order accuracy—can significantly boost customer satisfaction and loyalty. By focusing on these areas, businesses can reduce negative feedback and return rates, fostering a more positive customer experience. Furthermore, leveraging real-time tracking systems and providing transparent, personalized communication can enhance post-purchase support, reinforcing customer trust and encouraging repeat business.

The insights gained can also inform digital marketing campaigns. Highlighting improvements in product quality and delivery processes can serve as compelling selling points. Additionally, showcasing customer reviews and transparent return policies can help manage customer expectations and build confidence in the brand. By integrating these findings into their digital marketing strategies, e-commerce businesses can create more targeted and effective campaigns that resonate with customers, ultimately driving sales and growth.

While this study has provided valuable insights, there are several avenues for future research to build on the current findings. One recommendation is to validate these results using real-world data from various e-commerce platforms, ensuring the applicability and generalizability of the insights. Additionally, incorporating other data mining techniques, such as predictive modeling and customer segmentation, could provide a more comprehensive understanding of the factors driving customer satisfaction and loyalty.

Further research could also explore the longitudinal impact of specific interventions on customer satisfaction. By tracking changes in customer feedback over time in response to improvements in product quality, delivery processes, and customer support, researchers can gain deeper insights into the effectiveness of these strategies. Moreover, expanding the scope of analysis to include different demographic segments and market contexts can offer a more nuanced understanding of customer needs and preferences.

Declarations

Author Contributions

Conceptualization: C.I.; Methodology: C.I.; Software: W.C.S.; Validation: S.A.G.; Formal Analysis: C.I.; Investigation: W.C.S.; Resources: S.A.G.; Data Curation: W.C.S.; Writing—Original Draft Preparation: C.I.; Writing—Review and Editing: S.A.G.; Visualization: W.C.S. 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] D. E. Rose, "Digital Marketing Strategy in Enhancing Brand Awareness and Profitability of E-Commerce Companies," Aptisi Trans. Manag. Atm, vol. 8, no. 2, pp. 160–166, 2024, doi: 10.33050/atm.v8i2.2277.

[2] B. H. Sugiharto, "The Role of E-Commerce for MSMEs as a Digital Marketing Strategy in Facing Industrial Revolution 4.0," Productivity, vol. 1, no. 1, pp. 99–107, 2024, doi: 10.62207/80ndq458.

[3] Anak Agung Gde Bagus Udayana, "Investigating the Role of E-Commerce Application and Digital Marketing Implementation on the Financial and Sustainability Performance: An Empirical Study on Indonesian SMEs ," Int. J. Data Netw. Sci., vol. 8, no. 1, pp. 167–178, 2024, doi: 10.5267/j.ijdns.2023.10.007.

[4] S. Kabiraj and S. Joghee, "Improving Marketing Performance: How Business Analytics Contribute to Digital Marketing," Int. J. Technol. Innov. Manag. Ijtim, vol. 3, no. 1, pp. 9–18, 2023, doi: 10.54489/ijtim.v3i1.209.

[5] Dr. B. M. H. N. Dr. Rajalakshmi Anantharaman Madhukumar. B., "Analyzing the Role of Digital Marketing in Growth of E-Commerce in India: A Multiple Holistic Approach," Jier, vol. 4, no. 2, 2024, doi: 10.52783/jier.v4i2.835.

[6] A. S. Al-Adwan, "Revealing the Influential Factors Driving Social Commerce Adoption," Interdiscip. J. Inf. Knowl. Manag., vol. 14, pp. 295–324, 2019, doi: 10.28945/4438.

[7] A. Naafiarsha, H. Respati, and A. Fridiansjah, "E-Trust and Post-Purchase Experience on Repurchase Intention Through Customer Satisfaction: Study on Instagram Account @Jastipmlg," Int. J. Bus. Appl. Soc. Sci., pp. 50–58, 2020, doi: 10.33642/ijbass.v6n8p7.

[8] E. Lobacheva and N. E. Yadova, "The Impact of Digital Technology on E-Commerce Development in Russian Federation," Matec Web Conf., vol. 311, p. 02016, 2020, doi: 10.1051/matecconf/202031102016.

[9] M. V. Dykha, T. Ustik, O. Y. Krasovska, D. Pilevych, Z. Shatska, and T. Iankovets, "Marketing Tools for the Development and Enhance the Efficiency of E-Commerce in the Context of Digitalization," Stud. Appl. Econ., vol. 39, no. 5, 2021, doi: 10.25115/eea.v39i5.5234.

[10] M. Purnamasari, “Unveiling the Synergy: How Entrepreneurial Marketing and Product Quality Drive Purchase Decisions Through the Lens of Digital Marketing,” Eco-Buss, vol. 6, no. 3, pp. 1423–1434, 2024, doi: 10.32877/eb.v6i3.1181.
[11] N. T. Q. Nga and Đ. H. Vân, “The Impacts of Customer Satisfaction on Loyalty of the Youth (18-25 Years Old) Towards E-Commerce Trading Floors in Ho Chi Minh City,” Sci. Technol. Dev. J. - Econ. - Law Manag., vol. 4, no. 4, p. First, 2020, doi: 10.32508/stdjelm.v4i4.654.
[12] K. Xiaohui, D. Mather, D. L. Ott, E. S. Fang, P. Bremer, and M. Mirosa, “Fresh Food Online Shopping Repurchase Intention: The role of post-Purchase Customer Experience and Corporate Image,” Int. J. Retail Distrib. Manag., vol. 50, no. 2, pp. 206–228, 2021, doi: 10.1108/ijrdm-04-2021-0184.
[13] A. Kumar and B. Anjaly, “How to Measure Post-Purchase Customer Experience in Online Retailing? A Scale Development Study,” Int. J. Retail Distrib. Manag., vol. 45, no. 12, pp. 1277–1297, 2017, doi: 10.1108/ijrdm-01-2017-0002.
[14] V. V. D. P. Kotni, “Paradigm Shift From Attracting Footfalls for Retail Store to Getting Hits for E-Stores: An Evaluation of Decision-Making Attributes in E-Tailing,” Glob. Bus. Rev., vol. 18, no. 5, pp. 1215–1237, 2017, doi: 10.1177/0972150917710133.
[15] S. C. Cook and J. Yurchisin, “Fast Fashion Environments: Consumer’s Heaven or Retailer’s Nightmare?,” Int. J. Retail Distrib. Manag., vol. 45, no. 2, pp. 143–157, 2017, doi: 10.1108/ijrdm-03-2016-0027.
[16] Y. Cao, H. Ajjan, and P. Hong, “Post-Purchase Shipping and Customer Service Experiences in Online Shopping and Their Impact on Customer Satisfaction,” Asia Pac. J. Mark. Logist., vol. 30, no. 2, pp. 400–416, 2018, doi: 10.1108/apjml-04-2017-0071.
[17] S. Gnanavel et al., “Rapid Text Retrieval and Analysis Supporting Latent Dirichlet Allocation Based on Probabilistic Models,” Mob. Inf. Syst., vol. 2022, pp. 1–12, 2022, doi: 10.1155/2022/6028739.
[18] R. F. Mello, M. Andre, A. G. Pinheiro, E. Costa, and C. Romero, “Text Mining in Education,” Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 9, no. 6, 2019, doi: 10.1002/widm.1332.
[19] D. Tao, P. Yang, and H. Feng, “Utilization of Text Mining as a Big Data Analysis Tool for Food Science and Nutrition,” Compr. Rev. Food Sci. Food Saf., vol. 19, no. 2, pp. 875–894, 2020, doi: 10.1111/1541-4337.12540.
[20] B. Nie and S. Sun, “Using Text Mining Techniques to Identify Research Trends: A Case Study of Design Research,” Appl. Sci., vol. 7, no. 4, p. 401, 2017, doi: 10.3390/app7040401.
[21] R. Krishnamurthy and R. K. Balasubramanium, “Using Text Mining to Identify Trends in Oropharyngeal Dysphagia Research: A Proof of Concept,” Commun. Sci. Disord., vol. 24, no. 1, pp. 234–243, 2019, doi: 10.12963/csd.19579.
[22] F. Syuhada and R. A. Pratama, “Feature Extraction Technique for Text Mining Requirement for Reuse in Software Product Lines: A Systematic Literature Review,” Sainstech Innov. J., vol. 3, no. 2, pp. 87–95, 2020, doi: 10.37824/sij.v3i2.2020.230.