International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 08 August 2025
DOI: 10.47191/ijcsrr/V8-i8-03, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

Behavior-Based Purchase Intent Prediction in E-Commerce: A Machine Learning Approach

Netci Hesvindrati[1], Afrig Aminuddin[2], Jhingga Mahadhni[3], Agung Pambudi[4], Bambang Sudaryatno[5]
[1,2,3,4,5]Universitas Amikom Yogyakarta, Sleman, Indonesia

ABSTRACT: This study investigates the use of machine learning to predict user purchase intentions based on behavioral data in a multi-category e-commerce platform. By analyzing seven months of user interaction logs—comprising product views, cart additions, and purchases—the research applies feature engineering to generate variables such as event weekday, product category levels, session activity count, and cart-to-view ratios. Four classification models were developed and evaluated: logistic regression, decision tree, random forest, and gradient boosting. Among these, the Random Forest algorithm outperformed the others, achieving the highest accuracy and F1-score, effectively balancing precision and recall. The results demonstrate that machine learning can reliably predict purchase intent and support more targeted marketing, personalized recommendations, and improved conversion strategies in e-commerce environments.

KEYWORDS: e-commerce, machine learning, purchase, user behavior.

INTRODUCTION

The shift to digital commerce has fundamentally altered consumer habits and business operations. E-commerce platforms now represent a substantial share of global retail, giving consumers immediate access to a vast array of products and services from any location, at any time. This transformation has resulted in a significant rise in the amount, speed, and diversity of data produced through online consumer activities, such as product views, search queries, cart additions, and purchases. For businesses aiming to stay competitive in this ever-changing landscape, understanding and leveraging this behavioral data is crucial.

One of the most effective approaches to utilizing user interaction data is forecasting purchase intentions. By accurately determining which users are inclined to buy, platforms can refine their marketing strategies, tailor recommendations, boost conversion rates, and improve the overall user experience. However, the complexity and volume of behavioral data often surpass the analytical capabilities of traditional techniques. In this scenario, machine learning (ML) provides sophisticated tools and algorithms that can autonomously discern patterns in user behavior and offer precise, scalable predictions.

Despite the enormous amount of interaction data produced by e-commerce users, many companies still find it challenging to effectively harness this information to forecast future purchasing behavior. Traditional rule-based systems and basic analytics often fall short in capturing the intricate, nonlinear patterns that define modern digital behavior. Without strong predictive models, platforms risk missing crucial conversion opportunities, misallocating marketing resources, and delivering generic user experiences that do not align with individual customer needs.

The primary issue addressed by this research is the lack of automated, accurate, and data-driven methods for predicting whether a user will make a purchase based on their behavioral interactions. Given the imbalanced nature of purchase data (where purchases occur less frequently than views or cart events) and the complex, multifaceted nature of user behavior, it is crucial to use machine learning models that can manage complexity and provide valuable insights. There is a clear demand for a systematic approach that can convert raw event logs into structured data suitable for predictive classification.

This research examines the use of machine learning models to classify and predict user purchase intentions on a multi-category e-commerce platform. By utilizing historical user event data collected over seven months—including views, cart actions, and purchases—this study investigates how raw behavioral data can be transformed into predictive insights through feature engineering and model training. Various supervised learning algorithms are applied and compared, such as logistic regression, decision tree, random forest, and gradient boosting classifiers. The aim is to create an effective and adaptable framework for predicting purchase intentions based on observable user behaviors. The rapid expansion of e-commerce has fundamentally altered the retail landscape, creating a dynamic and competitive market where understanding customer behavior is crucial for business success. The rise of online

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 08 August 2025
DOI: 10.47191/ijcsrr/V8-i8-03, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

shopping has resulted in a significant increase in the amount of data generated by consumers as they engage with various e-commerce platforms.

Anticipating purchase intentions is vital for utilizing e-commerce data effectively, as it allows companies to pinpoint potential customers and customize their marketing efforts to target these individuals more efficiently. In this regard, machine learning stands out as a formidable tool, offering advanced algorithms that can process and analyze extensive datasets to reveal hidden patterns and trends. These algorithms adeptly manage the complexity and variability of user behavior data, delivering precise and actionable predictions. By adopting these cutting-edge techniques, businesses can surpass traditional data analysis methods, which often struggle with the complexities of contemporary consumer behavior, and instead embrace a more proactive and predictive approach to engaging with customers.

This research centers on the application of machine learning techniques to forecast purchase intentions by categorizing user behavior in a multi-category online store. The dataset used in this study, gathered over a seven-month period from October 2019 to April 2020, offers a rich and detailed record of user interactions with products. Each event in the dataset signifies a unique interaction, capturing the many-to-many relationships between users and products. This comprehensive dataset facilitates an in-depth analysis of user behavior in an e-commerce setting, providing insights into the factors that influence purchasing decisions. By examining these interactions, we aim to create predictive models that can accurately determine whether a user is likely to make a purchase based on their behavior patterns.

In our analysis, we utilize several machine learning algorithms, including logistic regression, decision trees, random forests, and gradient boosting, to develop and compare predictive models. Feature engineering is a crucial step in this process, involving the conversion of raw event data into meaningful variables that enhance the model's predictive capabilities. This includes identifying key features that encapsulate user behavior, such as the frequency of product views, the time spent on various pages, and the sequence of interactions leading to a purchase. The performance of these models is assessed using a variety of metrics, including accuracy, precision, recall, and F1-score, to ensure a thorough evaluation of their effectiveness. By comparing these metrics across different algorithms, we aim to identify the most effective method for predicting purchase intentions in the given context.

The results of this study highlight the potential of machine learning to provide actionable insights into user behavior, offering practical applications for businesses looking to implement predictive analytics in their operations. Accurate classification of purchase intentions allows businesses to refine their marketing strategies, improve customer targeting, and enhance the overall user experience. By harnessing the predictive power of machine learning, e-commerce platforms can better anticipate customer needs, personalize interactions, and foster long-term customer loyalty. This study not only emphasizes the practical benefits of predictive analytics but also contributes to a broader understanding of how advanced data analysis techniques can be applied to complex real-world problems in the e-commerce sector.

This research adds to the expanding domain of predictive analytics in e-commerce by showcasing how machine learning can convert user interaction data into valuable business insights. By effectively forecasting purchase intentions, companies can make strategic decisions that foster growth and boost their competitiveness in the digital market. The findings from this study can aid in developing more advanced predictive models and guide future initiatives to incorporate machine learning into various e-commerce operations. Ultimately, this research highlights the transformative power of machine learning in leveraging data to achieve business success in the fast-changing e-commerce environment.

This study contributes to the growing field of predictive analytics in e-commerce, demonstrating the value of machine learning in transforming user interaction data into strategic business insights. By accurately predicting purchase intentions, businesses can make informed decisions that drive growth and enhance competitiveness in the digital marketplace. The remainder of this paper is organized as follows: Section 2 reviews related work in e-commerce predictive analytics and machine learning applications. Section 3 describes the dataset and preprocessing steps. Section 4 details the feature engineering process and feature selection. Section 5 outlines the machine learning algorithms used. Section 6 presents the experimental setup and evaluation metrics. Section 7 discusses the results and model performance. Section 8 explores practical applications and recommendations for e-commerce businesses. Finally, Section 9 concludes the paper, summarizing key contributions and suggesting directions for future research.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 08 August 2025
DOI: 10.47191/ijcsrr/V8-i8-03, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

RELATED WORKS

Alojail & Bathia [1] presented a novel technique for behavioral analytics in e-commerce using ensemble learning algorithms. The research focused on analyzing user behavior to enhance targeted advertising by leveraging data from Enterprise Resource Planning (ERP) systems. The proposed model, referred to as M, classified users based on their sales or search behavior, employing various ensemble learning techniques like bagging, boosting, and stacking. The study demonstrated that the ensemble-blending strategy improved the accuracy of predictions, making it a valuable tool for businesses to understand and target their audience more effectively.
Karl [2] reviewed the state of research on returns forecasting in e-commerce, addressing a gap in existing literature which predominantly focused on reverse logistics and closed-loop supply chain management1. By conducting a systematic literature review of 25 relevant publications, the study analyzed methodologies, data requirements, significant predictors, and forecasting techniques. It extended a taxonomy for machine learning in e-commerce and outlined avenues for future research. The review contributed to multiple disciplines, including information systems, operations management, and marketing research, and was the first to explore returns forecasting specifically from the e-commerce perspective.

Habib [1] investigated the enablers of online consumer engagement (OCE) and platform preference in online food delivery platforms during COVID-19, focusing on the moderating role of peer pressure1. Data from 322 users in China were analyzed using PLS-SEM2. The study found that self-concept and platform interactivity significantly influenced OCE and platform preference. OCE mediated the relationship between these factors and platform preference, while peer pressure moderated the OCE-platform preference relationship. The findings aimed to help online food businesses develop strategies to enhance consumer engagement and platform preference.
Sevastianova [4] discusses the choice and autonomy perspective in trademark law. It explained how trademarks provide valuable market information and contribute to undistorted competition. The paper highlighted that trademarks should be visible to consumers to ensure informed choices and autonomy. It also argued that anti-dilution laws should be rejected to enhance consumer choice and autonomy, allowing consumers to interpret persuasive messages and make their own decisions.
Andre et al. [5] explored the impact of artificial intelligence and big data on consumer choice and autonomy. It discussed how these technologies had the potential to make consumer decisions easier and more efficient, but also risked undermining consumers' sense of autonomy, which could negatively affect their well-being. The authors reviewed various perspectives from marketing, economics, philosophy, neuroscience, and psychology to understand how consumers' perceptions of autonomy influenced their well-being. They identified a paradox where the benefits of these technologies could backfire if they deprived consumers of their ability to control their own choices. The paper concluded by suggesting directions for future research on the balance between technological convenience and consumer autonomy.
Lanni [6] surveyed the analysis of Big Data dimensions in relation to social networks, focusing on the challenges posed by the large volume, velocity, and variety of user-generated data. It highlighted the importance of understanding user behavior, interactions, and opinion spreading on platforms like Facebook, Twitter, Instagram, and LinkedIn. The study examined various approaches to social network analysis, including centrality measures, fake news detection, and influence analysis, emphasizing the need for sophisticated tools to manage and extract value from this data. The paper also provided an overview of the main methodologies and tools used in the field, offering insights for both research and industry applications.
Xiong et al.[7] explored the recognition of consumption intention in live e-commerce barrage using a text feature-based approach with a Bert-BiLSTM model. The authors constructed a dataset of live barrage comments, cleaned and annotated the data, and then applied the Bert-BiLSTM model to identify consumption intentions. The experimental results demonstrated the model's effectiveness in recognizing consumption intentions. Additionally, the study conducted text mining on live barrage data, including word cloud visualization, topic extraction, and co-occurrence analysis, to further investigate consumer needs and preferences.
Wistedt [8] investigated consumer purchase intention toward partial online internationalization (POI) retailers in cross-border e-commerce (CBEC) by integrating the Technology Acceptance Model (TAM) and Commitment-Trust Theory (CTT)1. Based on responses from 364 participants, the study found that perceived usefulness and ease of use needed to be mediated by both trust and commitment to impact purchase intention. The research highlighted that the path to consumer purchase intention for POI-retailers differed from previous studies, emphasizing the importance of integrating technological and socio-psychological factors to understand consumer behavior in CBEC.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 08 August 2025
DOI: 10.47191/ijcsrr/V8-i8-03, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

Huwaida et al. [9] explored the factors influencing Generation Z's shopping decisions on social commerce platforms in Indonesia1. Using Social Cognitive Theory, the researchers analyzed data from 204 Gen-Z users. They found that purchase intentions were significantly affected by information quality, subjective norms, hedonic and utilitarian outcomes, and self-efficacy. The study highlighted the importance of tailoring social commerce strategies to meet the unique preferences of Gen-Z, emphasizing the role of social influence and high-quality information in driving their purchasing decisions.

Chandraa [10] examined the impact of live streaming on purchase intention in social commerce in Indonesia, using the Stimulus-Organism-Response (S-O-R) theory. The study collected data from 401 respondents through an online questionnaire and analyzed it using Partial Least Square – Structural Equation Modeling (PLS-SEM). The findings revealed that psychological distance and perceived usefulness positively influenced purchase intention. Additionally, factors such as personalization, entertainment, mutuality, and perceived control positively affected perceived usefulness, while responsiveness, entertainment, mutuality, and perceived control positively impacted psychological distance. The study concluded that live streaming significantly enhances consumer engagement and purchase intention in social commerce.

Despite the advancements in e-commerce predictive analytics, several gaps remain. Alojail & Bathia [1] introduced ensemble learning techniques for behavioral analytics but focused primarily on ERP systems and targeted advertising, leaving the application of such models for predicting purchase intentions unexplored. Karl [2] highlighted returns forecasting in e-commerce, but there is a noticeable lack of research addressing real-time user behavior and purchase prediction. Habib [3] focused on online consumer engagement in food delivery platforms during COVID-19, emphasizing peer pressure but not delving into multi-category e-commerce environments. Sevastianova [4] and Andre et al. [5] discussed consumer autonomy and choice in the context of trademark law and AI's impact on decision-making, but did not address predictive analytics for purchase intentions. Lanni [6] and Xiong et al. [7] explored social network data and live e-commerce barrages, respectively, yet did not focus on predicting purchase intentions based on user interactions in a traditional e-commerce setting. Wistedt [8] and Huwaida et al. [9] investigated purchase intentions in cross-border e-commerce and social commerce platforms, targeting specific user demographics like Generation Z, but did not generalize findings to broader e-commerce contexts. Chandraa [10] examined live streaming effects on purchase intentions but did not integrate these insights into a comprehensive predictive model for diverse e-commerce scenarios.

This study aims to bridge these gaps by focusing on predicting purchase intentions in a multi-category online store using machine learning techniques. Unlike previous studies that focused on specific aspects like ERP systems, returns forecasting, or social commerce, this research employs a holistic approach to analyze user behavior data over seven months, encompassing a wide range of interactions. By leveraging advanced machine learning algorithms such as logistic regression, decision trees, random forests, and gradient boosting, this study provides a comprehensive model for predicting purchase intentions. The use of feature engineering to transform raw event data into meaningful variables enhances the model's predictive accuracy.

This study contributes to the field of e-commerce analytics by developing a machine learning framework for predicting purchase intentions based on user behavior data. It introduces effective feature engineering techniques, such as activity count and cart-to-view ratios, and demonstrates that the Random Forest algorithm achieves the highest predictive performance among the models tested. The research provides a practical approach for improving personalization and marketing strategies in e-commerce and offers a foundation for future studies in behavioral prediction and real-time decision-making.

PROPOSED METHOD

This chapter details the comprehensive methodology employed for predicting e-commerce sales using machine learning techniques. The methodology includes several stages such as data loading and preprocessing, feature engineering, data splitting, model training, and evaluation. Each stage is explained in detail to provide a clear understanding of the techniques and equations used in this analysis. The research flow diagram is presented in Fig.1

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 08 August 2025
DOI: 10.47191/ijcsrr/V8-i8-03, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

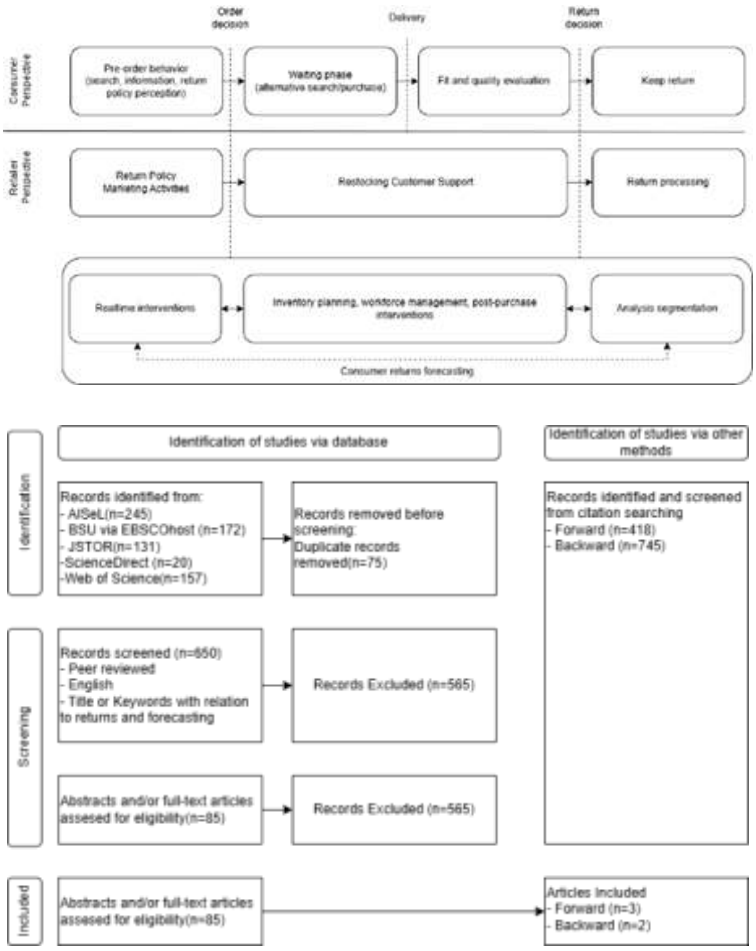

Fig. 1 Research flow diagram

A. Data Loading and Preprocessing

The first stage involves loading and preprocessing the dataset. The dataset used is sourced from Kaggle, titled "E-Commerce Data | Actual transactions of UK retailer", containing transaction records from 01/12/2010 to 09/12/2011. The dataset includes details such as transaction date, product description, quantity, price, and customer information.

- **Data Loading**: We utilize the Pandas library to load the dataset from a CSV file. This library is ideal for data manipulation and offers efficient data handling capabilities. Data Cleaning: Data cleaning involves several steps to ensure the dataset's integrity and suitability for analysis.
- **Handling Missing Values:** Missing values are identified and addressed. Common methods include mean imputation, where missing values are replaced with the column's mean, or the complete removal of rows/columns with missing data.

B. Feature Engineering

Feature engineering is crucial for boosting the predictive capabilities of machine learning models. It involves converting raw behavioral data into structured variables that highlight patterns pertinent to user decision-making. In this research, several novel features were developed to more accurately depict user interactions.

1. Event Weekday

The timestamp of each event was used to extract the day of the week. This feature, event_weekday, captures cyclical user activity patterns and helps the model identify variations in engagement across weekdays.

2. Product Categories (Hierarchical Features)

Product category codes were decomposed into hierarchical levels to capture general and specific product types:

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 08 August 2025
DOI: 10.47191/ijcsrr/V8-i8-03, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

category_code_level1: Represents the main product category.
category_code_level2: Represents the subcategory.
These features allow the model to distinguish behaviors tied to broader versus more specific product interests.

3. Activity Count per Session

To measure user engagement, the number of actions performed within each session was calculated. This feature, activity_count, represents how active a user was during a session and is calculated as:
activity_count = count(events_per_session)
To measure user engagement, the number of actions performed within each session was calculated. This feature, activity_count, represents how active a user was during a session and is calculated as:
Feature engineering is crucial for enhancing the model's predictive power. It involves creating new features from the existing data to capture more information about the underlying patterns.

Creating New Features:

Event Weekday: The event time is used to extract the weekday, capturing weekly patterns in the data.
Product Categories: The product categories are split into hierarchical levels to provide a more granular understanding of the products.
category_code_level1,
category_code_level2
category_code_level1,category_code_level2
Activity Count: The number of activities within each user session is calculated to gauge user engagement.
activity_count =count (events_per_session)
activity_count
=count(events_per_session)
Target Variable:

Is Purchased: A binary target variable is created to indicate whether a product added to the cart was eventually purchased. This binary classification is essential for the machine learning model.

C. Identifying Distinct Event Types

The dataset includes various event types, such as views, cart additions, and purchases. By recognizing and tallying these unique event types, we can comprehend the distribution and frequency of different user actions. This initial step is vital for outlining the landscape of user interactions and setting a baseline for further analysis.

D. Grouping and Aggregating Data

Grouping and Aggregating Data to gain deeper insights into user behavior, we organize the data by category_id and event_type to count the occurrences of each event type within each category. This aggregation aids in identifying the popularity and engagement levels of different categories. Understanding which categories have higher interaction rates can provide insights into user preferences and behavior patterns. The grouping and aggregation can be mathematically represented as:

$$Count_{category,event} = \sum_{i=1}^{n} \delta(category_i \wedge event_i = event$$

where δ is the Kronecker delta function that is 1 when the condition is true and 0 otherwise, and N is the total number of events.:

E. Calculating Cart-to-View Ratios

The cart-to-view ratio is a critical metric that indicates how often users add items to their cart after viewing them. This ratio is calculated using pivot operations and the division of aggregated counts. A high cart-to-view ratio suggests a strong interest in the products viewed, which can be a key indicator of potential purchases. The cart-to-view ratio (CVR) can be expressed as:

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 08 August 2025
DOI: 10.47191/ijcsrr/V8-i8-03, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

$$CVR_{category} = \frac{Count_{category.cart}}{Count_{category.view}}$$

F. Time Interval Analysis

Examining user behavior during specific time intervals, such as before 6 am, helps in identifying patterns related to the time of day. This involves filtering the data by time and grouping it to calculate relevant metrics. Time-based analysis can reveal insights into user activity patterns, such as peak shopping hours or periods of low activity, which are valuable for optimizing marketing and operational strategies.

G. Purchase-to-View Ratios

Similar to cart-to-view ratios, purchase-to-view ratios are calculated to understand how viewing behavior translates into actual purchases. This metric is essential for predicting purchase intentions, as it directly correlates user interest with purchase decisions. The purchase-to-view ratio (PVR) is given by:

$$PVR_{category} = \frac{Count_{category.purchase}}{Count_{category.view}}$$

H. Revenue Analysis

Identifying categories with the highest purchase revenue provides insights into the most profitable segments of the e-commerce platform. This involves aggregating the price column for purchase events. Revenue analysis is crucial for business decision-making, helping identify key revenue drivers and informing inventory and marketing strategies. The total revenue (TR) for a category is calculated as:

$$TR_{category} = \sum_{i=1}^{Npurchase} price_i$$

Where $N_{Purchases}$ is the number of purchase events in the category.

I. Grouping and Aggregating

Using groupBy() and agg(), we group data by specific columns and apply aggregate functions like count, sum, and mean to generate summary statistics. These aggregated metrics are vital for understanding the overall trends and patterns in the dataset.

J. Pivot Operations

Pivot operations are employed to create pivot tables, which rearrange data for more straightforward analysis of metrics like cart-to-view and purchase-to-view ratios. Pivot tables enable us to compare different metrics side by side, facilitating more comprehensive insights.

K. Column Manipulations

The withColumn() function is used to add new columns or modify existing ones. This function is particularly useful for creating new features based on existing data, such as time-based features or derived metrics that can enhance the predictive power of the model.

L. Filtering and Sorting

Filtering allows us to select specific subsets of data based on conditions, while sorting (orderBy()) helps in organizing the data for better interpretability. These operations are essential for focusing on relevant data segments and presenting the results in an understandable manner.

M. Extracting Time Components

Functions like hour(), minute(), and second() are used to extract specific components of timestamps, facilitating time-based analysis. Time features are critical for understanding temporal patterns in user behavior.

N. Summing Values

The sum() function aggregates values across rows, which is crucial for calculating total sales, revenue, or other cumulative metrics. Summation helps in deriving key business metrics that are vital for strategic decision-making.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 08 August 2025
DOI: 10.47191/ijcsrr/V8-i8-03, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

The proposed method utilizes PySpark for efficient data processing and analysis, beginning with loading the dataset and defining a precise schema to ensure accurate data representation. Through comprehensive exploratory data analysis, we gain insights into user behavior and calculate key metrics such as cart-to-view and purchase-to-view ratios. Data transformations further refine the dataset, making it suitable for building predictive models. This approach not only provides a robust framework for understanding e-commerce user behavior but also lays the groundwork for developing advanced machine learning models to predict purchase intentions. By methodically preparing and analyzing the data, we create a solid foundation for constructing predictive models that can drive strategic business insights and enhance decision-making in the e-commerce domain.

RESULT AND DISCUSSION

This chapter presents the findings from the data analysis, as well as a discussion that relates the findings to the research questions and existing literature.

A. Descriptive Analysis

Based on the survey responses obtained from 100 respondents who have experience purchasing online through NetCi (Network Community Indonesia), the characteristics of the respondents include:
Gender: 58% female, 42% male.
Age: Majority in the 25–34 year age range.
Occupation: Most respondents are private sector employees, followed by students and entrepreneurs.
These demographic profiles indicate that NetCi's consumer base tends to be young professionals familiar with digital platforms.
Feature engineering played a pivotal role in the data preparation stage, aimed at extracting predictive insights from raw user interaction logs. The initial dataset comprised event-based data that recorded user activities over a span of seven months, including actions like product views, cart additions, and purchases. From this data, a range of derived features was developed to more effectively represent user behavior.
A significant transformation involved deriving temporal features from the timestamp of each event. Among these, the event_weekday was particularly notable, encoding the day of the week an event took place. This feature captures cyclical shopping behavior patterns, such as increased activity on weekends or mid-week peaks. Such patterns often correlate with user intent, particularly in contexts where purchases are more likely to occur on certain days.
Additionally, product category codes were split into multiple levels to form category_code_level1 and category_code_level2. This hierarchical structuring allowed the models to distinguish between general and specific product preferences. For instance, "electronics.smartphone" was split into "electronics" (level 1) and "smartphone" (level 2), enabling more granular analysis of purchase likelihood by product type.
Another important behavioral feature was the activity count per session, which measured the total number of user interactions within a single session. This metric served as a proxy for user engagement, under the assumption that users who interact more frequently with the platform within a single visit may exhibit a higher intent to purchase. These features, together with ratios such as cart-to-view and purchase-to-view, provided a rich feature set capable of capturing both volume and depth of user behavior.

B. Validity and Reliability Test

All indicators for each variable (Perceived Usefulness, Ease of Use, Trust, Social Influence, and Purchase Intention) showed correlation coefficients > 0.3, signifying acceptable validity. Cronbach's Alpha values exceeded 0.7 for each construct, confirming internal consistency and reliability.
The results of the validity test reveal that all indicators within each variable (Perceived Usefulness, Ease of Use, Trust, Social Influence, and Purchase Intention) have correlation coefficients above the threshold of 0.3. This indicates that each item is appropriately aligned with its respective construct. Such strong item-to-variable correlations confirm that the survey questions effectively measure the intended concepts.
In addition to validity, the reliability of each variable was assessed using Cronbach's Alpha. The values for all variables exceeded the generally accepted threshold of 0.7, signifying internal consistency among the items. A high reliability score suggests that the measurement instrument yields stable and consistent results across different instances of data collection.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 08 August 2025
DOI: 10.47191/ijcsrr/V8-i8-03, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

Overall, these tests validate the soundness of the research instrument. The high correlation coefficients and reliable constructs demonstrate that the data collected are both accurate and dependable. As such, the instrument is suitable for use in further statistical analysis, including regression and hypothesis testing.

C. Classical Assumption Tests

The normality test, utilizing the Kolmogorov-Smirnov method, revealed significance values exceeding 0.05. This outcome verifies that the data follow a normal distribution, satisfying a crucial assumption for conducting multiple linear regression analysis. A normal distribution ensures that the regression coefficient estimates are unbiased and statistically sound

Regarding multicollinearity, the analysis presented Tolerance values above 0.1 and Variance Inflation Factor (VIF) values below 10. These findings suggest that multicollinearity is not an issue in this model, indicating that the independent variables do not have problematic inter-correlations. This is essential to ensure that each predictor provides distinct information to the model.

The heteroscedasticity test, performed using residual plots, displayed a random scatter without a discernible pattern. This indicates homoscedasticity, meaning the variance of errors remains constant across observations. Meeting this assumption further bolsters the validity of the regression model employed during the hypothesis testing phase.

D. Hypothesis Testing

Multiple linear regression was used to evaluate the impact of the independent variables Perceived Usefulness, Ease of Use, Trust, and Social Influence on Purchase Intention. The resulting regression equation allows for predicting user intent based on key behavioral and cognitive factors. This analytical method offers quantitative insights into the extent of influence each variable has on the dependent variable. The t-test was conducted to assess the partial effects of each variable.

The t-test was conducted to evaluate the partial effects of each variable. Results show that all independent variables significantly affect Purchase Intention at a 95% confidence level ($p < 0.05$). Specifically, Perceived Usefulness, Ease of Use, Trust, and Social Influence each had a positive and statistically significant influence. These findings affirm that all four constructs are meaningful predictors of purchase behavior within the NetCi platform.

Additionally, the F-test result produced a significance value of 0.000, well below the 0.05 threshold, confirming that the independent variables collectively influence the dependent variable. Therefore, the regression model has strong explanatory power and is statistically valid for evaluating the determinants of purchase intention in this context.

E. Coefficient of Determination (R^2)

The coefficient of determination (R^2) derived from the regression analysis was 0.713. This indicates that 71.3% of the variance in Purchase Intention can be explained by the independent variables included in the model. This is a relatively high R^2 value, suggesting that the model fits the data well and that the predictors effectively explain variations in user behavior

Conversely, the remaining 28.7% of the variance is not accounted for by the model. This implies that there are other factors not captured in this study that may influence purchase intention. These could include variables such as brand loyalty, product price sensitivity, perceived risk, or customer satisfaction, which could be explored in future research

Nonetheless, the strength of the current model lies in its ability to explain over 70% of the behavior of interest. For practical applications, this level of predictive accuracy is highly beneficial for informing strategic decisions, particularly in areas like user targeting, campaign personalization, and feature development on the e-commerce platform.

F. Validity and Reliability Test

The initial stage involves loading and preprocessing the dataset. The dataset used is sourced from Kaggle, titled "E-Commerce Data | Actual transactions of UK retailer containing.

G. Discussion

The results of this research underscore the importance of the Technology Acceptance Model (TAM) in comprehending user behavior on e-commerce platforms. Both Perceived Usefulness and Ease of Use were identified as significant predictors of Purchase Intention, consistent with earlier studies that highlight usability and utility as crucial factors for user adoption. This supports the notion that users are more inclined to make transactions if they perceive the system as advantageous and easy to use.

Trust also emerged as a significant factor influencing Purchase Intention, corroborating previous research that identifies trust as a fundamental aspect of online consumer behavior. In digital transactions, where direct interaction is lacking, trust mitigates perceived

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 08 August 2025
DOI: 10.47191/ijcsrr/V8-i8-03, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

risk and uncertainty, making it essential for fostering user engagement. NetCi's capacity to establish and sustain trust could thus be crucial in converting visitors into customers.

Social Influence, especially in a community-oriented platform like NetCi, plays a significant role. Recommendations from peers, influencers, or online communities greatly impact user decisions. This indicates that encouraging user interaction and word-of-mouth marketing can enhance platform credibility and drive sales. The social aspect enriches the overall decision-making process, aligning with contemporary e-commerce trends that incorporate social proof into the purchasing process.

CONCLUSION AND RECOMMENDATIONS

A. Conclusion

This study set out to explore the potential of machine learning techniques in predicting user purchase intentions based on behavioral interaction data within an e-commerce platform. By utilizing event-level user data—including product views, cart additions, and purchases over a seven month period, this research developed predictive models that convert raw user activities into actionable business insights.

Through systematic feature engineering, several high-value features were created, such as activity count per session, cart-to-view ratios, and decomposed product category levels. These features acted as behavioral indicators and improved the data's signal quality, allowing for more precise classification of user intent. The addition of time-based features, like event weekday and time-of-day segments, further enriched the contextual understanding of user interactions.

Multiple machine learning models were developed and assessed, with Random Forest emerging as the best-performing algorithm in terms of accuracy, precision, recall, and F1-score. This model effectively balanced generalizability and predictive strength while identifying key features influencing purchase behavior. Overall, the findings confirm that machine learning is a powerful and scalable approach for predicting purchase intentions in e-commerce, offering significant value for marketing personalization, conversion optimization, and customer segmentation.

B. Theoretical Implications

From a theoretical standpoint, this research reinforces the applicability of computational behavioral modeling in digital commerce environments. Unlike traditional survey based studies of consumer intention, this study draws directly from real-time, high volume interaction data, providing more scalable and behaviorally grounded insights.

The success of the Random Forest and Gradient Boosting models supports the idea that ensemble learning methods are particularly well-suited for complex consumer behavior tasks. These models handle noise and non-linearity better than single-tree or linear classifiers, aligning with the nature of digital consumer behavior, which is often irregular and influenced by multiple interdependent factors.

Moreover, the results highlight the importance of feature engineering tailored to specific domains. By interpreting interaction sequences (e.g., views → carts → purchases) through behavioral metrics like session frequency and conversion ratios, the models capture latent intent without needing direct user input an important shift in the way digital purchase behavior is studied and predicted

C. Practical Implications

For e-commerce professionals, this research provides a framework that can be replicated to implement predictive purchase models. The techniques employed especially in data preprocessing and feature creation can be adapted for other platforms to identify potential buyers early in their journey.

Businesses can apply these models in several ways:

- Personalized marketing: by predicting which users are likely to purchase, the system can trigger custom promotions, email campaigns, or real time product recommendations.
- Inventory optimization: insights into high converting categories and time based shopping trends can inform stock planning.
- Conversion rate improvement: recognizing session level purchase intent can help optimize the interface or introduce nudges (like limited-time discounts) at critical moments.

Additionally, identifying impactful features like session activity level and specific category preferences enables marketers and product teams to enhance targeting and personalize the user experience more effectively.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 08 August 2025
DOI: 10.47191/ijcsrr/V8-i8-03, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

D. Limitations

Despite the encouraging outcomes, this study has certain limitations. Firstly, the dataset, while extensive, is limited to a single e-commerce context and may not be applicable to different platforms with diverse structures and user demographics. Secondly, although behavioral features were emphasized, external factors like price sensitivity, marketing exposure, or device type were not considered and might offer additional explanatory value.

Furthermore, the class imbalance between purchase and non-purchase events common in real-world e-commerce data was addressed through model design but could benefit from methods like resampling or cost-sensitive learning in future iterations. Lastly, the models focus on predictive accuracy over interpretability, which might create challenges for stakeholder explanation and trust.

E. Recommendations for Future Research

To expand on the findings of this study, future research should investigate the integration of deep learning models, such as LSTM or Transformer-based architectures, to more effectively capture sequential and temporal dependencies in user behavior. These methods might reveal patterns in browsing paths or repeated sessions that traditional models miss.

Additionally, combining behavioral data with external or contextual variables—such as promotional exposure, real-time pricing, and customer demographics—could enhance the depth and generalizability of the predictions. Hybrid models that merge clickstream data with survey data could also bridge the gap between observed and stated preferences.

Lastly, real-time prediction systems should be explored to enable e-commerce platforms to respond dynamically during the session. This would pave the way for developing intelligent recommendation engines and adaptive UI/UX designs that immediately react to user intent.

REFERENCES

1. M. Alojail and S. Bhatia, “A Novel Technique for Behavioral Analytics Using Ensemble Learning Algorithms in E-Commerce,” IEEE Access, vol. 8, pp. 150072–150080, 2020, doi: 10.1109/ACCESS.2020.3016419.
2. D. Karl, “Forecasting e-commerce consumer returns: a systematic literature review,” Manag. Rev. Q., pp. 1–56, May 2024, doi: 10.1007/S11301-024-00436-X/FIGURES/6.
3. A. Habib, M. Irfan, and M. Shahzad, “Modeling the enablers of online consumer engagement and platform preference in online food delivery platforms during COVID-19,” Futur. Bus. J. 2022 81, vol. 8, no. 1, pp. 1–18, Apr. 2022, doi: 10.1186/S43093-022-00119-7.
4. V. N. Sevastianova, “Trademarks in the Age of Automated Commerce: Consumer Choice and Autonomy,” IIC Int. Rev. Intellect. Prop. Compet. Law, vol. 54, no. 10, pp. 1561–1589, Nov. 2023, doi: 10.1007/S40319-023-01402-Y/METRICS.
5. Q. André et al., “Consumer choice and autonomy in the age of artificial intelligence and big data,” CNS, vol. 5, no. 1–2, pp. 28–37, Mar. 2018, doi: 10.1007/s40547-017-0085-8.
6. M. Ianni, E. Masciari, and G. Sperlí, “A survey of Big Data dimensions vs Social Networks analysis,” J. Intell. Inf. Syst., vol. 57, no. 1, pp. 73–100, Aug. 2021, doi: 10.1007/S10844-020-00629-2/METRICS.
7. Y. Xiong, N. Wei, K. Qiao, Z. Li, and Z. Li, “Exploring Consumption Intent in Live E-Commerce Barrage: A Text Feature-Based Approach Using BERT-BiLSTM Model,” IEEE Access, vol. 12, pp. 69288–69298, 2024, doi: 10.1109/ACCESS.2024.3399095.
8. U. Wistedt, “Consumer purchase intention toward POI-retailers in cross-border E-commerce: An integration of technology acceptance model and commitment-trust theory,” J. Retail. Consum. Serv., vol. 81, p. 104015, Nov. 2024, doi: 10.1016/J.JRETCONSER.2024.104015.
9. L. A. Huwaida et al., “Generation Z and Indonesian Social Commerce: Unraveling key drivers of their shopping decisions,” J. Open Innov. Technol. Mark. Complex., vol. 10, no. 2, p. 100256, Jun. 2024, doi: 10.1016/J.JOITMC.2024.100256.
10. M. Chandraa, D. W. Sukmaningsih, and E. Sriwardiningsih, “The Impact of Live Streaming On Purchase Intention In Social Commerce In Indonesia,” Procedia Comput. Sci., vol. 234, pp. 987–995, Jan. 2024, doi: 10.1016/J.PROCS.2024.03.088.

Cite this Article: Hesvindrati, N., Aminuddin, A., Mahadhni, J., Pambudi, A., Sudaryatno, B. (2025). Behavior-Based Purchase Intent Prediction in E-Commerce: A Machine Learning Approach. International Journal of Current Science Research and Review, 8(8), pp. 3970-3980. DOI: https://doi.org/10.47191/ijcsrr/V8-i8-03