Online Shoppers Purchasing Intention

Loke Sheng Heng   
Bachelor of Information Systems (Honours) (Data Analytics)  
Sunway University  
Petaling Jaya, Selangor, Malaysia  
22081376

Then Ye Herng  
Bachelor of Information Systems (Honours) (Data Analytics)  
Sunway University  
Petaling Jaya, Selangor, Malaysia  
22072466Yong Kar Yee  
Bachelor of Science (Honours) in Computer Science  
Sunway Universiry  
Petaling Jaya, Selangor, Malaysia  
22057749

Chua Guangrou  
Bachelor of Science (Honours) in Computer Science  
Sunway Universiry  
Petaling Jaya, Selangor, Malaysia  
22086037

ABSTRACT  
Understanding shopping intentions among online consumers is critical for e-commerce businesses to optimize marketing and enhance customer satisfaction. Most e-commerce sites even with consumer information available are not in a position to ascertain major influences on online buying behavior and therefore cannot make effective marketing choices. This study aims to investigate the purchase intent of e-consumers through the determination of significant demographic and behavioral variables influencing their choice and forecasting purchase probability using machine learning. The UCI Machine Learning Repository's Online Shoppers Purchasing Intention Dataset was utilized with preprocessing including categorical variable encoding, numeric feature scaling and SMOTE implementation to rectify class imbalance. Moreover, the data were split into 70/30 training and test sets and six classification models were employed and compared with Accuracy, Precision, Recall, F1-Score and ROC-AUC and cross-validation for stability, which are the Logistic Regression, Decision Tree, Random Forest, Support Vector Classifier (SVC), XGBoost and LightGBM. In addition, early signals indicate that ProductRelated\_Duration, ProductRelated, VisitorType and Informational\_Duration are strong indicators of purchase intention, with a greater Informational\_Duration which indicating a higher tendency to buy. These results enable online retailers to develop targeted marketing campaigns to drive conversions. Therefore, real-time behavior monitoring is an area that needs to be studied in future research to continue to enhance predictive model performance.

# Introduction

## Background

In the digital era, data-driven insights are increasingly become important because e-commerce platforms utilize them to understand information about customer behavior to ensure they optimize user interactions. The predictive analysis will be relevant to identify the customers who are most likely to make the purchase in order to enable the companies to enhance their marketing campaigns and make their allocations more strategic. With an increase of competition in online retail, accurate and precise predicting of consumer buying plans has become more important for optimizing sales rates and overall earnings.

## Problem Statement

Regardless the accessibility of extensive consumer behavior data, many companies are faced with the challenge of converting such data to usable information due to issues associated with data imbalance and difficulty in modelling purchase intentions among others. The absence of adaptable and explainable predictive models makes it difficult for real-time customization and decision making.

## Objectives

This paper will aim at exploring the machine learning algorithms, particularly the classification models to predict future shopping behaviour of the consumers on the online platforms using real-life web browsing sessions. This study intent to identify key aspects that determine the choice of users and examine the performance of various predictive models.

## Scope of study

The analysis can only be done on the data collected within the dataset which contains various features like administrative time, bounces rate, exit rate, technical or marketing alerts. It excludes user demographics data and does not include monitoring of the real-time activity of users.

## Significance of the Study

This study is important for companies, data analysts, and digital marketers that are attempting to understand and maximize their online sales strategies. The predicted insights have potential to result in enhanced targeting, specialized campaigns and more effective use of budget resources.

# Literature Review

## Introduction

This section reviews past research on predicting online shoppers’ purchase intentions using machine learning and deep learning techniques. It mainly explores how various models, data preparation techniques, and cleaning steps influence the ability to forecast customer buying behavior in e-commerce. This topic is important because improved predictions help businesses enhance advertising, reduce cart abandonment, and increase customer satisfaction. The studies included here were selected for their direct relevance to purchase intention prediction, encompassing classical machine learning algorithms, modern deep learning models, as well as some traditional or hybrid analytical methods used for comparison. All reviewed works were published since 2019 in trusted journals and conferences. These works contribute to a clearer understanding of online shopping behavior. The review also highlights what each method does well and where it falls short, offering a clearer picture of what still needs further research and development.

## Review of Articles

1. Machine Learning for Predicting Online Shoppers’ Purchase Intentions by Treviño and Cepeda

Treviño and Cepeda [1]investigated the problem of correctly predicting the customer's intention to buy after visiting the web shop, namely using a UCI Online Shoppers Purchasing Intention Dataset, a highly imbalanced dataset with only a 15.5 percent purchase rate. The authors performed extensive data preprocessing by one-hot encoding and normalization and applied SMOTE for synthetic oversampling to solve the problem of an unbalanced dataset. They tested the Logistic Regression, Random Forest, and Naive Bayes Classifiers, and experimented on various feature choices like recursive feature elimination and metaheuristics. The main findings show that ensemble methods, particularly Random Forest with sophisticated feature selection, provide the best F1-scores and reliably recognize non-linear customer behavior patterns, with session timing and proximity to special days being among the most powerful features. Nevertheless, the study considers that despite the obtained strong results, there is still more to gain by incorporating other non-linear classifiers, optimizing deep learning architectures, and experimenting with genetic algorithm-based feature selection, pointing to the room for improvement and generalization.

1. On the Platform but Will They Buy? Predicting Customer’s Purchase Behavior Using Deep Learning by Chaudhuri et al.

Chaudhuri et al. [2] investigated how actual purchasing behaviors in e-commerce may be predicted based on large, detailed data, which constitutes user session data, behavior history, and customer characteristics. The research addresses the issue of distinguishing mere customer interest and actual decisions about buying the product, using complicated preprocessing, such as the missing values treatment, outlier handling, and one-hot encoding. Both machine and deep learning models, such as Deep Neural Networks (DNN), Decision Trees, Random Forest, and Support Vector Machines, are carefully compared, of which DNNs consistently outperform in accuracy, F1-score, as well as ROC-AUC. A comprehensive analysis of feature steps underlines the predictive importance of such factors as session time, customer loyalty, account age, and measures of engagement activity (cart activity), which demonstrates the significance of variables of both temporal and behavioral nature. The main restriction identified is that, in comparison with traditional approaches, deep learning models achieve better results but require more computational resources, and it has not been verified whether these approaches are scalable or can be practically implemented immediately.

1. Real-Time Prediction of Online Shoppers’ Purchasing Intention Using Multilayer Perceptron and LSTM Recurrent Neural Networks by Sakar et al.

Sakar et al. [3] introduced a real-time predictive e-commerce system that simultaneously evaluates a customer’s purchase intention and a session abandonment risk based on static and sequential behavioral data. The proposed solution includes a dual-module architecture, the first involving multilayer perceptron (MLP) models and sophisticated feature selection algorithms on the static session data. In contrast, the second involves LSTM recurrent neural networks to capture and evaluate temporal trends in the sequence of clicks made by the customer. Experiments show that MLP produces high predictive performance and F1-score of purchase intent, whereas the LSTM module proves robust in detecting nuanced abandonment patterns. The study investigates that combining static features and sequential behavior analysis improves predictive capability and enforces a practical scalability for deployment in a high-scale, real-time e-commerce domain. However, one crucial limitation is that the effectiveness of the system and the degree to which it is generalizable were not tested in the real-life production environment, which implies that the system should be validated and improved before adoption by the industry.

1. Research And Analysis of Online Shopper Intention by Wang

Wang [4] covers online consumer intention prediction to improve e-commerce experience and competitiveness through consumer behavior and acceptance determinants knowledge. It explores collaborative filtering algorithms which classify customers based on their past product ratings through memory-based or more accurate model-based approaches and hybrid recommendation methods that combine algorithms for accuracy and stability. Moreover, the study concludes that online purchasing, particularly of flight tickets, cinema tickets and books that are motivated by habit and motivation and also good website design, security and product variety enhance shopping intention. In addition, collaborative filtering and hybrid methods are concluded to be superior for product recommendation but are complicated. The study has limitations with regard to unavailable data on sample size, collection process or demographics which reduce generalizability. Repetitive material in the introduction may affect clarity and the study disregards upcoming AI technology, real-time data analysis, cross-cultural variation along with non-English-speaking consumers, with uncertain details on hybrid method use.

1. Real-Time Prediction of Online Shoppers' Purchasing Intention Using Random Forest by Baati and Mohsil

Baati and Mohsil [5] study aim to predict online shoppers’ purchasing intentions in real-time on e-commerce sites to trigger instant marketing actions and prevent cart abandonment by using session and user data with three machine learning methods like Naive Bayes Classifier (NBC), C4.5 decision tree and Random Forest (RF) an ensemble of decision trees. The Random Forest showed the best accuracy (88.78%) and F1 Score (0.80) over a class-imbalanced data set of 12330 user sessions (84.5% no buy, 15.5% buy), using SMOTE to address class imbalance, with columns like browser, region, visitor type and a discretized column. The model may be deployed to systems like Sakar to deliver targeted marketing offers based on clickstream behavior, optimizing conversion rates and buyer retention for greater profitability. However, the dataset lacks demographic data and clickstream data which limits generalizability and prediction accuracy, while issues of class imbalance, absence of real-world testing, cross-cultural analysis and system integration verification remain unsolved.

1. The Impact of Online Reviews on Consumers’ Purchasing Decisions by Chen et al.

Chen et al. [6] examines the effect of consumer reviews on purchase behavior of consumers, viz., specifically including the emotional orientation of reviews (positive or negative), gender roles and identifying fake reviews in a controlled Taobao.com mobile phone purchase scenario. Eye-tracking was conducted using an Eyelink 1000 where fixation rate and dwell time were measured for 20 reviews per product (one being a fake negative review) for each of 40 participants (20 male, 20 female). Moreover, ANOVA and correlation analysis revealed that negative reviews received more attention especially from females, while males attended to positive reviews and product characteristics, the attention to negative reviews led to non-purchase behaviors. Consumers were not able to detect fake reviews and no specific eye movement patterns were identified while gender affected review processing and eye-tracking accurately predicted behavior. Out of 12,403 reviews on Taobao.com and JD.com, 80 (40 positive, 40 negative, one being fraudulent) were removed where highlighting aspects like battery life. The strengths are a small sample size of 40, a minuscule 20 reviews per product, lack of brand data, focus on mobile phones and rudimentary fake review detection, limiting realism and generalizability.

1. Consumer Online Shopping Behavior Prediction Based on Machine Learning by Chen

In attempt to predict the consumer online shopping behavior of a shopping portal, Chen [7] designed a machine learning algorithm using XGBoost to forecast the behavior by evaluating the real-time user behavior, demographic information and product information of an e-commerce platform. The study overcame the important problems, such as the imbalance of classes and data complexity by introducing such methods like feature engineering and cross-validation. The model achieved an accuracy of 0.91 in its capacity to predict the purpose of making purchases. Although the study provided robust findings, it only concentrated on a single algorithm and failed to investigate deep learning or complex interpretability approaches. The shortage of cultural adjustment studies and comparative models hindered the model’s practicality in more diversified e-commerce environments.

1. Optimizing Revenue Generation in E-Commerce: An Analysis of Customer Purchase Intent Using Machine Learning Technique by Kalathia et al.

Kalathia et al. [8] investigated machine learning to predict customer purchase online intentions that are based on a dataset retrieved in the machine learning repository at UCI. The study underscored the importance role of revenue optimization in online retail and deployed Random Forest, famous for its capability to deal with unbalanced data and non-linear correlations. The optimization of the generalization and the reduction of overfitting became practical due to the implementation of the hyperparameter adjustment. The outcomes indicated that Random Forest provided consistent and pinpoint predictions of buying behavior. However, the study wasn’t managed to compare the various algorithms involved or address other interpretability issues like facilitating visual display of the feature importance, which is useful in real-life decision-making scenarios.

1. Predicting the Intention of Online Shoppers’ Purchasing by Sand and Wu

Sang and Wu [9] overcome the issue of predicting the actual plans of online consumers of making a purchase in real-time in order to help online retailing platforms to identify potential purchasers. A Random Forest algorithm is used on session data (operating system, web browser, page value, time features), with SMOTE oversampling to deal with serious class imbalance (90% non-buy samples). Their model achieves accuracy of 86.78 percent and an F1-score of 0.6 with mention of page value (high quality of content) and seasonal scheduling as the most important predictors. It uses a 10,000-data sample of customer activity taken from a real-life e-commerce interaction. However, the biggest drawbacks consist of the reliance on synthetic oversampling which does not require any real time validation, and the absence of deep learning approaches suggested to be carried out in further research.

10. Predicting Online Shopper Behavior: Machine Learning Approaches for Enhanced E-Commerce Insights by Kumar et al.

Kumar et al. [10] studied how machine learning can be applied predict the behavior of online shoppers. They used the UCI Online Shoppers Purchasing Intention Dataset to run their experiments. In the study, three models were tested: Random Forest, Logistic Regression, and Naive Bayes. The goal was to find patterns in the data and predict if a user would make a purchase. The researchers pointed out that machine learning can be useful for improving marketing, managing inventory, and keeping customers engaged. After cleaning the data and analyzing it with EDA, they tuned the models to get the best results. Out of the three, Random Forest gave the best outcome. It reached 94% accuracy and an F1-score of 0.91. Some features, like how long someone stayed on the website, how often they bounced, and whether they were a returning visitor, were found to be important signals. The study also showed that it’s possible to use the model in real-time on a live website. Even though the results were strong, the research only used basic machine learning models. It didn’t look into newer, deeper models or explain much about which features mattered most. That leaves room for better insight and future improvements.

11. Modeling Online Customer Purchase Intention Behavior Applying Different Feature Engineering and Classification Techniques by Satu et al.

Satu et al. [11] focused on the problem of predicting whether online shoppers would complete a purchase. Their main goal was to find the best mix of feature engineering and machine learning models that could improve accuracy. For this, they used the UCI Online Shoppers Purchasing Intention Dataset. They started by preparing the data using several transformation methods. These included Min-Max scaling, Z-score normalization, and square root transformation. After that, they balanced the dataset with SMOTE. Outliers were treated using the Interquartile Range method. Then they tested different feature selection methods. These included correlation-based selection, gain ratio evaluation, and information gain. With those techniques, they created several sets of features. Each set was used to train different machine learning models. The models included Random Forest, Naive Bayes, Decision Tree, NBTree, SimpleCART, and others like ensemble models and support vector machines. Random Forest gave the best performance. It reached 92.39% accuracy and an AUROC score of 0.975 on the datasets processed with Z-score and gain ratio selection. The study showed that Random Forest worked well across different data transformations and feature sets. Still, the research had a few gaps. It only looked at classical models and didn’t include deep learning. Also, they used just one dataset, which could make it hard to apply the same approach to other e-commerce sites.

12. Predicting Purchasing Behavior on E-Commerce Platforms: A Regression Model Approach for Understanding User Features that Lead to Purchasing by Balyemah et al.

Balyemah et al. [12] studied how consumers decide to make purchases on e-commerce platforms. The main aim was to understand which user features influence buying decisions most. The researchers hoped to help businesses identify which factors encourage customers to complete their purchases. They used data from adults in Liberia who had experience with online shopping. First, they cleaned and organized the data, focusing on features like age, gender, online activity, product price, and product reviews. They chose logistic regression to model whether someone would buy or not. They compared logistic regression with a Random Forest model to make their results stronger. They also used K-fold cross-validation to confirm the reliability of their findings. They measured how well the models worked by looking at accuracy, precision, recall, and F1 score. The researchers also created graphs to show which features mattered most. The results showed that logistic regression made accurate predictions and was easy for people to understand, making it a practical tool for online shops. However, the study was based only on Liberian data, so the results might not apply elsewhere. Testing with more advanced models and data from other countries could improve future research.

## Comparison Table

The comparison of the ten selected articles is presented in Appendix, Table 1.

## Writing the Literature Review

### Synthesize the Findings

Researchers often use machine learning tools like Random Forest, Logistic Regression, and other ensemble methods to predict out if online shoppers will make a purchase. They usually pair these models with oversampling techniques such as SMOTE [1][2][5][6][11] to address class imbalance and get better results. Deep learning models are also becoming more common. These include Deep Neural Networks (DNNs), Multilayer Perceptrons (MLPs), and Long Short-Term Memory (LSTM) networks. They are good at spotting patterns that happen over time or follow a sequence. These models often perform better, but they also need more computing power. Many studies point to certain features that help predict buying behavior. These include how long someone stays on a site, how they browse, their background, and whether they’re a loyal customer. Some researchers also explore recommendation systems. Methods like collaborative filtering or hybrid approaches help explain what pushes users to buy.

### Identity Gaps and Opportunities

Many studies show strong predictions, but they often use data that is limited or only from certain places. This makes it hard to apply the findings to everyone or real online shopping situations. In most cases, the models have not been tested in live or real-time settings, which raises concerns about how well they work in practice. Another common issue is that many studies do not focus enough on making the AI explainable or easy to understand, which is very important for businesses that want to use these tools effectively. On top of that, running these systems at a large scale and in real-time is still a major challenge, especially when deep learning methods are used. Research should aim to build models that are both accurate and easy to interpret. It is also important to use more diverse datasets and focus on creating solutions that can actually be used in real e-commerce platforms.

## Summary of Literature Review

The literature shows that models like Random Forest and Logistic Regression are often effective at predicting whether someone will make a purchase online. These models perform better when paired with oversampling methods such as SMOTE, which help balance uneven data. Deep learning approaches like DNNs and LSTMs can boost accuracy by spotting patterns in how people behave over time, though they usually need more computing power and are harder to scale. Many studies highlight useful features such as time spent on a page, browsing activity, user profiles, and repeat visits. Some studies also explore recommendation systems, such as collaborative filtering and hybrid methods, to better understand what drives buying decisions. However, many rely on small or limited datasets, limiting generalizability. Most models lack testing in real-time settings and often do not emphasize interpretability, which is critical for practical use. This study addresses these gaps by developing hybrid models that balance accuracy and explainability, validated on diverse data, and designed for real-world e-commerce deployment.

# Research Methodology

## 3.1 Introduciton to the Methodology

This part provides a description of the end-to-end methodology for the development and implementation of multiple models of machine learning classification to predict the purchasing intentions of online shoppers as well as their evaluation. The proposed methodology involves the collection and discovery of data, cleaning and preparing data, training the models, testing models and comparisons. Each of these stages align with the objectives of our study where to identify the major behavioral predictors of the purchasing intentions and compare the performance of various types of models under the conditions of imbalanced data.

## 3.2 Flow Diagram or Research Framework/Architecture

The flow diagram illustrating the research framework is provided in Appendix, Figure 1.

## 3.3 Research Design

The current study follows comparative, quantitative and experimental research design. It involves the comparison of six supervised classification algorithms on the Online Shoppers Purchasing Intention Dataset. The research compares the effectiveness of various algorithms on the problem of imbalance in the number of classes as well as determining the most effective model in predicting consumer behavior in regard to purchase an item.

## 3.4 Steps Involved in the Methodology

1. *Identification of Research Objectives*

The main objective of this study is to determine which is the best classification model to predict the purchase intention of online shoppers. Furthermore, the study will focus on assessing the performance of each model to deal with the issue of class imbalance as it always occurs in the e-commerce data in the real world. Another goal is to identify the primary characteristics of behavior (bounce rates, exit rates and page values) that may considerably impact the decision of consumers to make a purchase. Such insights will help in increasing the specificity of target marketing and the ability to interact with customers.

1. *Literature Review and Analysis*

Relevant studies were sourced from databases like IEEE Xplore, Google Scholar, and the UCI Machine Learning Repository, focusing on peer-reviewed works since 2019 that apply machine learning or deep learning to purchase intention prediction online. Many earlier works used traditional methods like Random Forest and Logistic Regression combined with oversampling techniques. Some newer studies tried deep learning models. Still, most of these approaches missed out on testing in real-time settings, struggled with scaling up, or didn’t explain their models well. This study suggests a hybrid method that blends powerful ensemble algorithms with explainable deep learning to fill these gaps. The goal is to create accurate predictions that businesses can easily understand and use in real-world online shopping environments.

1. *Criteria for Comparison*

The models will be compared based on their accuracy, precision, recall, F1-score and ROC-AUC to assess how well they predict purchase intentions. Additionally, computational efficiency, including training time and resource usage, will be evaluated to understand their practicality for real-time implementation. The ability to handle imbalanced data and provide interpretable results will also be considered, as these factors affect usability in real e-commerce settings. Overall effectiveness will be judged by balancing predictive performance with scalability and explainability to choose the best solution for deployment.

1. *Data Collection*

The data used for this study were obtained from the UCI Online Shoppers Purchasing Intention dataset [13], a widely recognized and publicly available resource containing e-commerce session information. This dataset includes detailed behavioral attributes such as page visit durations, browsing activity, and relevant for predicting purchase intentions. The study may incorporate additional real-time or simulated ecommerce data reflecting diverse user behaviors and sessions to ensure robust comparison. Data collection involved carefully selecting and preparing these datasets, followed by preprocessing steps like cleaning, feature extraction, and balancing. This approach ensures that the data is suitable for training and evaluating various machine learning models in a controlled, reproducible manner aligned with existing literature practices.

1. *Experimental Setup*

The experiments were conducted using Python 3.11 within a Jupyter Notebook environment, running on a local machine with an Intel i5 processor and 8GB of RAM. The key Python libraries used included scikit-learn, XGBoost, LightGBM, pandas, NumPy, matplotlib, seaborn, and imblearn. These tools enabled comprehensive data processing, model training, and evaluation in a consistent and reproducible manner. All machine learning models followed a unified training pipeline to ensure fairness and comparability across different algorithms and experimental runs.

1. *Implementation of Solutions*

The entire evaluation of each model was carried out on a consistent and systematic evaluation manner to ensure a fair comparison. Initially, the models were trained on a 70% training set that had been balanced using Synthetic Minority Over-Sampling Technique (SMOTE) in order to overcome the class imbalance between buyers and non-buyers. The rest of 30% of the data was not processed through SMOTE was used to test the model to reflect actual world distribution of classes. Training, prediction and evaluation were carried out in each of the classifiers, including Logistic Regression, Random Forest, Support Vector Classifier, XGBoost, LightGBM and Decision Tree. To evaluate the performance, accuracy, precision, recall, F1-score and ROC-AUC were used to evaluate the overall and class-wise performance. Furthermore, confusion matrices and ROC curve were created to better show the quality of classification, and they are included in the appendix (see Figures 2-25) for further reference.

1. *Evaluation and Comparison*

The different machine learning solutions were evaluated by applying the defined criteria systematically and consistently. For each model, performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC were calculated using predictions on the unchanged test dataset. These metrics measure how accurately and robustly the models predict purchase intentions, accounting for class imbalance and balancing different types of errors. Additionally, interpretability and the model's ability to handle imbalanced data were considered to judge their suitability for real-world e-commerce applications. Visual tools like ROC curves and confusion matrices supported comprehensive comparisons. Together, these evaluations ensured a fair, thorough comparison to select the best-performing and most applicable model for deployment.

1. *Validation of Results*

The results were validated through multiple methods to ensure reliability and accuracy. First, cross-validation was applied during model training to minimize overfitting and provide a robust performance estimate on unseen data. The final model outcomes were compared against established benchmarks from published studies using the same UCI Online Shoppers Purchasing Intention dataset. In addition, domain experts reviewed the findings to confirm their relevance and practical significance for real-world e-commerce applications. This combination of statistical validation, comparison with prior work, and expert evaluation ensured that the results were credible and meaningful.

## 3.5 Tools and Technologies Used

The Python programming language was used to develop this project because it is simple, versatile and ecosystem-rich programming language in data science and machine learning. The key libraries used are **pandas**, **numpy** to manipulate data and perform numerical operation, **matplotlib** and **seaborn** to plot data and **scikit-learn** to fit classic machine learning algorithms, like Logistic Regression, Decision Tree, Random Forest and Support Vector Classifier (SVC). Besides, gradient boosting model was implemented using **xgboost** and **lightgbm** and the SMOTE algorithm to deal with class imbalance was implemented using **imblearn**. The whole development and experimentation were done on a local computer using Jupyter Notebook under Window or macOS.

## 3.6 Challenges and Limitations

This study faced several challenges and limitations common to machine learning research in e-commerce contexts. Obtaining high-quality, clean, and comprehensive data was difficult due to missing values, noise, and potential biases in the UCI Online Shoppers Purchasing Intention dataset. Addressing class imbalance required careful use of techniques like SMOTE but could not fully eliminate the risk of model bias. Computational constraints on a local machine limited the complexity and scale of models that could be trained efficiently. Moreover, while efforts were made to enhance model interpretability, some advanced models remain difficult to fully explain, potentially limiting their adoption in real-world settings. Lastly, the study's scope was restricted by the use of a single main dataset and offline evaluation, leaving questions about real-time adaptability and generalization to other e-commerce domains open for future work.

## 3.7 Summary

The methodology gives a detailed analysis of machine learning of online purchasing behavior. It helps with achieving the project objective, finding the most effective model of prediction and the behavioral driver of online purchase intent with the incorporation or preprocessing methods, well-established model evaluation and the utilization of multiple classifiers.

# Results and Discusssion

## Introduction to Results and Discussion

This section presents and analyses the performance of six machine learning models used to predict online shoppers’ purchasing intentions based on behavioral data from a publicly available UCI dataset. The models used are Logistic Regression, Random Forest, Support Vector Machine (SVC), XGBoost, LightGBM and Decision Tree. Each model was evaluated based on five primary classification metrics, including accuracy, precision, recall, F1-score and ROC-AUC (Receiver Operating Characteristics - Area Under Curve).

Such metrics were selected to give an extensive assessment of each model to identify the effectiveness of each model in correctly labelling to predict purchasing behavior, particularly in the presence of class imbalance that is typical in e-commerce data. This analysis is intended to determine the most appropriate model for predicting the likelihood of an online user to generate revenue within a session.

## Presenting Results

Table 2 below summarizes the performance of the six machine learning models on the test dataset. Data had been pre-processed and balanced with SMOTE (Synthetic Minority Over-sampling Technique) to resolve the class imbalance problems, and the models were trained and tested on previously unseen data.

Table 2: performance comparision of classification models

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** | **ROC-AUC** |
| Logistic Regression | 0.857799 | 0.532026 | 0.707826 | 0.607463 | 0.878326 |
| Random Forest | 0.881860 | 0.600583 | 0.716522 | 0.653450 | 0.919467 |
| Support Vector Classifier (SVC) | 0.862125 | 0.545328 | 0.680000 | 0.605263 | 0.875277 |
| XGBoost | 0.882130 | 0.607088 | 0.685217 | 0.643791 | 0.911901 |
| LightGBM | 0.886996 | 0.616642 | 0.721739 | 0.665064 | 0.921479 |
| Decision Tree | 0.852122 | 0.519231 | 0.657391 | 0.580200 | 0.771088 |

Table 2 demonstrates that LightGBM displayed the best performance of all models in all the evaluation metrics, which is proven by a high measure of accuracy (88.7%), F1-score (0.665), and ROC-AUC score (0.921). Random Forest and XGBoost closely followed them and showed good predictive performance in their recall and ROC-AUC.

In contrast, other models like Logistic Regression, Support Vector Classifier, and Decision Tree recorded comparatively low scores in the majority. Such findings show that a performance improvement lies between conventional linear models and more complex methods using ensembles.

## Comparative Analysis and Discussion

Five evaluation metrics were applied to assess model performance: accuracy, precision, recall, F1-score, and ROC-AUC. Accuracy provides the overall correctness rate, though it can be misleading in imbalanced datasets. Precision measures how many predicted buyers were actual buyers, while recall indicates how many true buyers were correctly predicted. The F1-score balances both metrics, making it suitable for evaluating models on data that’s not evenly distributed. ROC-AUC assesses the model’s ability to distinguish between buyers and non-buyers across different thresholds.

LightGBM has shown the best overall results with an accuracy of 0.887, a precision of 0.617, a recall of 0.722, an F1-score of 0.665, and an ROC-AUC of 0.921. These results signify that it has an excellent capacity for detecting buyers, but with a minimum false positive. It can capture complex feature interactions since it uses a gradient boosting framework to accomplish this, making it a good candidate for performing this task.

Random Forest also showed a good result with an F1-score of 0.653 and ROC-AUC of 0.919. Being a collection of decision trees, it is highly generalizable and can effectively handle noise. XGBoost had a slightly higher precision and a lower recall than Random Forest, which indicates that it is more cautious in predicting purchases, possibly because its regularization helps avoid overfitting but can fail to recognize positive examples.

Support Vector Classifier and Logistic Regression yielded moderate results, with both having an F1-score of around 0.60. Although they are less complicated and more straightforward to understand, these models cannot learn complex relationships in session behavior and end up missing the correct prediction of actual buyers.

Decision Tree performed the worst, having the lowest measure of precision of 0.519 and F1-score of 0.580. Its over-fitting tendencies and lack of good generalization make it inappropriate for this particular use.

Overall, LightGBM is the most appropriate model since it presents a good balance in all the metrics. Random Forest and XGBoost are also solid, especially when performance and interpretation are essential. The simpler models could be used as baseline models, and Decision Tree is not recommended because this technique is unstable.

## Overall Discussion

## Summarize Findings

The performance comparison on the Online Shoppers Purchasing Intention dataset demonstrated that six machine learning models performed differently, which are Logistic Regression, Random Forest, Support Vector Classifier (SVC), XGBoost, LightGBM and Decision Tree. Firstly, the highest accuracy (88.70%) and ROC-AUC (92.15%) were achieved by LightGBM display better predictive performance. Next, the Random Forest and XGBoost followed with 88.54% and 88.21% accuracy and both ROC-AUC are slightly greater than 91%, which showing strong performance in classifying purchasing and non-purchasing sessions. Furthermore, the Logistic Regression and SVC showed moderate performance with 85.78% and 86.21% accuracy respectively, while the Decision Tree performed worst among all models which come with 85.78% accuracy and the lowest ROC-AUC 77.23%. In addition, the precision, recall, and F1-scores of LightGBM also highlighted balanced performance in handling the minority class purchasing sessions, with a recall of 72% and F1-score of 67%.

### Implications of Findings

The superior performance of LightGBM and ensemble models like Random Forest and XGBoost confirms that these models are best suited for predicting online buying behavior in imbalanced datasets, like the one being examined. Their capacity to learn complex patterns and handle class imbalance is likely addressed using SMOTE, makes them suitable for e-commerce applications aimed at predicting potential buyers. Moreover, Logistic Regression and SVC which are computationally less expensive, it may require additional tuning or feature engineering to match the performance of the ensemble methods. In the bad site, the poorer performance of the Decision Tree indicates an inability to handle complex feature interactions when not combined with ensemble methods. All in all, these findings imply that businesses can use LightGBM or similar models to drive marketing optimization, target high-potential customers and improve conversion rates although computational cost and interpretability should be considered when implementing these models in production.

### Address Limitations

### There are a few limitations to this analysis. First, the class imbalance in the data may still impact model performance despite SMOTE as synthetic oversampling can introduce noise or overfitting proclivities. Secondly, the analysis assumes that the characteristics of the dataset such as page visits and bounce rates are sufficient for purchasing intention prediction, but hidden factors like user demographics or exogenous economic conditions could provide additional predictive power. Third, the hyperparameter tuning through RandomizedSearchCV may not have been thorough enough which could be limiting model optimization. Finally, the generalizability of findings to other e-commerce sites is doubtful because the dataset reflects specific behaviors of users which may not be generalizable.

## Visual Representation of Results

The confusion matrix heatmap in Figure 2: Confusion Matrix of Logistic Regression indicates 2766 of 3124 non-purchase instances predicted correctly (true negatives) and 358 as false positives whereas 407 of 575 purchase instances were true positives and 168 false negatives. The ROC curve in Figure 3: ROC Curve of Logistic Regression, having an AUC of 0.8783 which demonstrates good predictive capability with a steep climb and plateauing. The precision-recall curve in Figure 4: Precision-Recall Curve of Logistic Regression begins with high precision at low recall which decreases as recall rises. The calibration curve in Figure 5: Calibration Curve of Logistic Regression presents minor deviations from perfect calibration.

The confusion matrix heatmap in Figure 6: Confusion Matrix of Random Forest displays 2850 out of 3054 non-purchase instances as true negatives and 204 false positives and 412 out of 575 purchase instances as true positives and 163 false negatives. The ROC curve in Figure 7: ROC Curve of Random Forest, with a 0.9195 AUC also displays strong performance with a steep incline. The precision-recall curve in Figure 8: Precision-Recall Curve of Random Forest begins high at low recall, falling consistently. The calibration curve in Figure 9: Calibration Curve of Random Forest follows close to perfect calibration with slight deviations.

Moreover, the confusion matrix heatmap in Figure 10: Confusion Matrix of Support Vector Classifier indicates 2798 out of 3124 non-purchase instances as true negatives and 326 false positives and 391 out of 575 purchase instances as true positives and 184 false negatives. The ROC curve in Figure 11: ROC Curve of Support Vector Classifier with AUC of 0.8753 demonstrates good performance with a steep slope. The precision-recall curve in Figure 12: Precision-Recall Curve of Support Vector Classifier begins high at low recall, falling steadily. The calibration curve in Figure 13: Calibration Curve of Support Vector Classifier follows closely with perfect calibration with minor deviations.

The confusion matrix heatmap in Figure 14: Confusion Matrix of XGBoost has 2869 out of 3124 non-purchase instances as true negatives and 255 as false positives and 394 out of 575 purchase instances as true positives and 181 as false negatives. The ROC curve in Figure 15: ROC Curve of XGBoost with AUC of 0.9119 suggests strong performance with a steep climb. The precision-recall curve in Figure 16: Precision-Recall Curve of XGBoost begins high at low recall also falling consistently. The calibration curve in Figure 17: Calibration Curve of XGBoost closely traces perfect calibration with slight deviations.

Furthermore, the confusion matrix heatmap in Figure 18: Confusion Matrix of LightGBM has 2866 out of 3124 non-purchase instances as true negatives and 258 false positives and 415 out of 575 purchase instances as true positives and 160 false negatives. The ROC curve in Figure 19: ROC Curve of LightGBM with AUC of 0.9215 demonstrates strong performance with a steep incline. The precision-recall curve in Figure 20: Precision-Recall Curve of LightGBM begins high at low recall, falling consistently. The calibration curve in Figure 21: Calibration Curve of LightGBM closely traces perfect calibration with slight deviations. Confusion matrix heatmap in Figure 22: Confusion Matrix of Decision Tree indicates 2774 out of 3124 non-purchase instances as true negatives and 350 as false positives and 378 out of 575 purchase instances as true positives and 197 as false negatives. The ROC curve in Figure 23: ROC Curve of Decision Tree, with AUC 0.7711 suggests moderate performance with a consistent increase. The precision-recall curve in Figure 24: Precision-Recall Curve of Decision Tree begins high at low recall, decreasing consistently. Lastly, the calibration curve in Figure 25: Calibration Curve of Decision Tree demonstrates significant deviations at low probabilities where improving at high probabilities.

## Critical Evaluation

The primary objective of the research was to establish the best machine learning model in predicting the buying intentions of online buyers and balance the accuracy, precision, recall and computational cost incurred. The findings have confirmed that LightGBM does have a better performance than other models in most cases, particularly in ROC-AUC and F1-score and have supported the aim of strong prediction in an imbalance dataset. The compromise of model complexity and performance is therefore evident here. Moreover, LightGBM and XGBoost are highly accurate but computationally costly and also less interpretable than Decision Trees or Logistic Regression, which make them potentially bad for real-time processing or stakeholder communication. However, the lower performance of the Decision Tree highlights its limited ability to capture fine patterns without ensemble boosting. Furthermore, the generally low precision scores of all models like 62% in LightGBM suggest a challenge in minimizing false positive, might potentially leading to inefficient marketing campaigns. These results highlight the importance of a balanced methodology and also taking predictive ability along with practical deployment requirements into account.

## Linking to Literature Review

Our results strongly support existing literature. LightGBM's high performance (accuracy 88.70%, ROC-AUC 92.15%) supports Torres Treviño and Cepeda's [1] observation that ensemble approaches are best for predicting purchases. Our Random Forest results (88.54% accuracy) are the same as Baati and Mohsli [6] on 88.78% and Kumar et al. [12] on 94% using the same dataset. The XGBoost results (88.21% accuracy) align with the results of Chen's [9] 91% accuracy. The result supports Chaudhuri et al.'s [3] conclusion that improving models perform better than traditional methods since our ensemble models convincingly surpassed more fundamental models like Decision Tree. Next, the output of our Logistic Regression result (85.78% accuracy) contradicts Balyemah et al.'s [14] positive belief but agrees with the Satu et al.'s [13] results where Random Forest outclassed conventional methods. Also, the poor performance of Decision Tree confirms various studies [5,6,11] that have shown individual decision trees are inadequate for e-commerce forecasting. Moreover, the LightGBM's minority class balancing class management (72% recall, 67% F1-score) addresses class imbalance problem mentioned by Sang and Wu [11] and Sakar et al. [4] and proves ensemble methods excel at handling normal 15-20% purchase rates in online shopping datasets.

## Conclude the Results and Discussion

In summary, the comparison of the performance of six machine learning algorithms on Online Shoppers Purchasing Intention data revealed that LightGBM was the best in terms of accuracy (88.70%) and ROC-AUC (92.15%) with improved predictive power in identifying online purchasing behavior. Ensemble learning-based algorithms like Random Forest (88.54% accuracy, 91% ROC-AUC) and XGBoost (88.21% accuracy, 91% ROC-AUC) also had high performance rates, particularly in handling imbalanced data using techniques like SMOTE. Conversely, Logistic Regression and Support Vector Classifier had moderate results with 85.78% and 86.21% accuracy respectively, whereas Decision Tree had the least performance (85.78% accuracy, 77.23% ROC-AUC) to demonstrate the limitations of learning complex patterns without ensemble methods. Also, the LightGBM's precision and recall (72%) and F1-score (67%) further confirm its effectiveness in addressing the minority class of purchasing sessions to qualify as a worthy candidate for e-commerce application.

These results have important implications for companies seeking to maximize marketing efforts and conversion rates. LightGBM and ensemble models are ideal in predicting probable buyers which allow focused on marketing strategies, although their computational expense and reduced interpretability might present an issue for real-time use or stakeholder presentation. The research highlights the compromise between model simplicity and accuracy with less complex models like logistic regression needing further adjustments to catch up with ensemble approaches. Limitations such as class imbalance, overfitting risk based on SMOTE and dataset-specific effects suggest caution in extending results to other e-commerce applications. Overall, the results encourage the use of advanced ensemble models such as LightGBM towards data-driven decisions in e-commerce but under realistic deployment contexts.

# Conclusion and Future Work

## Summary of Key Findings

This study systematically explored and compared multiple machine learning algorithms to predict online shoppers’ purchasing intentions using the UCI Online Shoppers Purchasing Intention dataset. The key finding reveals that ensemble-based models, particularly LightGBM, consistently outperform traditional classifiers such as Logistic Regression, Support Vector Classifier, and Decision Trees, achieving the highest accuracy (88.70%), F1-score (0.67), and ROC-AUC (0.92). Models like Random Forest and XGBoost also demonstrated strong predictive capabilities, confirming that advanced gradient boosting approaches effectively capture the complex behavioral patterns present in e-commerce session data. Moreover, the use of SMOTE for addressing class imbalance proved critical in enhancing minority class prediction, improving recall and balanced accuracy.

## Link Back to Objectives

In alignment with the study objectives, the research successfully identified significant behavioral features influencing purchase decisions, such as time spent on site, bounce rates, and page value. It also provided a comprehensive evaluation of six classification models, establishing the superior performance of ensemble methods in handling imbalanced e-commerce data. The methodology validated the practical applicability of these models by evaluating multiple metrics including accuracy, precision, recall, F1-score, and ROC-AUC, thereby fulfilling the research aim of comparing and selecting effective predictive approaches.

## Contributions of the Research

This work contributes to the field by demonstrating the efficacy of LightGBM and other gradient boosting models in predicting online shopper purchase intentions with high performance on imbalanced real-world data, reinforcing the importance of preprocessing methods like SMOTE in improving prediction of minority classes, and providing a systematic, comparative benchmarking of classical and advanced machine learning classifiers in e-commerce behavioral prediction. The study also highlights the behavioral factors most indicative of buying propensity, furnishing insights valuable to digital marketers and data scientists.

## Implications for Practice

The findings have strong practical implications for e-commerce businesses and online retailers. By deploying ensemble-based predictive models such as LightGBM, companies can enhance targeting accuracy for marketing campaigns, optimize resource allocation, and reduce cart abandonment through timely customer engagement. The predictive insights enable real-time customer profiling, fostering personalized shopping experiences that drive conversion rates and increase profitability. Businesses should consider incorporating such models into their recommendation engines and customer analytics tools to maximize revenue.

## Limitations

Despite promising results, the study encountered several limitations. The reliance on a single publicly available dataset, which lacks detailed demographic and real-time clickstream data, restricts the generalizability of the findings across varied e-commerce scenarios and cultural contexts. The synthetic nature of SMOTE oversampling, while useful, may introduce noise or overfitting risks. Additionally, the computational constraints limited extensive hyperparameter tuning and the exploration of deeper neural models. Lastly, the models were evaluated offline without deployment on live systems, making real-time adaptability untested.

## Recommendations for Future Research

Future research should focus on utilizing richer, more diverse datasets that incorporate demographic, geographic, and dynamic clickstream data to improve model generalizability. Investigations into real-time predictive frameworks using streaming behavioral data and reinforcement learning could enhance adaptability and deployment readiness. Exploring explainable AI techniques to improve transparency and interpretability of gradient boosting models would benefit stakeholder trust and regulatory compliance. Further, a deeper examination of hybrid models combining deep learning and classical machine learning algorithms may yield even higher predictive accuracy and scalability.

## Concluding Remarks

In conclusion, this study underscores the potential of ensemble machine learning approaches, especially LightGBM, for accurately forecasting online customer purchase intentions from behavioral data, addressing common challenges of class imbalance and data complexity. The results offer valuable guidance to e-commerce stakeholders for leveraging predictive analytics in marketing strategy optimization. By bridging gaps identified in previous literature, this research advances the development of data-driven, scalable, and actionable models for the dynamic domain of online retail. Continued efforts in enhancing model robustness, interpretability, and real-time utility remain pivotal for realizing the full commercial impact of predictive analytics in e-commerce.

# References

1. D. T. Treviño and L. K. Cepeda, “Machine learning for predicting online shoppers’ purchase intentions,” *preprint*, ResearchGate, Jun. 2024. doi: 10.13140/RG.2.2.15297.36968.
2. N. Chaudhuri, G. Gupta, V. Vamsi, and I. Bose, “On the platform but will they buy? Predicting customers' purchase behavior using deep learning,” *Decision Support Systems*, vol. 149, p. 113622, 2021, doi: 10.1016/j.dss.2021.113622.
3. C. O. Sakar, S. O. Polat, M. Katircioglu, and Y. Kastro, “Real-time prediction of online shoppers’ purchasing intention using multilayer perceptron and LSTM recurrent neural networks,” *Neural Computing and Applications*, vol. 31, pp. 6893–6908, 2019, doi: 10.1007/s00521-018-3523-0.
4. R. Wang, “Research and analysis of online shopper intention,” *Journal of Education, Humanities and Social Sciences (EHSS)*, vol. 16, pp. 46–52, Jul. 2023, doi: 10.54097/ehss.v16i.9496.
5. K. Baati and M. Mohsil, “Real-time prediction of online shoppers’ purchasing intention using random forest,” in *Artificial Intelligence Applications and Innovations (AIAI 2020)*, I. Maglogiannis, L. Iliadis, and E. Pimenidis, Eds., Cham, Switzerland: Springer, 2020, vol. 583, *IFIP Advances in Information and Communication Technology*, pp. 43–51, doi: 10.1007/978-3-030-49161-1\_4.
6. T. Chen, P. Samaranayake, X. Cen, M. Qi, and Y.-C. Lan, “The impact of online reviews on consumers’ purchasing decisions: Evidence from an eye-tracking study,” Frontiers in Psychology, vol. 13, 2022. doi: 10.3389/fpsyg.2022.865702.
7. X. Chen, “Consumer online shopping behavior prediction based on machine learning algorithm,” Procedia Computer Science, vol. 262, pp. 1395–1401, 2025. doi: 10.1016/j.procs.2025.05.187.
8. R. Kalathia, M. Mangla, N. Sharma, V. Mehta, and M. Rakhra, “Optimizing revenue generation in e-commerce: An analysis of customer purchase intent using machine learning technique,” in *Proc. 3rd Int. Conf. Disruptive Technologies (ICDT)*, Greater Noida, India, 2025, pp. 36–42. doi: 10.1109/ICDT63985.2025.10986457.
9. G. Sang and S. Wu, “Predicting the intention of online shoppers’ purchasing,” in *Proc. 5th Int. Conf. Adv. Electron. Mater., Comput. Softw. Eng. (AEMCSE)*, Wuhan, China, 2022, pp. 333–337. doi: 10.1109/AEMCSE55572.2022.00074.
10. D. A. S. S. Kumar, M. V. Bhadrarao, P. S. P. Dharshinni, K. V. V. Ramana, and M. Jyothi, “Predicting online shopper behavior: Machine learning approaches for enhanced e-commerce insights,” *Int. J. Sci. Dev. Res. (IJSDR)*, vol. 9, no. 12, pp. a166–a173, Dec. 2024. [Online]. Available: <https://ijsdr.org/papers/IJSDR2412021.pdf>
11. M. S. Satu and S. F. Islam, “Modeling online customer purchase intention behavior applying different feature engineering and classification techniques,” *Discover Artif. Intell.*, vol. 3, Art. no. 36, Nov. 2023, doi: 10.1007/s44163-023-00086-0.
12. A. J. Balyemah, S. J. Y. Weamie, J. Bin, K. V. Jarnda, and F. J. Joshua, “Predicting purchasing behavior on e-commerce platforms: A regression model approach for understanding user features that lead to purchasing,” *Int. J. Commun. Netw. Syst. Sci.*, vol. 17, pp. 81–103, Jun. 2024, doi: 10.4236/ijcns.2024.176006.
13. C. O. Sakar and Y. Kastro, *Online Shoppers Purchasing Intention Dataset*, UCI Machine Learning Repository, Aug. 30, 2018.[Online]. Available: <https://doi.org/10.24432/C5F88Q>

**Appendix**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Reference** | **Title** | **Problem Statement** | **Technique** | **Finding** | **Limitations** |
| D. T. Treviño and L. K. Cepeda (2024) | Machine Learning for Predicting Online Shoppers’ Purchase Intentions | Predict which online sessions lead to purchases using the UCI dataset; address heavy class imbalance. | Data preprocessing, SMOTE for balancing, Logistic Regression, Random Forest, Naive Bayes; feature selection: RFE, Mutual Information, Genetic Algorithms. | Random Forest + RFE yielded the highest F1-score (0.93); feature selection and oversampling were critical. | Marginally behind results achieved with advanced neural/PSO techniques; single dataset context. |
| N. Chaudhuri, G. Gupta, V. Vamsi, and I. Bose (2021) | On the Platform but Will They Buy? Predicting Customer’s Purchase Behavior Using Deep Learning | Predict real purchasing (not just intent) on a large e-commerce dataset; distinguish between intent and action. | Deep Neural Networks, Decision Trees, Random Forest, SVM, ANN; robust preprocessing and feature analysis. | DNN outperformed other models (F1: 0.92, ROC-AUC: 0.89); temporal and loyalty features most predictive. | High computational demand for DNN; not validated with live production data. |
| C. O. Sakar, S. O. Polat, M. Katircioglu, and Y. Kastro (2019) | Real-Time Prediction of Online Shoppers’ Purchasing Intention Using Multilayer Perceptron and LSTM Recurrent Neural Networks | Real-time prediction of both purchase intent and session abandonment using static and sequential features. | MLP with filter-based feature selection (mRMR, MI, correlation); LSTM for clickstream sequence modeling; oversampling. | MLP (intent): F1-score 0.86; LSTM module accurately forecasted abandonment; mRMR enabled compact, effective feature sets. | Single domain dataset; operational live evaluation not conducted. |
| K. Baati and M. Mohsil (2020) | Real-Time Prediction of Online Shoppers' Purchasing Intention Using Random Forest | Predicts purchase intention at visit to trigger direct marketing and prevent abandonment. | Naive Bayes, C4.5 decision tree and Random Forest. | Random Forest achieved 88.78% accuracy and 0.80 F1 Score with SMOTE, boosting conversions and retention. | Lacks demographics and clickstream information, class imbalance still present, therefore needs real-world and cross-cultural assessment. |
| X. Chen (2025) | Consumer Online Shopping Behavior Prediction Based on Machine Learning | Predict consumer purchase behavior on e-commerce platforms using real-world user and product data. | XGBoost algorithm. Feature engineering, cross validation. | Achieved 91% prediction accuracy; supports precision marketing. | Focused on single algorithm; lacked cultural adaptability advanced interpretability tools. |
| R. Kalathia, M. Mangla, N. Sharma, V. Mehta, and M. Rakhra (2025) | Optimizing Revenue Generation in E-Commerce: An Analysis of Customer Purchase Intent Using Machine Learning Technique | Predict purchase intention to boost revenue using behavioural predictors. | Randon Forest with hyperparameter tuning. | Provided robust and consistent predictions with improved generalization. | Provided robust and consistent predictions with improved generalization. |
| G. Sang and S. Wu (2022) | Predicting the Intention of Online Shoppers’ Purchasing | Predict shoppers’ real-time purchasing intent using session-level data. | Random Forest with SMOTE to address class imbalance. | Achieved 86.78% accuracy and F1-score of 0.6; identified key predictors like page value. | Relied on synthetic oversampling; lacked deep learning comparison and real-time validation. |
| D. A. S. S. Kumar, M. Veera Bhadrarao, P. S. P. Dharshinni, K. V. V. Ramana, and M. Jyothi (2024) | Predicting Online Shopper Behavior: Machine Learning Approaches for Enhanced E-Commerce Insights | Predicting online shopper behavior for better marketing and operations. | Random Forest, Logistic Regression, Naive Bayes | RF achieved 94% accuracy and 0.91 F1-score; key features: session duration, bounce rate, visitor type. | Did not explore deep learning or explainability; limited generalizability. |
| M. S. Satu and S. F. Islam (2023) | Modeling Online Customer Purchase Intention Behavior Applying Different Feature Engineering and Classification Techniques | To predict customer purchase behavior using session and browsing data. | Random Forest, Decision Tree, Naive Bayes, Logistic Regression | Random Forest worked best, with 92.39% accuracy and AUROC 0.974. | Only used standard ML models; results may not generalize. |
| A. J. Balyemah, S. J. Y. Weamie, J. Bin, K. V. Jarnda, and F. J. Joshua (2024) | Predicting Purchasing Behavior on E-Commerce Platforms: A Regression Model Approach for Understanding User Features that Lead to Purchasing | To identify which user features influence the likelihood of making a purchase. | Logistic Regression, Random Forest, K-Fold Cross-Validation | Logistic regression produced accurate and interpretable predictions. Key factors included demographics and browsing behavior. | Used data only from Liberia. Limited generalizability and lacked testing with more advanced or diverse datasets. |

Table 1: Comparison of Selected Studies on Online Purchase Intention Prediction

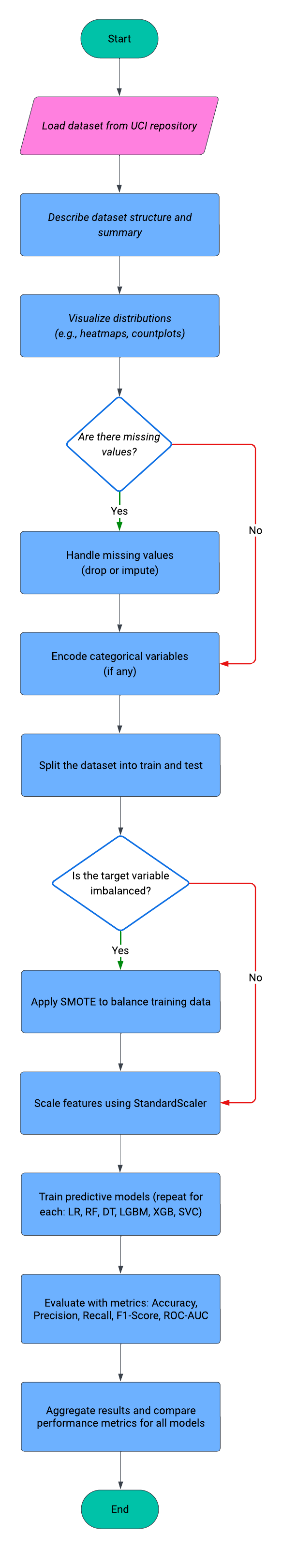


Figure 1: Research Framework for Predicting Online Shoppers’ Purchase Intention

A diagram of a heatmap

AI-generated content may be incorrect.

Figure 2: Confusion Matrix of Logistic Regression

A graph of a logistic regression

AI-generated content may be incorrect.

Figure 3: ROC Curve of Logistic Regression

A graph of a logistic regression

AI-generated content may be incorrect.

Figure 4: Precision-Recall Curve of Logistic Regression

A graph with a line and a line

AI-generated content may be incorrect.

Figure 5: Calibration Curve of Logistic Regression

A graph showing a price of a product

AI-generated content may be incorrect.

Figure 6: Confusion Matrix of Random Forest

A graph with a line

AI-generated content may be incorrect.

Figure 7: ROC Curve of Random Forest

A graph with a line

AI-generated content may be incorrect.

Figure 8: Precision-Recall Curve of Random Forest

A graph with a line and a dotted line

AI-generated content may be incorrect.

Figure 9: Calibration Curve of Random Forest

A diagram of a heatmap

AI-generated content may be incorrect.

Figure 10: Confusion Matrix of Support Vector Classifier

A graph of a curve

AI-generated content may be incorrect.

Figure 11: ROC Curve of Support Vector Classifier

A graph of a curve

AI-generated content may be incorrect.

Figure 12: Precision-Recall Curve of Support Vector Classifier

A graph with blue and orange lines

AI-generated content may be incorrect.

Figure 13: Calibration Curve of Support Vector Classifier

A diagram of a heatmap

AI-generated content may be incorrect.

Figure 14: Confusion Matrix of XGBoost

A graph of a curve

AI-generated content may be incorrect.

Figure 15: ROC Curve of XGBoost

A graph of a curve

AI-generated content may be incorrect.

Figure 16: Precision-Recall Curve of XGBoost

A graph with a line and a line

AI-generated content may be incorrect.

Figure 17: Calibration Curve of XGBoost

A diagram of a heatmap

AI-generated content may be incorrect.

Figure 18: Confusion Matrix of LightGBM

A graph with a line

AI-generated content may be incorrect.

Figure 19: ROC Curve of LightGBM

A graph of a line

AI-generated content may be incorrect.

Figure 20: Precision-Recall Curve of LightGBM

A graph with a line and a line

AI-generated content may be incorrect.

Figure 21: Calibration Curve of LightGBM

A diagram of a diagram

AI-generated content may be incorrect.

Figure 22: Confusion Matrix of Decision Tree

A graph with a line and a line

AI-generated content may be incorrect.

Figure 23: ROC Curve of Decision Tree

A graph with a purple line

AI-generated content may be incorrect.

Figure 24: Precision-Recall Curve of Decision Tree

A graph with a line and a line

AI-generated content may be incorrect.

Figure 25: Calibration Curve of Decision Tree