

Predict the online shopper's Intentions by using different type of Machine learning Algorithms

Vandana, Dr. Vishal Verma, Mr. Bajrang

¹MCA Student, ²Professor, ³Asst. Prof.

Deptt. of Comp. Sc. and Application, Chaudhary Ranbir Singh University, Jind, Haryana, India

1. Abstract

In today's online world, e-commerce plays an important role. To understand the customer behaviour is very important for business sale. This research paper concentrates to predict whether a customer will make purchase or not that based on their online activity on website. This research paper implements different type of machine learning algorithms (logistic regression, Decision tree and random forest).

By comparing these algorithms random forest gives highest prediction accuracy.

Keywords- User behaviour data, e-commerce platform and machine learning algorithms.

2. Introduction

In recent years, there has been rapid growth of e-commerce. It is a very easy way for people to shop and directly interact with businesses. Here, a large number of visitors arrive at the site, do not purchase the product, and leave. In handling these problems, machine learning plays an important role.

Predicting online shoppers' intentions has become an important area of research in the field of Machine Learning and e-commerce analytics. With the rapid growth of online shopping platforms, understanding whether a user is likely to make a purchase helps businesses improve sales strategies and customer experience. Online user behavior, such as page views, time spent, and clicks patterns, generates large amounts of data that can be analyzed for meaningful insights.

A dataset that contains various types of user behavior such as page value, bounce rate, exit rate and visitor type. In this dataset, various preprocessing steps, including duplicate removal, feature engineering, encoding, and scaling, were applied.

With the rapid growth of e-commerce, understanding customer behavioral intentions has become very difficult. Whether a user visits a site but does not purchase, the business wants to predict it.

There are three types of machine learning algorithms used:

Random Forest, Logistic regression and decision tree. By comparing these algorithms, random forest predicts highest accuracy (~90%).

3. Motivation

The main ambition of this study is to understand the online user behavior and to predict whether a customer will make purchase or not. At present, large number of users visits a site but only small

number of users complete the purchases. This is the main challenge for business; machine learning offers an effective way to instantly analyze the data and improve prediction accuracy.

Literature Review

Y. Su, X. Wang, and T. Tang (2025) focused on optimizing e-commerce platform operations using user behavior data. These strategies aim to improve the shopping experience and boost the platform's operating efficiency. Analyzing users' browsing and purchase histories helps companies recommend products they'll actually like. That, in turn, boosts conversion rates and the average amount customers spend. [1].

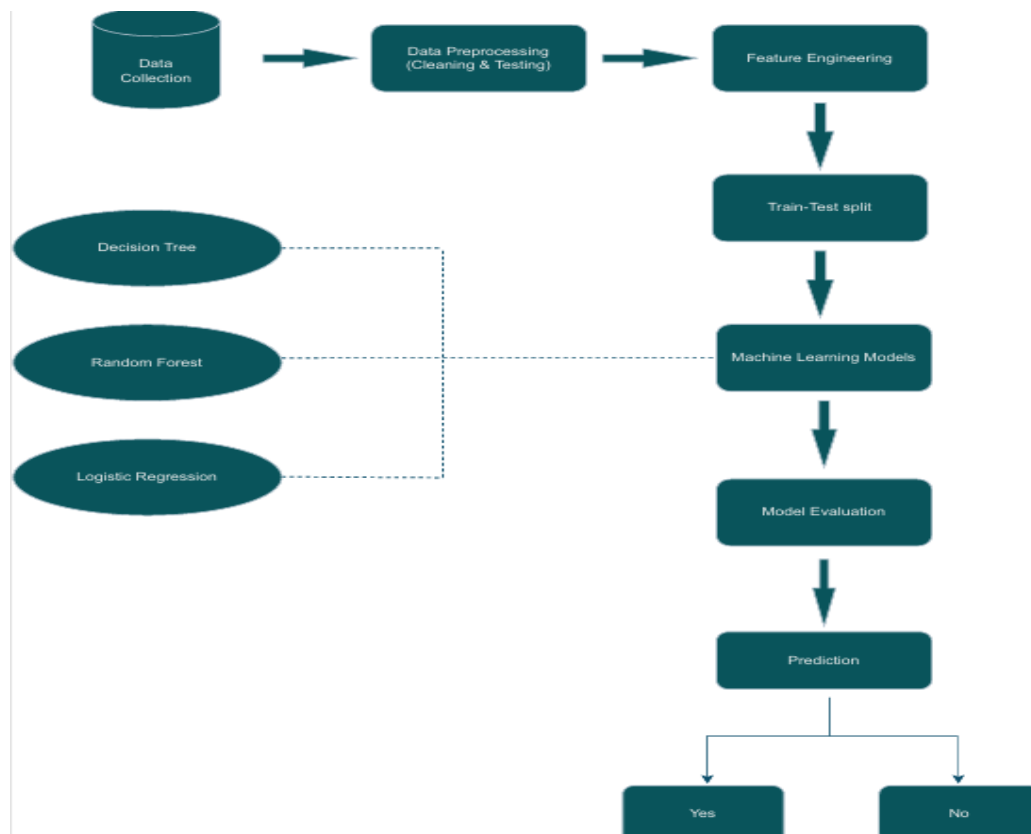
M. Gupta and T. S. R. Sai (2025) highlights the role of user interaction metrics such as session duration, page views and clicks. A user who spends more time on the site is more likely to purchase a product [2].

M. Jain (2025) explored many fundamental concepts, implementation strategies, and associated challenges. Which impacts the model's performance [3]?

J. L. M. Loureiro Abrantes et al. (2025) investigate which factors impact cart abandonment and purchasing decisions in online shopping. This research also underscores the significance of predictive models that integrate elements such as pricing, session details, and user location [4].

H. Netci et al. (2025) highlight the importance of analyzing user interaction that includes cart addition, product views and purchases. The authors evaluate various classification models. This contains four algorithms: RF, linear regression, gradient boosting, and decision trees [6].

Flow chart:



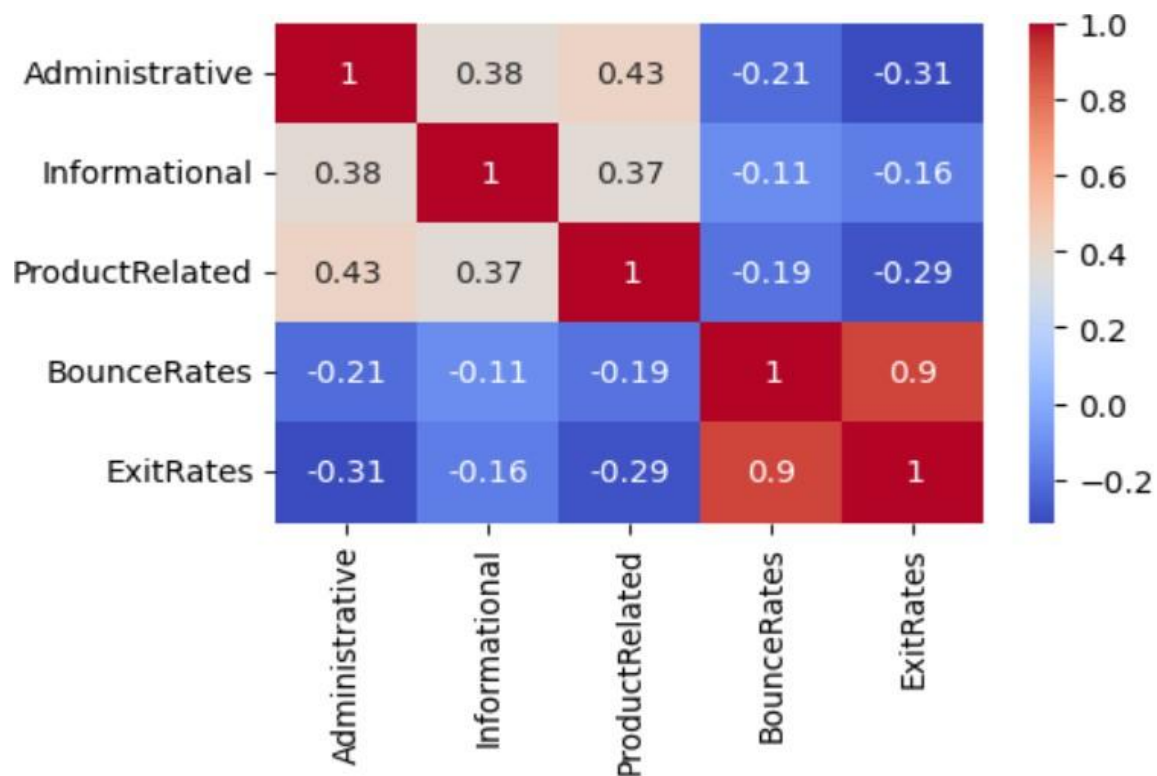
4. Methodology

4.1 Data Collection: The dataset is taken from kaggle, which includes information about users on e-commerce website. In our dataset there are 12,330 records and 18 features. The user contains different type of attributes linked to the user behavior, browsing actions and their purchasing patterns.

Revenue is the target variable, which shows that whether a user buy a product or not.

4.2 Data Preprocessing: This is the most important step in machine learning. It is used to clean our data and improve the performance of our model.

Steps include: Handling missing values, remove duplicates convert the categorical data into numerical data and feature scaling.

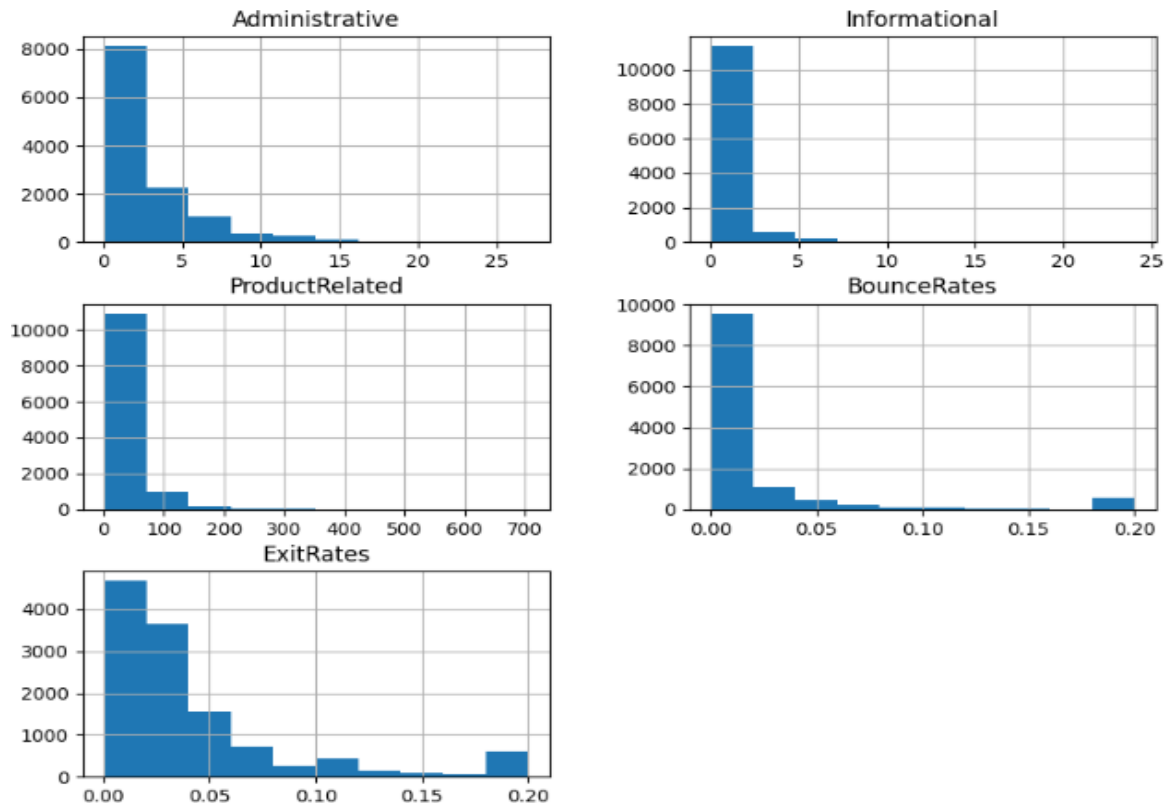


4.3 EDA (Exploratory data analysis): It is used to identify hidden patterns, to understand the structure of the dataset, find relationship between variables that help to improve performance of model.

4.4 Feature Selection: In this step, select the important feature that impact customer behavior.

Most important features in the dataset are:

- Administrative
- Informational
- Product Related
- Bounce rate
- Exit rate
- Revenue (Target variable)



4.5 Data Splitting: Data splitting is used to divide dataset into subset that is training phase and testing phase. The ratio of training and testing is in 80:20 ratio.

4.6 Research Development: There are three types of machine learning algorithms (Logistic regression, Random Forest, Decision tree) is used.

4.7 Model Evaluation: It is used to check accuracy and performance of the model. There is different type of techniques is used:

- Accuracy
- Precision
- Recall
- F1-score

5. Results:

These three models are successfully implemented on this dataset to predict the online user intension whether a customer purchase a product or not. The dataset is analysed, preprocess and used to train different type of classification model. This model is divided into two training and testing sets that is in 80:20 ratio. On this dataset there are type of machine learning algorithms is used that is Logistic Regression, Random Forest and Decision tree and evaluating using Accuracy, Recall, F1-score and Recall.

By comparing these three ML algorithms Random Forest gives best performance. By comparing these algorithms, random forest predicts highest accuracy (~90%). Random forest helps to reduce overfitting and handle complex relationship among these important features. It also helps to improve the performance and prediction reliability.

Performance Comparison:

Algorithm	Accuracy	Precision	Recall	F1-score
Logistic Regression	86.4%	84.9%	83.5%	84.2%
Decision Tree	88.3%	87.1%	86.4%	86.7%
Random Forest	90.8%	90.2%	89.6%	89.9%

6. Discussion:

By comparing these algorithms (Logistic Regression, Random Forest and Decision tree) in all evaluation metrics. This model helps to understand complex relationship and also find the relationship between the user's intention by using important features such as Exit Rate, Visitor type, Product Related pages, Bounce rate which improve the overall performance of the prediction accuracy.

7. Future Scope:

The latest research uses machine learning algorithms such as Logistic Regression, Random Forest and Decision tree. In future studies, different type of deep learning technique such as CNN (convolutional neural network) and ANN (Artificial neural network) can help to improve the performance, prediction accuracy and also help to predict the user intention.

References:

1. Y. Su, X. Wang, and T. Tang, "Research on operation optimization strategies of E-commerce platform based on user behavior data," in Proceedings of the 2025 International Conference on Artificial Intelligence and Digital Finance (AIDF 2025), Chengdu, China, June 2025. doi: 10.1145/3764727.3764758.
2. M. Gupta and T. S. R. Sai, "Integrating consumer engagement analytics and predictive modelling to enhance e-commerce decision-making and marketing strategies," Journal of Science Engineering Technology and Management Science, vol. 2, no. 12, Dec. 2025.
3. A. M. Jain, "The role of predictive analytics in e-commerce conversion rate optimization," Journal of Computer Science and Technology Studies, vol. 7, no. 2, pp. 114–121, Apr. 2025, doi: 10.32996/jcsts.2023.5.4.25.
4. J. L. M. Loureiro Abrantes, N. de L. Figueiredo, B. M. Ferreira, and L. F. Martinez, "An analysis of consumer purchase behavior following cart addition in e-commerce utilizing explainable artificial intelligence," Feb. 2025.
5. Swetha, T., Roopa, R., Sajitha, T., Vidhyashree, B., Sravani, J., & Praveen, B., "Forecasting Online Shoppers Purchase Intentions with Cat Boost Classifier," in Proceedings of the 2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT), 2024, pp. 1–6,
6. H. Netci, A. Afrig, M. Jhingga, P. Agung, and B. Sudaryatno, "Behavior-Based Purchase Intent Prediction in E-Commerce: A Machine Learning Approach," International Journal of Current Science Research and Review, vol. 8, no. 8, pp. 3970–3980, 2025,

