A Web-Based Analytics-Based Intelligent Machine Learning Framework for Forecasting Consumer Purchase Behaviour

Dr. K. Satyam[1], Kalamala Kavya[2]

[1]Associate Professor, Department of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, Andhra Pradesh, India.

[2] Post Graduate, Department of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, Andhra Pradesh, India.

Abstract

Forecasting consumer purchase behaviour has become crucial for businesses hoping to promote sales and customer engagement. This paper presents a machine learning-based method for analysing and predicting consumer purchase decisions using behavioural data. The proposed method uses factors including user visits, time spent on the platform, cart additions, and previous purchase history to calculate the likelihood of a purchase. A Gradient Boosting Classifier is employed as the primary model due to its exceptional accuracy and ability to handle complex patterns in structured data. The model is integrated into a Flask-based web application that anticipates and visualises significant influencing factors in real time using an interactive dashboard. Experiments have shown that behavioural traits, particularly cart additions and time spent, have a significant impact on purchase intent. The technology provides helpful data that businesses can utilise to improve customer happiness and develop targeted marketing strategies. This study demonstrates the value of machine learning techniques in understanding consumer behaviour and supporting data-driven decision-making**.**

Keywords

Customer purchase prediction, machine learning, gradient boosting, customer behavior analysis, web analytics, classification, Flask application, predictive modeling, feature importance, e-commerce analytics

I.Introduction

Understanding consumer behaviour has become much more important in recent years due to the explosive growth of e-commerce and internet platforms. Since this information may enhance marketing tactics, customise user experiences, and boost overall income, businesses are always looking for efficient ways to predict whether a client will make a purchase. The intricacy of contemporary user interactions may not be well captured by traditional analytical tools, which frequently rely on historical trends and basic statistical techniques. Large amounts of client data may now be analysed to find hidden patterns that affect purchasing decisions thanks to the development of machine learning. To produce accurate predictions, machine learning models can take into account a variety of behavioural indicators, including time spent on a platform, number of visits, product interactions, and past purchase history.

These models are more flexible and can keep getting better as new information comes in. This paper proposes a machine learning-based system that uses behavioural data to forecast customer purchasing decisions. Users can enter customer-related information and get real-time predictions thanks to the system's web-based application design. The Gradient Boosting technique is chosen out of all the models that were assessed because of how well it handles structured data and increases prediction accuracy. The suggested system offers insights into the significance of several elements impacting the decision-making process in addition to forecasting purchasing behaviour. This facilitates data-driven decision-making and aids companies in comprehending the critical elements influencing consumer engagement.

II. Problem Statement

Understanding and forecasting consumer purchasing behaviour is still a major challenge for firms in today's digital marketplace. Due to the effect of numerous behavioural elements, it is difficult to extract useful insights and properly forecast purchase decisions from the vast amounts of consumer contact data that are accessible. Conventional methods frequently fall short of making forecasts in real time and are unable to pinpoint the crucial elements driving consumer choices. Thus, an effective and intelligent system that can use machine learning techniques to properly forecast buy intent and analyse client behaviour is required.

III. Dataset Description

The study's dataset was created artificially to mimic actual consumer behaviour in an online retail setting. About 2000 entries with various attributes that reflect customer demographics and interaction patterns are included in the dataset.

The dataset contains the following features:

- Age – Represents the age of the customer
- Gender – Categorical feature indicating male or female
- Location – Customer's geographic region
- Device Type – Device used (mobile, desktop, etc.)
- Number of Visits – Total visits made by the customer
- Time Spent – Duration of user interaction on the platform
- Pages Viewed – Number of pages browsed
- Cart Additions – Number of items added to the cart
- Last Purchase Days Ago – Recency of previous purchase
- Discount Used – Indicates whether discounts were applied
- Customer Rating – Feedback score provided by the user

The target variable is Purchase Status, which indicates whether the customer completed a purchase (Yes/No). The dataset is balanced to ensure fair model training and evaluation.

IV. System Architecture

Three primary components make up the proposed system's modular architecture: a web-based user interface, a machine learning model, and data processing.

1. Data Layer: Manages the creation and preprocessing of datasets, including feature encoding and scaling.
2. Model Layer: Contains training and assessment of machine learning models, including Random Forest, Gradient Boosting, and Logistic Regression.

3. Application Layer: An intuitive interface for real-time predictions is connected to the trained model via a Flask-based web application.

User input is the first step in the system workflow, after which it is processed and sent to the trained model. The customer's likelihood of making a purchase is then predicted by the model, which presents the outcome along with insights.

V. Experimental Results

Several machine learning methods, such as Gradient Boosting, Random Forest, and Logistic Regression, were used to assess the suggested system's performance. The Gradient Boosting Classifier outperformed the other models in terms of feature interpretation and prediction accuracy.

An 80:20 train-test split was used to train the model using the fake dataset. The classification performance was evaluated using standard evaluation measures like accuracy, precision, recall, and F1-score.

Because of its capacity to manage intricate, non-linear correlations between customer behaviour features, the Gradient Boosting model attained the greatest accuracy.

Real-time forecasts and result visualisation were made possible by the system's further integration into a web-based interface. The anticipated purchase choice and feature importance analysis are included in the output, which improves the system's interpretability and decision-making utility.
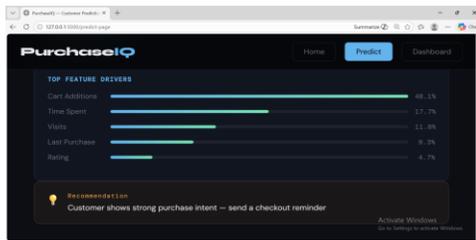
VI. Results and Analysis

Fig: Analysis Page

The findings of the experiment demonstrate how important behavioural characteristics are for forecasting consumer purchasing decisions. From the visualisation of feature importance:

• The most significant predictor was found to be cart additions (48.1%), suggesting that consumers who actively add products to their basket are quite likely to finish a transaction.

• Time Spent (17.7%) is a significant factor, indicating that increased interaction raises the likelihood of a purchase.

• Visits (11.8%) show that users are interested in and frequently use the platform.

• Last Purchase (9.3%) indicates that purchasers who have recently made a purchase are more likely to do so again.

• When compared to behavioural factors, rating (4.7%) has a comparatively smaller impact.

Based on forecasts, the algorithm also offers practical suggestions. To boost conversion rates, for example, consumers with high buy intent can be targeted with promotional offers or checkout reminders. These results show that when it comes to forecasting purchasing behaviour, user engagement indicators have a greater influence than demographic characteristics. The outcomes confirm that the Gradient Boosting model is successful in deriving significant insights from structured data.

VII. Conclusion

The suggested solution uses behavioural data analysis and machine learning approaches to accurately forecast consumer buying behaviour. The Gradient Boosting Classifier outperformed the other models in terms of accuracy and dependability. Real-time prediction and user engagement are made possible by the model's integration with a Flask-based web application. Cart additions and time spent are the most important elements in determining purchase intent, according to feature importance research. Businesses can enhance their marketing strategy and client interaction by utilising the system's actionable insights. All things considered, this work shows how machine learning may be used practically to improve decision-making in e-commerce settings.

References

[1] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 2016, pp. 785–794.

[2] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[3] D. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, Feb. 2011.

[4] M. Grinberg, Flask Web Development: Developing Web Applications with Python, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2018.

[5] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2011.