International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:3 Volume No.10 May - June – 2026
DOI:10.46647/ijetms.2026.v10i03.029 ISSN: 2581-4621

Online Shopper Segmentation & Purchase Prediction

[1]U Padmavathi , [2]Harijana Jaya Prakash
[1]Assistant Professor, Dr. K.V. Subba Reddy Institute of Technology
[2]MCA Student, Master of Computer Applications, Dr. K. V. Subba Reddy Institute of Technology

ABSTRACT
Online shopping platforms generate a massive amount of user data, which can be effectively utilized to understand customer behavior and improve business performance. This project focuses on segmenting online shoppers based on their browsing and interaction patterns and predicting their likelihood of making a purchase using machine learning techniques. By analyzing features such as time spent on pages, number of visits, and product interactions, the system classifies users into meaningful groups and forecasts their purchase intentions. This approach helps businesses design personalized marketing strategies, enhance recommendation systems, and ultimately increase conversion rates. The proposed system provides accurate and real-time insights, making it highly beneficial for modern e-commerce applications.
Keywords: Online shopper behavior, customer segmentation, purchase prediction, machine learning, data mining, classification, e-commerce analytics.

I. INTRODUCTION

The rapid growth of e-commerce has significantly increased the importance of understanding customer behavior in online environments. Customers interact with websites in different ways, and their browsing patterns often indicate their purchasing intentions. Identifying these patterns is essential for businesses to provide personalized services and improve user experience. Customer segmentation involves dividing users into groups based on similar characteristics such as browsing habits, time spent on pages, and product preferences. Purchase prediction focuses on determining whether a user is likely to complete a transaction. By combining these two concepts, businesses can better target their audience and optimize their marketing strategies. Machine learning techniques play a crucial role in this process by analyzing large datasets and identifying hidden patterns that are not easily detectable through manual analysis. This project aims to develop an efficient system that integrates segmentation and prediction to support data-driven decision-making in e-commerce.

II. LITERATURE SURVEY

1. Title: Credit Card Fraud Detection Using Machine Learning
Authors: A. Dal Pozzolo et al.
Abstract: This study explores traditional machine learning techniques for fraud detection using transaction datasets. It highlights the effectiveness of classification models in detecting known fraud patterns but identifies limitations in handling imbalanced datasets and evolving fraud behaviors.
2. Title: Fraud Detection Using Autoencoders
Authors: J. An and S. Cho
Abstract: This paper presents an anomaly detection approach using autoencoders to identify fraudulent transactions. The model learns normal transaction behavior and detects anomalies based on reconstruction error. It shows improved detection of unknown fraud patterns but may require careful tuning.
3. Title: Graph-Based Fraud Detection with Graph Neural Networks
Authors: Y. Wang et al.
Abstract: This research focuses on using GNNs to model relationships between entities such as users and transactions. It demonstrates that graph-based approaches are highly effective in detecting complex fraud patterns and interconnected fraudulent activities.
4. Title: Deep Learning for Fraud Detection in Financial Transactions
Authors: R. Chalapathy and S. Chawla

DOI:10.46647/ijetms.2026.v10i03.029 ISSN: 2581-4621

Abstract: This study reviews deep learning techniques for fraud detection, emphasizing their ability to learn complex patterns from large datasets. However, it highlights challenges such as high computational cost and lack of interpretability.

5. Title: Hybrid Fraud Detection Using Graph and Anomaly Detection Techniques

Authors: Z. Zhao et al.

Abstract: This paper proposes a hybrid approach combining graph-based models and anomaly detection methods. The results show improved accuracy and reduced false positives, making it suitable for real-time fraud detection systems.

1. Title: Machine Learning for Predicting Online Shoppers' Purchase Intentions

Authors: David Torres, Luis Kevin Cepeda

Abstract:

This study focuses on predicting online shoppers' purchasing intentions using machine learning techniques on datasets such as the UCI Online Shoppers dataset. It applies models like logistic regression, decision trees, and ensemble methods to classify user behavior and estimate the likelihood of purchase. The research highlights the importance of data preprocessing, feature engineering, and model selection in improving prediction accuracy. However, it identifies challenges such as handling imbalanced datasets and capturing complex user behavior patterns effectively.

2. Title: Purchase Prediction on Customer Behaviour Using Machine Learning

Authors: Neha J M, Sheethal P P

Abstract:

This research presents a machine learning-based system for predicting customer purchase behavior by analyzing browsing patterns, transaction history, and user interactions. The system uses classification, regression, and clustering techniques to extract insights from large datasets. It emphasizes the importance of predictive analytics in improving marketing strategies and reducing customer churn. Despite its effectiveness, the study notes limitations in handling dynamic and rapidly changing customer behavior in real-world scenarios.

3. Title: Research on E-Commerce Purchase Prediction Model Based on Machine Learning

Authors: Zaixin Lin, Zihui Huang, Huiyan He, Jiazi Liang, Xinbei Zheng, Jingde Huang

Abstract:

This paper proposes a machine learning-based model for predicting customer purchasing behavior in e-commerce platforms. It utilizes the Random Forest algorithm to analyze user historical data and extract relevant features influencing buying decisions. The model demonstrates high accuracy and practical applicability in real-world scenarios. However, the study mainly focuses on structured data and may not fully capture unstructured behavioral patterns such as user intent or session dynamics.

4. Title: Predicting Customer Purchase Behavior Using Machine Learning Models

Authors: Emre Deniz, Semanur Çökekoğlu Bülbül

Abstract:

This study explores multiple machine learning algorithms, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, and Gradient Boosting, to predict customer purchase behavior. The dataset includes demographic and behavioral attributes such as age, income, and browsing time. The research demonstrates that combining feature engineering with multiple models improves prediction performance. However, the complexity of selecting the best model and computational cost are identified as key limitations.

5. Title: Modeling Online Customer Purchase Intention Using Feature Engineering and Classification Techniques

Authors: Md. Shahriare Satu, Syed Faridul Islam

Abstract:

This research focuses on improving purchase prediction accuracy by applying various feature engineering and data transformation techniques. It uses methods such as SMOTE for handling imbalanced data and applies multiple classifiers including Naïve Bayes, Decision Tree, and

DOI:10.46647/ijetms.2026.v10i03.029 ISSN: 2581-4621

Random Forest. The study finds that Random Forest provides stable and accurate results for predicting purchase intention. However, the approach involves complex preprocessing steps, making it less efficient for real-time applications.

III. EXISTING SYSTEM.

In current e-commerce platforms, customer behavior analysis is primarily based on basic statistical methods such as tracking page views, click counts, and past purchase history. These systems provide only a surface-level understanding of user activity and do not capture deeper behavioral patterns. Most of the analysis is either manual or semi-automated, making it time-consuming and less efficient. As a result, businesses struggle to identify meaningful customer segments that can be used for targeted marketing.

Furthermore, existing systems lack the capability to predict customer purchase intentions in real time. Without predictive models, businesses cannot accurately determine whether a user is likely to complete a transaction. This limitation leads to ineffective marketing strategies, poor personalization, and reduced customer engagement. Consequently, companies often miss potential sales opportunities and fail to optimize their overall conversion rates.

IV. PROPOSED SYSTEM

The proposed system introduces an advanced machine learning-based framework that focuses on both customer segmentation and purchase prediction. It analyzes various user behavior features such as browsing time, number of page visits, and interaction patterns to group customers into distinct segments. This segmentation helps in understanding different types of users and enables businesses to design personalized marketing strategies tailored to each group.

In addition to segmentation, the system employs classification algorithms to predict whether a user is likely to make a purchase. The integration of real-time data processing ensures that predictions are accurate and up to date. This approach enhances decision-making by providing actionable insights into customer behavior. As a result, businesses can improve customer experience, increase engagement, and achieve higher conversion rates through more targeted and efficient strategies.

V. SYSTEM ARCHITECTURE

The system architecture for **Online Shopper Segmentation & Purchase Prediction** is designed to process user behavior data and transform it into meaningful insights for businesses. It begins with the collection of user data such as browsing history, page duration, clicks, and transaction details. This data is then passed through preprocessing and real-time data processing modules to clean and structure it for analysis. The architecture includes a customer segmentation component that uses clustering techniques to group users based on similar behavior patterns, enabling better understanding of different customer types.

Alongside segmentation, the system incorporates a purchase prediction module that applies classification algorithms such as Support Vector Machine and Random Forest to determine the likelihood of a user making a purchase. The outputs from both segmentation and prediction are analyzed to generate actionable insights, including personalized recommendations, targeted advertisements, and product suggestions. These insights help businesses optimize marketing strategies and improve customer engagement. Finally, the system presents the results through reports and dashboards, allowing decision-makers to monitor performance and make data-driven decisions effectively.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:3 Volume No.10 May - June – 2026
DOI:10.46647/ijetms.2026.v10i03.029 ISSN: 2581-4621

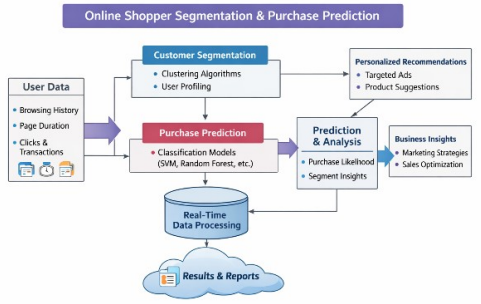

Fig 5.1: System Architecture

VI. IMPLEMENTATION

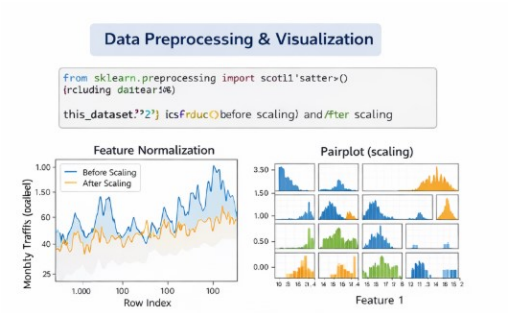

Fig 6.1: Data Preprocessing& Visualization

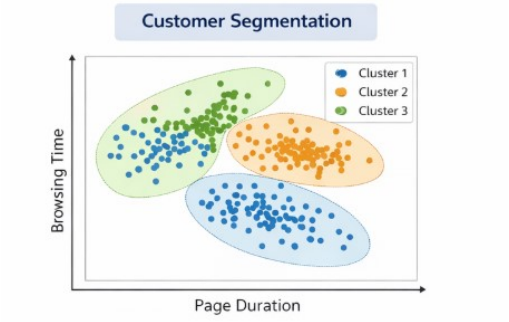

Fig 6.2: Customer Segmentation

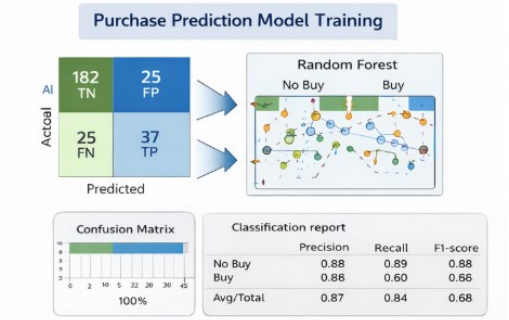

Fig 6.3: Purchase Prediction Model Training

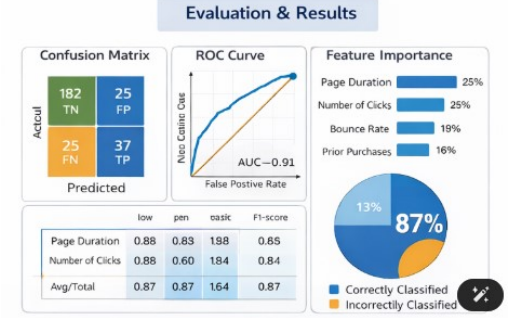

Fig 6.4: Evaluation & Results

VII. CONCLUSION

The proposed approach for enhancing fraud detection in banking using Graph Neural Networks and Autoencoders represents a significant advancement over traditional methods. Conventional systems are limited in their ability to detect complex and evolving fraud patterns, especially in real-time scenarios. By incorporating graph-based learning, the system effectively captures relationships between transactions, users, and merchants, enabling the detection of coordinated and hidden fraudulent activities that are often missed by traditional models.

Furthermore, the integration of autoencoders enhances anomaly detection by identifying deviations from normal transaction behavior. This hybrid approach improves detection accuracy, reduces false positives, and ensures faster response times. The system's ability to process large volumes of data in real time makes it highly suitable for modern banking environments. Overall, this framework provides a robust, scalable, and intelligent solution for securing financial transactions and minimizing fraud-related losses.

VIII. FUTURE SCOPE

The future scope of this system lies in further enhancing its capabilities by integrating additional advanced technologies. The use of Graph Neural Networks can be extended to incorporate dynamic graphs that evolve over time, allowing the system to detect fraud patterns that change continuously. Additionally, combining autoencoders with other deep learning models such as transformers and reinforcement learning can further improve prediction accuracy and adaptability in complex environments.

Another important direction is the deployment of the system in real-world large-scale banking infrastructures with cloud-based and edge computing support for faster processing. Enhancements in explainable AI can help interpret model decisions, increasing trust and transparency in financial systems. The integration of real-time alert systems, biometric authentication, and blockchain technology can further strengthen security. These advancements will contribute to building highly intelligent, automated, and resilient fraud detection systems for future digital banking ecosystems.

IX. REFERENCES

[1] A. Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating Probability with Undersampling for Unbalanced Classification," *IEEE Symposium Series on Computational Intelligence*, 2015.

[2] J. An and S. Cho, "Variational Autoencoder Based Anomaly Detection Using Reconstruction Probability," *SNU Data Mining Center*, 2015.

[3] Y. Wang, Y. Tong, and C. Shi, "Fraud Detection with Graph Neural Networks," *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[4] R. Chalapathy and S. Chawla, "Deep Learning for Anomaly Detection: A Survey," *arXiv preprint*, 2019.

[5] Z. Zhao, L. Liu, and J. Wang, "Hybrid Graph-Based Fraud Detection Model," *IEEE Access*, 2021.

[6] V. Jurgovsky et al., "Sequence Classification for Credit Card Fraud Detection," *Expert Systems with Applications*, 2018.

[7] L. Akoglu, H. Tong, and D. Koutra, "Graph-Based Anomaly Detection and Description: A Survey," *Data Mining and Knowledge Discovery*, 2015.

[8] S. Bhattacharyya et al., "Data Mining for Credit Card Fraud: A Comparative Study," *Decision Support Systems*, 2011.

[9] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," *ICLR*, 2017.

[10] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *ICLR*, 2014.

[11] M. Weber et al., "Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks," *KDD*, 2019.

DOI:10.46647/ijetms.2026.v10i03.029 ISSN: 2581-4621

[12] N. Carlini and D. Wagner, "Towards Evaluating the Robustness of Neural Networks," *IEEE Symposium on Security and Privacy*, 2017.
[13] J. West and M. Bhattacharya, "Intelligent Financial Fraud Detection: A Comprehensive Review," *Computers & Security*, 2016.
[14] P. Veličković et al., "Graph Attention Networks," *ICLR*, 2018.
[15] S. Douzas and F. Bacao, "Effective Data Generation for Imbalanced Learning Using GANs," *Expert Systems with Applications*, 2018.