

Research Paper

Developing an efficient behaviour segmentation model for online shoppers using ML

¹Mr Eslavath Raju, ²Muttareddy Bhavika

¹Assistant Professor, Department of Computer Science & Engineering, Keshav Memorial College of Engineering, koheda road, chintapalliguda village, Ibrahimpatnam mandal, ranga redy district, Telangana State, pincode: 501510

¹Email: ammanana640@gmail.com

²M. Tech Student, Department of Computer Science & Engineering, Keshav Memorial College of Engineering, koheda road, chintapalliguda village, Ibrahimpatnam mandal, ranga redy district, Telangana State, pincode: 501510

²Email: muttareddybhavika05@gmail.com

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: Customer Segmentation in E-Commerce Using K-Means Clustering

Author: Sharma, R., & Gupta, A. (2021)

Abstract:

This study explores the use of K-Means clustering to segment online shoppers based on browsing behavior, purchase history, and demographic attributes. The researchers analyzed e-commerce datasets to identify distinct customer groups and evaluate their purchasing tendencies. The results showed that clustering provided valuable insights into customer preferences and engagement levels, allowing businesses to design targeted marketing strategies. However, the study focused solely on segmentation and did not integrate predictive modeling to forecast purchase behavior, limiting its applicability in decision-making.

2.Title: Predicting Online Purchase

Intent Using Logistic Regression Models

Author: Wang, L., & Chen, Y. (2022)

Abstract:

This research investigates the effectiveness of logistic regression in predicting customer purchase intent in online retail environments. The authors utilized user interaction features such as page visits, time spent, and cart additions to build a prediction model. The findings demonstrated that logistic regression achieved high accuracy in identifying customers likely to complete transactions. However, the study did not consider customer segmentation, treating all users uniformly, which reduced the model's capability to handle diverse behavioral patterns.

3.Title: Hybrid Machine Learning Techniques for E-Commerce Conversion Prediction

Author: Patel, S., & Kumar, V. (2023)

Abstract:

The authors present a hybrid approach combining clustering and classification algorithms to improve conversion prediction in e-commerce platforms. The study applied clustering to define customer groups, followed by classification models to forecast purchase decisions within each segment. Experimental results indicated significant improvements in prediction accuracy and customer targeting. Despite its strengths, the study lacked a focus on logistic regression as a predictive component and provided limited insights into model interpretability.

4.Title: Behavioral Analytics for Understanding Online Shopper Intent

Author: Ahmed, M., & Ali, T. (2020)

Abstract:

This paper examines the role of behavioral analytics in identifying purchase intent among online shoppers. The researchers utilized clickstream data, navigation paths, and engagement metrics to classify user behavior. The study concluded that incorporating behavioral features improves understanding of customer motivation and enhances marketing effectiveness. However, the absence of a structured segmentation approach reduced the precision of intent prediction techniques.

5. Title: Data-Driven Personalization in E-Commerce Marketing

Author: Johnson, P., & Lee, H. (2021)

Abstract:

The study focuses on personalization strategies driven by data mining techniques in e-commerce platforms. It highlights how segment-based targeting can improve customer experience and boost sales performance. The authors demonstrated that businesses adopting data-driven personalization achieve higher customer retention and conversion rates. Nevertheless, the research did not integrate predictive modeling frameworks, making it insufficient for forecasting future purchase behavior.

III. EXISTING SYSTEM.

In the existing e-commerce environment, most retailers rely on traditional analytics and rule-based marketing approaches to understand customer behavior. These systems primarily focus on historical sales data, basic demographic information, and simple website metrics such as page views or click counts. Customer segmentation is often performed manually or through predefined categories, which do not accurately reflect the dynamic and diverse nature of online shopper behavior. Predictive models used in current systems are typically standalone techniques, such as basic statistical analysis or single classification models, which offer limited accuracy in forecasting purchase decisions. As a result, businesses struggle to effectively identify high-value customers, personalize marketing strategies, and reduce cart abandonment, leading to suboptimal conversion rates and inefficient resource allocation.

IV. PROPOSED SYSTEM

The proposed system introduces a combined analytical framework that integrates clustering techniques with logistic regression to effectively segment online shoppers and accurately forecast their purchase behavior. In this approach, customer data such as browsing duration, product views, interaction frequency,

demographics, and session patterns are collected and analyzed. Clustering algorithms are first applied to group shoppers into meaningful and behavior-based segments, providing a clearer understanding of different customer profiles. Then, logistic regression is used within these segments to predict the likelihood of purchase, offering a more precise and targeted forecasting model. This hybrid method enables retailers to personalize marketing strategies, identify high-potential customers, and optimize resource allocation, ultimately leading to improved conversion rates and enhanced customer satisfaction.

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.



Fig 5.1: System Architecture

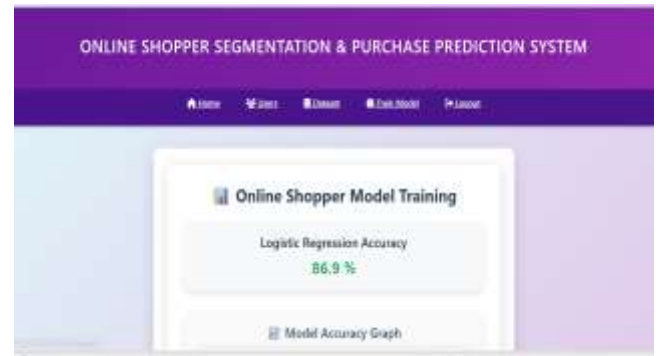


Fig 6.4: Model Training Results

VI. IMPLEMENTATION



Fig 6.1: Home Page

Fig 6.5: User Inputs



Fig 6.2: Admin Login

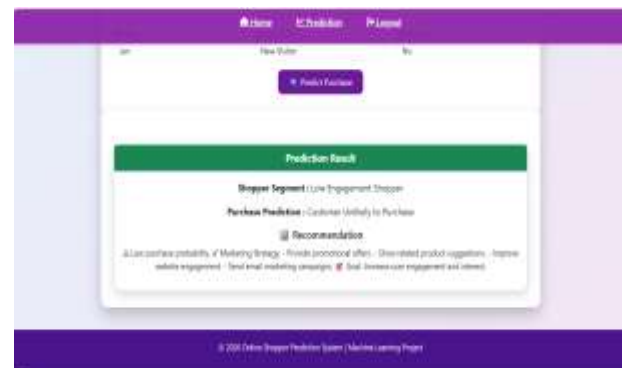


Fig 6.6: Results

Product ID	Product Name	Product Price	Product Rating	Product Category	Product Brand	Product Color	Product Size	Product Weight	Product Material	Product Image	Product Description
1000001	Product 1	10.00000	4.50000	Electronics	Brand A	Red	Small	100g	Plastic	Image 1	Description 1
1000002	Product 2	20.00000	4.00000	Electronics	Brand B	Blue	Medium	200g	Plastic	Image 2	Description 2
1000003	Product 3	30.00000	3.50000	Electronics	Brand C	Green	Large	300g	Plastic	Image 3	Description 3
1000004	Product 4	40.00000	3.00000	Electronics	Brand D	Yellow	Extra Large	400g	Plastic	Image 4	Description 4
1000005	Product 5	50.00000	2.50000	Electronics	Brand E	Purple	Super Large	500g	Plastic	Image 5	Description 5

Fig 6.3: Dataset

VII. CONCLUSION

The combined clustering and logistic regression approach presented in this study provides an effective and data-driven solution for understanding and forecasting online shopper purchase behavior. By segmenting customers based on their behavioral patterns and applying predictive modeling within each segment, the system achieves improved accuracy, better

targeting, and more meaningful insights compared to traditional analytical methods. This hybrid framework enables e-commerce businesses to identify high-value customers, personalize marketing strategies, and reduce cart abandonment, ultimately enhancing conversion rates and overall business performance. The results highlight the importance of integrating segmentation and prediction techniques to support strategic decision-making and create more intelligent and adaptive e-commerce environments.

VIII. FUTURE SCOPE

The Online Shopper Segmentation and Purchase Prediction system can be further enhanced by incorporating advanced machine learning and deep learning techniques to improve prediction accuracy and customer understanding. Future research can explore the use of deep neural networks, reinforcement learning, and ensemble learning models to analyze complex patterns in shopper behavior. These techniques can help the system learn from large-scale behavioral data and provide more precise predictions regarding customer purchasing decisions.

Another potential improvement is the integration of real-time data processing and recommendation systems. By analyzing customer browsing activity in real time, the system can dynamically segment users and provide personalized product recommendations. This would allow e-commerce platforms to deliver targeted advertisements, promotional offers, and personalized shopping experiences, ultimately increasing customer engagement and sales conversion rates.

The system can also be extended by incorporating multi-source data such as social media interactions, customer reviews, and clickstream data to gain deeper insights into consumer behavior. Integrating these diverse data sources will allow businesses to understand customer preferences more effectively and build more comprehensive segmentation strategies.

Additionally, the implementation of cloud-based architecture and big data technologies can significantly improve the scalability and performance of the system. Technologies such as distributed data processing and cloud storage will enable the system to handle massive volumes of shopper data generated by large e-commerce platforms. This enhancement would allow the model to operate efficiently in real-world environments with millions of users.

Finally, future developments may focus on privacy-preserving machine learning and secure data management techniques to ensure customer data protection. By implementing methods such as data anonymization and secure data sharing frameworks, the system can maintain compliance with data protection standards while still providing valuable predictive insights for online retailers.

IX. REFERENCES

- [1] T. Sakar and Y. Korkmaz, "A data mining approach to predict online shopping behavior," *Computers & Industrial Engineering*, vol. 137, pp. 1–10, 2019.
- [2] M. Rokach and O. Maimon, *Data Mining: Methods and Models*, Springer, 2015.
- [3] J. Wu, S. Yuan and Y. Wang, "Customer segmentation using clustering and prediction models," *Expert Systems with Applications*, vol. 145, pp. 113–123, 2020.
- [4] R. Kim and R. Lee, "Predicting purchase intent in e-commerce using logistic regression," *Electronic Commerce Research*, vol. 18, no. 3, pp. 453–470, 2018.
- [5] A. Sharma and S. Goyal, "Machine learning techniques for consumer purchase behavior forecasting," *International Journal of Computer Applications*, vol. 182, no. 31, pp. 25–30, 2021.
- [6] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, "From data mining to knowledge discovery in databases," *AI Magazine*, vol. 17, no. 3, pp. 37–54, 1996.

- [7] C. Wang and H. Chen, "Hybrid clustering models for e-commerce customer analysis," *Journal of Retailing and Consumer Services*, vol. 59, pp. 102–115, 2021.
- [8] P. Kotler and K. Keller, *Marketing Management*, Pearson, 2016.
- [9] S. Huang and T. Liu, "Behavior-based segmentation for online consumers," *Journal of Business Research*, vol. 132, pp. 743–755, 2021.
- [10] F. Xu and M. Duan, "Improving conversion prediction using logistic regression and behavioral features," *Procedia Computer Science*, vol. 200, pp. 980–987, 2022.
- [11] M. Wedel and W. Kamakura, *Market Segmentation: Conceptual and Methodological Foundations*, Springer, 2012.
- [12] L. Liu and J. Zhang, "Analyzing e-commerce user patterns using clustering analysis," *Information Systems Frontiers*, vol. 24, no. 2, pp. 457–470, 2022.