Proceedings of the International Conference on Inventive Computation Technologies (ICICT-2025)
IEEE Xplore Part Number: CFP25F70-ART; ISBN: 979-8-3315-1224-8

Online Shopping Market Analysis using CHAID Algorithm

Ms. K Saranya[1], M Issac Markson[2], S Khalith[3], M Mithun[4]
Assistant Professor[1], Student[234].
saraninnovator@gmail.com [1], kit.25.21bad021@gmail.com [2], kit.25.21bad033@gmail.com [3], kit.25.21bad029@gmail.com [4]
Department of Artificial Intelligence and Data Science, KIT-Kalaignarkarunanidhi Institute of Technology, Kannampalayam (POST), Coimbatore.

ABSTRACT:

With the modern speed of the digital marketplace, online businesses have to be aware of consumer behavior in order to compete in the online retail business. This project analyses the online shopping marketplace by applying the CHAID algorithm, determining many factors that lead towards consumer buying behavior. The analysis is carried out based on a very large dataset capturing customer demographics, browse patterns, purchase history, and preferred products. Using CHAID, this research identifies significant interactions between these variables and segmented the market into various subgroups of consumers, each coming with typical characteristics and behaviors. Such consumer groups would be very enlightening regarding the ways in which age, location, or browsing habits affect shopping behavior. Segmentation allows businesses to establish focused, personalized marketing strategies that can be related to increased customer engagement, proper product placement for each customer, leading to increased customer satisfaction and sales. Along with CHAID, other prominent advanced algorithms are SVM, ID3, GBDT, XGBoost, and many more; while the techniques of clustering aim at enhancing the accuracy of classification and forecasting. These models further refine the understanding of consumer patterns and behaviors, enabling more precise predictions of future trends. Ultimately, this project provides a blueprint for businesses to realize data-driven insights into improving online shopping experience, which impacts both customer retention and profitability.

Keywords: CHAID algorithm - Market analysis - Customer segmentation - E-commerce trends - Decision tree analysis - Data mining - Predictive analytics

I. INTRODUCTION

In such a rapidly digitizing modern economy, shopping online has become an integral part of the behavior of consumers. This has pushed businesses to be very agile in such rapidly changing landscapes. The number of transactions and browsing sessions happening every day is huge, thus making the consumer data a 'sea' that a business needs to dive into and learn from. Due to the great importance of such an interaction, a business needs this complex understanding of data about its consumers' buying behavior to develop competitive strategies, optimize its business operations, and enhance customer engagement.

Several machine learning algorithms have been applied to analyze the behavior of consumers and predict the model's outcome; among them are Support Vector Machines (SVM), Iterative Dichotomiser 3, Gradient-Based Decision Trees (GBDT), and XGBoost. None of them is fully free of drawbacks; they are computationally complex, are not easy to interpret, and generally fail to identify multivariate interactions between the variables. Classification and segmentation might be more interpretable and efficient with categorical and ordinal data by the Chi-squared Automatic Interaction Detector algorithm. This project aims to demonstrate the efficiency and accuracy of the CHAID algorithm in comparison to other machine learning classification techniques for online shopping market analysis.

CHAID is a very good decision-tree based algorithm for finding statistically significant relationships between variables and stratifying data into more or less homogeneous groups. Being applied to areas where the discovery of independent variable interactions is crucial, it can easily display multi-level splits that other methods are more likely to miss. Importance The finding of key factors driving the purchase decision of the consumer in this problem will be obtained through the use of the CHAID algorithm on a comprehensive dataset based on customer demographics, browsing patterns, product preferences, and purchase histories. Further segmentation of the online shopping market into distinct consumer groups will be used for more effective targeting of marketing strategies and related promotions.

By using comparative analysis, this project would evaluate how effective CHAID was in comparison with other classification algorithms, including its predictiveness, computational efficiency, and capacity for handling complex interactions. This should try to achieve the goal that CHAID is not only efficient in terms of segmentation and decision-making but also establish it as a power tool by optimizing marketing strategies by businesses involved in this market within the competitive online environment.

II. LITERATURE STUDY

This literature review addresses varying machine learning techniques used during the behavior analysis in online shopping, from classification algorithms that include SVM, Random Forest, and XGBoost to clustering and recommendation models. It reviews how such approaches are generally utilized in matters of predictive buying intention, recommendation personalization, and increasing customer satisfaction while somehow addressing the limitations related to scalability and precision.

Abdullah-All-Tanvir, Iftakhar Ali, et al., 2023 presented a number of machine learning models to predict the purchase intentions, including SVM, Random Forest, MLP, Decision Trees, and XGBoost. They have demonstrated a range of accuracy from 85% to 88% with an example that proved the predictive strength of these techniques. They are very complex and resource hungry when scaled up to large applications [1]. Kangming Xu et al. (2024) used BERT models and nearest neighbor algorithms to classify product categories and improve personalized recommendation. That paper looks at improving the scalability of a system so that it can dependably generate accurate and personalized recommendations for e-commerce users all over the globe via the used platforms [2]. Lijuan Xu and Xiaokun Sang (2022) developed a recommendation system using the models of GBDT, Logistic Regression, Random Forest, and XGBoost. Such a system resulted in precision rates between 0.50 and 1.00 and thus is very effective for the classification task of customers into clusters as well as would provide the target users with relevant personalized recommendations depending on the likeliness [3]. Ahmad Aldelemy and Raed A. Abd-Alhameed (2023) explored whether SVM, RBF, Multilayer Perceptron Neural Networks, and LDA models could be employed appropriately to predict customer purchasing behavior. The prediction accuracy results of developed models are moderate; between 0.600 to 0.799 precision rates, hence there is still room for improvement to a closer result [4]. Yonis Gulzar and team (2023) have used different types of clustering algorithms including K-Means as well as Fuzzy C-Means to develop recommendation systems. The primary objective behind the development of the recommendation system is to present relevant information of products to customers so that they are satisfied. Consequently, this leads to efficient recommendation systems on e-commerce websites [5]. Ganjar Alfian and Muhammad Qois Huzyan Octava combined RFID technology with machine learning models in the form of Multilayer Perceptron and iForest. They applied Adaptive Synthetic Sampling (ADASYN) to improve the accuracy of the model related to customers' behavior [6]. Ying Fei Lim et al. reviewed hybrid recommender systems that hybridize content-based and collaborative filtering techniques for reviewing. Such a model could mitigate single-method concerns of systems while improving recommendations and taking accuracy to the next level for online shoppers [7]. Li Li, Lin Yuan, and Juanjuan Tian (2023) focus on the way users' interactions influence customer satisfaction based on big data algorithms. The outcomes found in the study prove that more user interaction will aid in eliciting maximum customer satisfaction and loyalty toward an online shopping site [8]. Theresa Maria Rausch and colleagues (2020) have applied decision trees, classification, and regression models to predict and prevent abandonment at checkout. This helps determine the possible causes for cart abandonment in a manner that may help e-commerce retailers comprehend their dilemmas and have the critical opportunity to enhance their conversion and ensure that something is done to put an end to cart abandonment once and for all [9]. I-Chin Wu and Hsin-Kai Yu (2020) have used MRPs and LSA to examine the behavior of users in terms of shopping behaviors. Based on this study, MRP patterns in shopping behavior that are most important have been explained, thereby providing better insight into the shifting needs of customers over time.[10]

III. EXISTING SYSTEM

The system in place has focused efforts on classification algorithms, which are very traditional; they use algorithms such as Support Vector Machines (SVM), Random Forest (RF), K-Means Clustering, Iterative Dichotomiser 3, and Linear Discriminant Analysis.

A. Support Vector Machines (SVM):

SVM is a very strong classification technique applied to binary classification. In online shopping, SVM classifies customer behavior based on some of the features such as demographics, browsing patterns, or purchasing history. While SVM offers very strong performance, especially for high-dimensional data, it may be less efficient in bigger datasets because of the cost related to the training of the model. Moreover, in order to be optimized for various applications, SVMs often require tuning of carefully chosen parameters and selection of appropriate kernels.

B. Random Forest (RF):

Random Forest is another ensemble learning method, comprised of multiple decision trees learned to classify customer purchasing behavior by combining the predictions based on several decision trees to make a final decision. Even though Random Forest provides robustness and high accuracy, especially for complex datasets, sometimes, it's hard to interpret due to its "black-box" nature. It also requires more computational resources as the number of trees increases.

C. K-Means Clustering:

Perhaps the most widely used algorithm in the e-commerce use case for customer segmentation is the K-Means Clustering algorithm. It clusters customers into groups based on similar behavior or characteristics. One of the issues with K-Means is that it is simple and easy to implement and limits the complexity in determining the optimum number of clusters. Another problem is the discovery of a complex relationship between variables since clusters using K-Means are assumed to be spherical and equally sized, which may not be the case in real-world data.

D. Iterative Dichotomiser 3 (ID3):

It creates a decision tree algorithm that classifies data by splitting it recursively by the most informative features. In the proposed system, ID3 is applied to classify customers' behavior based on their online shopping patterns. Still, ID3 overfits in the case where data get large or is noisy; hence its overall performance is affected. Besides, it does not classify the continuous variables well and requires discretization.

E. Linear Discriminant Analysis (LDA):

LDA is a classifying method that models the relationship between classes by finding a linear combination of features which best separate the different categories. In the existing system, LDA has been applied for customers' classifications according to their purchasing habits. LDA performs well with separable data, but it fails in cases involving non-linear relationships and complex decision boundaries wherein lots of online shopping data reside.

IV. PROPOSED SYSTEM

A. Data Collection

Data accumulation is a crucial constituent of the proposed system for the analysis of online shopping behavior using

CHAID. It comprises extensive and diversified datasets from various online shopping platforms to present a comprehensive view of customer behavior. The data that will be collected includes customer demographics, such as age, gender, income level, and location, which will help in understanding the broader characteristics of the customer base. Further, the page visited, time spent on the site, and search queries of customers must be crucial in identifying any patterns of use.

The purchase history would be crucial in data collection, as it would outline the products purchased and how often they have been purchased; or, at times, the rate of abandonment in the cart. Such data represents past experiences that clearly can indicate trends and drivers of customer decisions and preferences over time. This data is also collected to gather product preference information, providing types of categories, brands, and price ranges appealing to different customer segments. Such variables are effective inputs in predicting future purchases and personalized product recommendation.

The collected data will be real time via web scraping, APIs, or directly through e-commerce platforms. This allows the data not to be out of date, but current with customer behavior. The data will then be stored in scalable cloud-based databases such as AWS RDS or Google BigQuery to address the flexibility and scalability in handling large volumes of information. The data will undergo preprocessing, which will involve dealing with missing values, outliers, and inconsistencies before applying the use of the CHAID algorithm.

Continuous variables such as age and income can be binned into categories so that it meets the requirements of the CHAID algorithm. Categorical variables will be appropriately encoded to maintain their meaning during analysis, as CHAID handles the categorical variable quite efficiently. The integrated and prepared data serve as a basis for effective segmentation and predictability of customer behavior by the CHAID algorithm.

B. CHAID Algorithm Implementation

The CHAID (Chi-Squared Automatic Interaction Detector) algorithm takes a central place in the proposed system, especially in the classification and segmentation of the online shopper in his/her behaviors while purchasing online.

CHAID is a decision tree-based algorithm that finds significant relationships between categorical variables and hence finds extreme performance in datasets containing demographic data, product preferences, and purchasing patterns. What makes CHAID different from many traditional classification algorithms, however, is that it can directly accommodate categorical variables. Accordingly, it has no need for heavy preprocessing steps such as label encoding or one-hot encoding, among others. CHAID splits the data and segmentates it using recursive chi-squared tests of independence; each split would have a statistical justification and would reveal meaningful interactions between variables. First, it begins creating the CHAID starting with the goal variable; in this case, this might be a customer's choice of purchase or type of goods to acquire. Predictor variables involve characteristics about age, gender, browsing history, or previous purchases are used to determine which factors were the most statistically significant towards determining these outcomes. The output of CHAID produces a decision tree in which every node represents a split of data based on the most relevant variable; thus, the system will be able to identify distinct customer segments with common behaviors. For example, the algorithm may establish that younger customers prefer a specific category of products, or frequent shoppers are more likely to respond to time-sensitive promotions. One of the other very important advantages of CHAID is its ability to detect multi-level interactions between variables, thus providing deeper insights into how different factors have combinations that can influence customer behavior. In addition, decision trees generated by CHAID are extremely interpretable, and businesses can easily understand results and put data-driven decisions in place. Using the strengths of CHAID, this proposed system will result in more accurate customer segmentation and better targeting of marketing strategies with personalization of experience at shop-front.

C. Real-Time Consumer Behavior Prediction

Real-time consumer behavior forecasting is one of the key functionalities that enhance the potential for classification and response to customer activity during its actual occurrence on the e-commerce website. This function aims to utilize data that is continuously captured from users as they interact with the website, such as the various types of products viewed by users, time they spend on those pages, and responses users provide in regard to recommendations. The system, therefore, will shortly identify, upon real-time analysis of these behaviors, which customer group the user falls into, by predefined segments generated using the CHAID algorithm. For instance, if one is spending a lot of their time browsing discounted products, the system will classify that customer as price-sensitive and immediately adjust the offers and product suggestions on the platform to meet this behavior. The decision tree trained on historical data with the use of the CHAID algorithm acts as the foundation for these real-time predictions. The users' interactions are analyzed against a decision tree in the context of their site engagements to predict probable outcomes such as potential purchase, cart abandonment, or product interest. This also enables the system to generate targeted recommendations and relevant promotions for the customers' time-sensitive needs, so it dramatically enhances the user experience. To do this, real-time data processing frameworks such as Apache Kafka or AWS Lambda are used to process streaming data that will invoke predictions at an instant. These technologies could be combined to allow the platform the dynamic changing of its content - particular promotions can be offered to individuals or loyalty-based discounts for regular shoppers, for example, based on expected behavior. Real-time prediction of consumer behavior increases engagement, but it also helps in business optimization efforts so marketing can be increased directly to maximize conversion. In response to the consumer's behavior real time, the system evaluates whether at all the users will be given the most appropriate shopping experience hence inducing them to buy to begin with.

D. CHAID-Driven Customer Segmentation

Customer segmentation based on the CHAID algorithm is one of the key features through which the proposed system will present an in-depth, data-driven analysis of various distinguished customer groups within an online shopping environment. In contrast, typical forms of segmentation methods frequently rely on demographic data with low complexities or trends in general behavior. On the other hand, the CHAID algorithm allows for segmenting customers based on statistically significant interactions with various multiple variables such as demographics, browsing patterns, purchase history, and product preferences. Thus, the CHAID algorithm would eventually produce a decision tree where the set of customers is split into homogeneous segments sharing common characteristics affecting their shopping behavior. For instance, the algorithm may output: young customers, who are, for example, relatively very active in searches related to fashions, are likely to make frequent purchases of trending items. Another example is the algorithm could identify a segment of low-priced item buyers-they seem to buy more during sales events. All these segments can provide actionable insight on buying behavior and help businesses to create more targeted marketing campaigns with enhanced personalization of shopping experiences. By understanding the unique traits of each grouping, firms can target their strategies specifically to address the needs and demands of each group. This may then point to greater customer satisfaction and retention rates as well as conversion rates. One of the principal advantages of CHAID-driven segmentation is interpretability. The fact that CHAID produces an intelligible as well as graphical decision tree makes it easier for marketers as well as business analysts to understand what characterizes the factors of each group. This openness makes business nimbly shift strategies based on new consumer or market behavior. In addition, because new data are continually being produced, the CHAID model- based segmentation can be updated continuously to keep business strategies appropriately aligned with the most relevant up-to-date customer insight. This dynamic, information-rich approach enhances customer engagement and drives business growth by providing a better understanding of consumer behavior.

E. CHAID Model Evaluation

One of the crucial steps of an assessment in comparison with other machine learning algorithms, especially in analyzing behaviors in online shopping, is comparing the effectiveness of the model. An evaluation will assess how efficient the CHAID model is in terms of accuracy and precision, recall, and even its computational efficiency, especially when applied in the customer segmentation and prediction of behavior related to consumers. The overall evaluation was compared on CHAID with the common algorithms used within classification tasks by traditional models such as SVM, GBDT, and Random Forest, common algorithms found in usage within any e-commerce website. Initially, the evaluation on the metrics of the CHAID model was established on precision, recall, and the F1 score as they indicate how well a model classifies distinct customer segments and predicts the behavior of customers. Precision refers to the ratio of true positives or right customers made from all the positive predictions, whereas recall refers to the ratio of actual positives correctly identified by the model. F1-score will balance out the precision and recall for a fair representation of the model's classification ability. Computational efficiency analysis is also carried out in the comparison between the CHAID in the performance of accommodating large data sets with categorical variables and the computationally expensive models like Random Forest. Against most "black box" algorithms, however, one of the great advantages in favor of CHAID is interpretability. The decision trees developed through CHAID are easy to interpret and thus enable businesses to understand various customer segments and factors behind purchasing decisions themselves. This increases the degree of transparency through which marketers and data analysts may develop their decisions. The performance metrics of the model are monitored using a visual dashboard built with tools such as Tableau or Power BI. This allows for constant evaluation and tracking of how the system performs over time, which is essential for stakeholders to change their strategies based on changing dynamics.

F. Continuous Learning and Model Improvement

Some of the facets of the continuous learning process and model improvement for keeping the CHAID algorithm effective and relevant to the proposed system for the analysis of online shopping behavior involve the update by the model on consumer preferences and the trend in the market. Adaptability to update is a value-added benefit toward accuracy and actionability. Continuous learning would be the continuous retraining of the CHAID model on new data becoming available due to continuous interaction with customers, thus incorporating new behaviors and trends in shopping. This is made easier through an automated data pipeline merging new data into the training process of the model, thus enabling the CHAID algorithm always to be aligned with the latest insights emanating from customers. The system uses cloud infrastructure like AWS or Google Cloud for the purpose of continuous learning since it offers exactly that scalability and computational resources that will come in handy when working with big data. Scheduled jobs using tools like AWS Lambda or Google Cloud Functions automate all data retrieval, preprocessing, and model retraining for a tremendous reduction in manual efforts. For example, this automation will allow the system to update the CHAID model at certain pre-set intervals or specific events, such as drastic changes in purchasing behavior when seasonal sales or other marketing activities take place. Further, model improvement is encouraged through iterative feedback loops in which the performance of the CHAID algorithm is regularly gauged against established metrics like accuracy, precision, and recall. Areas of discrepancies or improvements identified can be addressed through model tuning, be it the criteria for splitting used in decision trees or adjustment of input feature utilized for segmentation. The adaptive process does not only improve prediction of customers but also contributes towards creating a data-driven culture of decisions within the organization. Ultimately, this can make the system responsive and agile as businesses continue to appropriate the shifts in consumer behaviors and the dynamics of the marketplace.

G. System Scalability and Performance

System scalability and performance form a crucial basis in the design and implementation of the proposed CHAID-based analysis system for the online behaviors of customers.

Because the dynamics of e-commerce change, the system was designed to be scalable while large volumes of data fluctuations in user interactions are present. Then, using a cloud-based infrastructure, such as Amazon Web Services or Google Cloud Platform, it ensures that computational assets will quickly scale up and down with the demand provided. This then grants the system to be adaptable to fluctuations in data traffic, as might occur during major promotional events or the holiday shopping season, without losing speed or accuracy in processing data. From a number of strategies, the CHAID algorithm is optimized in performance. It can directly handle categorical data; therefore, it reduces the need for exhausting preprocessing steps necessary for others. This efficiency pays when working with large datasets by speeding the training and prediction process of the model. Also, a CHAID- generated decision tree is light in terms of its size. So, it requires fewer storage-related resources and the computing power as well, compared to other even more complicated ensemble methods like Random Forest or Gradient Boosting. Another important thing is that the system applies asynchronous data processing techniques that allow real-time predictions without big delays caused by this action. The system can continuously ingest customer interaction data using stream processing frameworks like Apache Kafka and immediately update recommendations and marketing offers owing to continuous analysis. The in-built performance monitoring tools in the system monitor system load, response time, and processing efficiency. This continuous supervision can help to find performance bottlenecks and provide insight into optimization to ensure responsiveness and efficiency when the volume of data and complexity of operations increase. In any case, a focus on scalability and performance will mean that such a proposed system can deliver accurate insights and recommendations in real time and therefore improve user experience as a whole in the online shopping environment.

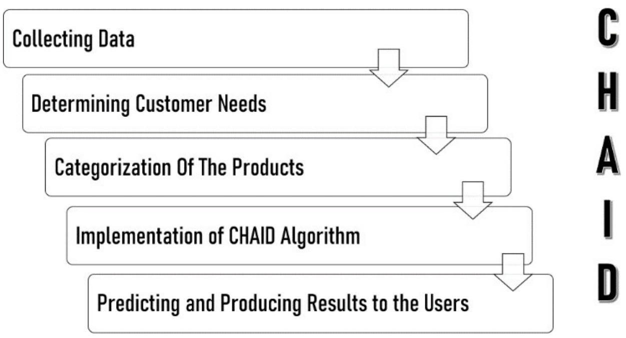

Fig. 1 CHAID Analysis

V. RESULTS & DISCUSSION

ANALYSIS RESULTS

Accuracy: 0.9637

F1-Score (for 'Yes' class): 0.6091

Precision (for 'Yes' class): 0.5154

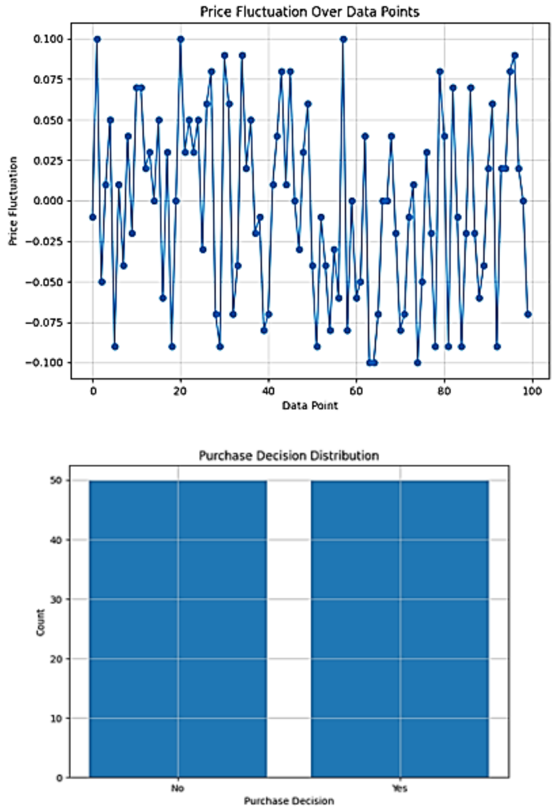

Fig 1.2 Data Analysis and Model Outcome

The overall outcome of the model using the CHAID algorithm is a high accuracy of 96.37%, which proves the model to be efficient in classifying consumer actions within the online shopping industry. Upon analyzing class-specific performance, the F1-score for the 'Yes' class, most probably capturing an important customer choice like purchase intent, was 0.609, and precision for the same class was 0.515. These numbers indicate that the model is performing well in terms of general classification but has low precision in actually classifying true positive cases. This may be a result of a class imbalance in the dataset, lack of effective feature selection, or noise in the data input. The lower value for precision means that the false positive rate is higher, and as such, the model frequently misclassified non-buy customers into buyers. This may have implications for marketing campaigns if companies base promotions on these forecasts. The decent F1-score points to the possibility of further fine-tuning, perhaps with ensemble learning methods such as XGBoost or Gradient Boosted Decision Trees (GBDT), as indicated in the literature review. Further diversifying and expanding training data may also assist in generalizing patterns for the model.

In spite of such restrictions, the analysis obtained from CHAID segmentation may prove useful for e-commerce companies. By integrating the model's output into an interface embedded in the web, stakeholders are able to observe customer behaviors and continue to optimize their strategies for personalized recommendations, product placement optimization, and targeted advertising. In the future, real-time analytics and deep learning-based methods could further improve the accuracy and reliability of the system, rendering it a better system of online shopping market analysis.

VI. CONCLUSION

This paper proposes an integrated system for the prediction and analysis of online shopping behaviors using the CHAID algorithm. As CHAID uses categorical data well in terms of discovering multi-level relationships, it is more accurate and understandable in the aspect of customer segmentation compared with traditional ML models. This will ensure that the system is response both to the immediate consumer behavior and evolves over time with market trends and customer preferences. It is highly scalable and can be efficient enough in terms of dealing with large and dynamic datasets, more characteristic of e-commerce. Using cloud-based infrastructure, automated data pipelines, and real-time processing frameworks, it is ensured that it remains uninterruptible in high traffic times, for example, when there are promotional offers. The continuous learning and model improvement components ensure the CHAID model stays relevant, giving its business users actionable insights on how to improve their marketing efforts and customer engagement. Overall, this project demonstrates huge value addition of compositing the high-quality data-driven analysis within machine learning techniques toward improving decision- making in the environment surrounding online shopping. In addition to refining customer segmentation and providing much-needed personalized experiences, the system complements the growth in conversion rates and sales along with the improvement of customer satisfaction. Such an approach is quite apt and equips businesses with more than just the inherent tools to stand against the dynamic fluctuations of the digital marketplace. This project holds great promise to shape how business will be conducted over the Internet by online retailers, who could be able to connect in a richer and more individually tailored way with their customers.

REFERENCES

[1] Abdullah-All-Tanvir, Ali, I., Khandokar, A.K.M., Islam, M., Islam, S., & Shatabda, S. (2023). A gradient boosting classifier for purchase intention prediction of online shoppers. Heliyon, 9, e15163.

[2] Xu, K., Zhou, H., Zheng, H., Zhu, M., & Xin, Q. (2024).
Intelligent classification and personalized recommendation of e- commerce products based on machine learning. ArXiv.

[2] Xu, L., & Sang, X. (2022). E-commerce online shopping platform recommendation model based on integrated personalized recommendation. Hindawi Scientific Programming, Volume 2022.

[3] Aldelemy, A., & Abd-Alhameed, R. (2023). Binary classification of customer's online purchasing behavior using machine learning. Journal of Techniques, 5(2).

[4] Gulzar, Y., Alwan, A. A., Radhwan, M., & Oumrani, M. (2023). OCA: Ordered clustering-based algorithm for e-commerce recommendation system. Sustainability, 15(2947).

[5] Alfian, G., & Octava, M. Q. H. (2023). Analysis of customer shopping behavior using RFID and machine learning models. Information, 14(551).

[6] Lim, Y. F., Haw, S. C., Ng, K. W., & Anaam, E. A. (2023). Hybrid-based recommender system for online shopping: A review. Journal of Engineering Technology and Applied Physics, 2023.

[7] Li, L., Yuan, L., & Tian, J. (2023). Influence of online e-commerce interaction on consumer satisfaction based on big data algorithm. Heliyon, 9, e18322.

[8] Rausch, T. M., Derra, N. D., & Wolf, L. (2020). Predicting online shopping cart abandonment with machine learning approaches. International Journal of Market Research, 2020.

[10] Wu, I.-C., & Yu, H.-K. (2020). Sequential analysis and clustering to examine online shopping intentions based on need states. Information Processing & Management, 57, 2020.