WCRG Global Research
WORLD COMET RESEARCH GROUP

International Journal of Modern Innovations and Emerging Trends
Vol. 1, No. 2, 2018
Doi: 10.67228/30715628/IJMIET-2018PIIJ6M3
PP. 01-15

Original Article

AI-Driven Customer Behavior Analysis in E-Commerce Platforms

Dr. Rajesh Kumar Sharma

Professor, University of Delhi, India.

Received: 08-04-2018 Revised: 08-25-2018 Accepted: 09-01-2018 Published: 09-04-2018

ABSTRACT

The rapid growth of e-commerce has generated massive customer interaction data, creating opportunities for intelligent systems to analyze and predict consumer behavior. Traditional analytical methods often struggle to capture the complex and nonlinear nature of customer decision-making. In this context, Artificial Intelligence (AI) has emerged as a transformative tool, enabling advanced behavioral analytics through machine learning, data mining, and predictive modeling. This paper explores AI-based customer behavior analysis in e-commerce, focusing on improving personalization, recommendation systems, and customer retention. It utilizes techniques such as supervised learning, unsupervised clustering, and deep learning to analyze purchasing patterns, browsing behavior, and customer segmentation. The proposed framework integrates multi-source data, including clickstream data, transaction history, demographic details, and product interactions. By applying feature engineering and behavioral modeling, the system identifies hidden patterns in customer decisions. Predictive models like logistic regression, decision trees, and neural networks are used to forecast purchasing and churn behavior. A key contribution is the hybrid architecture combining clustering and classification methods, enabling better customer profiling and adaptive response systems.The study emphasizes the importance of data preprocessing—such as normalization, noise removal, and handling missing values—for improving model performance. Results show that AI-based models outperform traditional statistical approaches in accuracy, scalability, and adaptability. Additionally, AI enhances customer experience through personalized recommendations and targeted marketing. Overall, the paper highlights AI as a powerful enabler in modern e-commerce analytics, offering scalable and efficient solutions for understanding complex customer behavior. Future work may focus on real-time analytics and integrating reinforcement learning to further improve customer engagement.

KEYWORDS

Artificial Intelligence, E-Commerce Analytics, Customer Behavior Analysis, Machine Learning, Predictive Modeling, Recommendation Systems, Data Mining, Customer Segmentation.

ISSN: 3071-5628
PP: 01-15.

1. INTRODUCTION

1.1. Background

The high growth in the field of digital technologies has essentially altered the context of business transactions, where e-commerce platform has presented as the prevailing and very competitive business environment. As the world continues to go online, use of mobile gadgets and cloud computing continues to increase, businesses are currently being conducted in environment where high volumes of information are constantly being produced through customer contacts. This data contains structured data (records of transactions and user-profiles) and unstructured data (product reviews, social media feedbacks, social ping lists, browsing history and click streams data). The analysis of such heterogeneous and large amount of data has become fundamental in organizations those who want to know what customers prefer, to improve user experience and to optimize the information in marketing and sales. The ability to properly utilize such data can help businesses to offer personalized recommendations, enhance customer experiences, and ensure they gain competitive advantages in the market. Nevertheless, the classical methods of analysis, mostly based on statistical techniques, are considerably restricted with the scope and the volumes of the contemporary e-commerce data. Such approaches tend to be linearistic and have difficulty in reflecting the dynamic nature of customer behaviour as well as nonlinearity. The growing complexity and contextuality of consumer interactions necessitate the growing need to have smarter systems that can learn automatically by examining data and perceiving underlying patterns and change at specific trends. Here, artificial intelligence and machine learning have become one of the most influential resources that can cope with these issues, allowing making predictions more accurately and make better decisions based on the data in e-commerce situations.

1.2. Role of Artificial Intelligence in E-Commerce

Artificial Intelligence (AI) can change the face of e-commerce as it is a facilitating technology that brings new computational frameworks capable of processing large volumes of customer information and identifying hidden trends. Through machine learning algorithms, systems have the capability of learning by engaging the past information, making them better over time without a formal program reference. Such an ability provides predictive analytics, which is used to predict the future behavior of a customer as well as prescriptive analytics which proposes optimal business actions. Consequently, AI also enables e-commerce platforms to increase operational efficiency and customer experience and make real-time data-driven decisions.

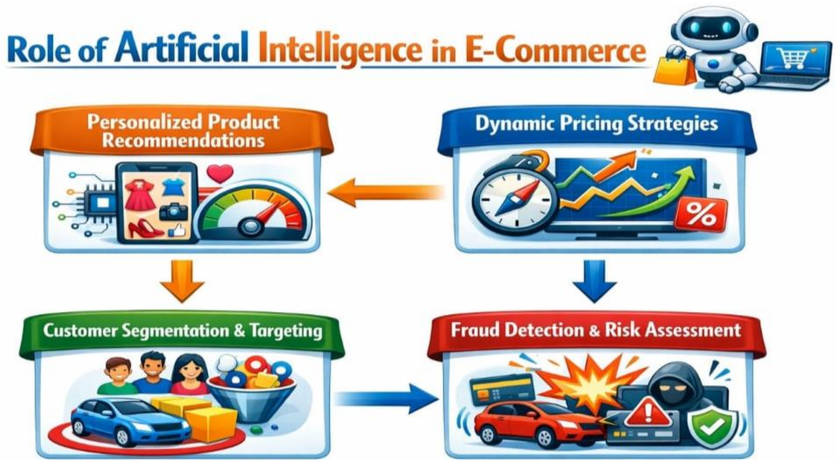

Fig 1 - Role of Artificial Intelligence in E-Commerce

ISSN: 3071-5628
PP: 01-15.

1.2.1. Personalized Product Recommendations

AI-based personalized recommendation systems use the behavior and preferences of people and prior interaction to recommend suitable products to different customers. These systems use collaborative filtering, content-based filtering and deep learning techniques to give personalized recommendations that suit customer interests. Individualized suggestions do not only improve a shopping experience but also elevate the customer engagement, the rate of conversion, and sales at large as they display to users those products that they have more chances of buying.

1.2.2. Dynamic Pricing Strategies

Dynamic pricing is an AI-based strategy that can change the pricing of products in real-time depending on such factors as demand, competition, customer behavior, or conditions in the market. The machine learning model is used to grade the historical sales data and outer variables in a manner that it will arrive at the best pricing model that maximizes the sales and competitiveness. Such a strategy can enable e-commerce companies to react to the changes in the market rapidly, provide highly personalized discounts, and use a surge pricing strategy in case of high demand, which should eventually enhance profitability.

1.2.3. Customer Segmentation and Targeting

AI makes it possible to do superior customer profiling because it allows subdivision of the customers according to their behavior, likes and preferences their demographics and their buying habits. Machine learning algorithms will be able to identify specific customer segments and unearth concealed connections in the data. The segmentation enables enterprises to create focused marketing programs, to customize messages, and provide customers with the appropriate content to the intended groups. Consequently, businesses will be able to retain customers better and gain more interaction with them, as well as acquire greater marketing efficiency.

1.2.4. Fraud Detection and Risk Assessment

In the context of securing e-commerce sites, AI is of crucial importance in preventing fraud and in evaluating risks in real time. Machine learning models are used to examine the patterns of transactions, behavior of consumers, and anomalies to discover suspicious activity in an unauthorized transaction or account takeover. These systems constantly learn new information making them enhance their capability of detecting emerging fraud trends. Determining fraud is effective to prevent businesses and most importantly customer confidence through at least assurance of a secure shopping environment.

1.3. AI-Driven Customer Behavior Analysis in E-Commerce Platforms

The analysis of customer behavior with the assistance of AI has already become a very important part of the contemporary e-commerce platforms helping companies obtain new insights into the interaction of customers with the online market. Through the use of superior machine learning and data mining systems, e-commerce systems are able to handle large volumes of data due to customer usage, such as browsing behaviors, click trails, order history, search history, and product reviews. Such insights enable the businesses to know what the customers are purchasing but also why they make some choices and hence be able to make informed and strategic decisions. It is because AI models can find making predictions related to customer preferences, purchase intentions, and subsequent behavior, hidden patterns and correlations in this data can be identified, and accurate predictions are possible. An example of this is predictive analytics to predict customer needs and recommendation systems to provide them with product suggestions that are unique to their needs in real-time, improving user experience and boosting conversion rates. Furthermore, user

ISSN: 3071-5628
PP: 01-15.

behavior and demographics can be analyzed using AI and used to divide customers into categories to target their marketing and engagement approaches. In comparison to conventional resources that have historically been used to analyze data, AI systems can operate with constant learning, and adjusting to new information, which is why it is highly useful in unstable and rapidly changing online shops. Such flexibility helps the businesses to be quick at responding to revising customer trends and economic circumstances. Moreover, AI aids in automation during decision making and human intervention may be minimized, as well as operational efficiency is enhanced. Nevertheless, issues like privacy of data, the transparency of models, and the ethical aspects also should be tackled to make sure that AI technologies are used responsibly. All in all, customer behavior analysis by AI enables e-commerce internet stores to provide personalized, more efficient, and data-driven services to customers, which ultimately increases customer satisfaction and business performance.

2. LITERATURE SURVEY

2.1. Early Approaches to Customer Behavior Analysis

The first methods of customer behavior analysis used were mainly based on the classical methods of statistics and rule-based systems to analyze the consumer data. Several methods were popularly used to reveal association between the purchasing variables with customer demographics and purchase patterns, including regression (and logistic regression) and association rule mining (e.g., Apriori algorithm). These tools allowed companies to do simple customer segmentation, demand forecasting and market basket. As an example, association rules were used to find out some of the commonly purchased products hence assisting in cross-selling. Nevertheless, the techniques relied mainly on structured sets of data and presetting assumptions making them rather inflexible. They had a hard time keeping up with the growing amount, pace, and diversification of data that the current e-commerce websites produce. Besides, they could not model nonlinear relationships and dynamically behave as a user making them less predictive in complex settings.

2.2. Machine Learning in E-Commerce

Introduction of machine learning became the breakthrough in the field of customer analysis behavior as it is able to pick up patterns out of data without need to write about the programs. We used supervised learning algorithms like decision trees, random forest and support vector machines (SVM) in the classification tasks such as customer churn prediction and purchase intention analysis. In the meantime, the k-means clustering and hierarchical clustering (unsupervised type of learning) became popular in the context of customer segmentation, and it enabled companies to cluster users according to observed behavior patterns. These frameworks showed enhanced scalability and flexibility over the traditional approaches. Also, machine learning made it easy to do feature engineering and automatic pattern recognition and made predictions more accurate. Nevertheless, problems of overfitting, data imbalance and the necessity of large labeled datasets remained. Moreover, in spite of the fact that machine learning models enhanced operations, they were not always transparent, and it was *not* an easy task to explain to businesses the rationale behind predictions.

2.3. Recommendation Systems

Recommendation systems have become an agency of the contemporary e-commerce systems and a crucial determinant in customer interactions and sales. The previous systems used were based on collaborative filtering which uses the user item interactions to propose the products based on the similar interests of the users. Content-based filtering, however, emphasizes on item properties and user profile to developed personal recommendations. Although the two methods have their advantages, they also have their downs which include cold-start problem and sparsity of data. In

ISSN: 3071-5628
PP: 01-15.

order to resolve these limitations, there has been a development of hybrid recommendation systems, which integrates both collaborative and content-based recommendation methods to enhance the accuracy and strengths in recommendation. The sophisticated techniques also involve the use of matrix factorization and neighborhood models in order to increase the scalability. The e-commerce giant organizations have embraced the use of these systems to deliver shopping experiences that are personalized. Nonetheless, they have been successful, but there are still issues of carrying out diversity in recommendations, preventing excessive specialization, and guaranteeing real-time flexibility.

2.4. Deep Learning Techniques

Current developments in artificial intelligence have seen the introduction of deep learning algorithms in the analysis of customer behaviour in e-communication. Transformers, convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep neural networks have been shown to be superior with respect to high-dimensional and unstructured data that include images, text, and clickstream logs among other types. Such models are able to model complex and nonlinear relationships and time-dependencies on user behavior and therefore, make more correct predictions and customized recommendations. As an example, RNNs and long short-term memory (LSTM) networks are especially suitable at encoding sequential user contacts, whereas CNNs are applied to the recommendation of visual products. Deep learning is also end-to-end learning, which does not require manual feature engineering. Nevertheless, the models are expensive in terms of computational capabilities and training on large datasets. What is more, their black-box nature puts into question interpretability and trust, particularly in the context of making important decisions in business.

2.5. Research Gaps

Although the analysis of customer behavior has achieved extensive progress, there still remain a number of gaps in the research that are inhibiting the potential of AI-driven e-commerce systems. Real-time data processing is one of the greatest challenges because most of the currently available models do not focus on the processing of streaming information and the provision of prompt insights. The other important problem is the interpretability of the model, in which by using complex machine learning and deep learning models, the generated models are not very transparent and as a result, stakeholders cannot easily interpret the generated predictions and trust them. Assimilation of multi-source data, structured, semi-structured, and unstructured data among different platforms is a complicated undertaking because of the deviation and data heterogeneity. Also, the performance of recommendation systems is still influenced by the necessity to deal with sparse datasets, in particular, when new customers or products appear. To mitigate such challenges, more effective, explicable, and scalable models that can be run in a dynamic, real world setting are needed.

ISSN: 3071-5628
PP: 01-15.

3. METHODOLOGY

3.1. System Architecture

Fig 2 - System Architecture

3.1.1. Data Collection

The first step of the proposed system is the process of collecting the data pertaining to customers, which will involve acquiring data on the information related to customers in various areas within the e-commerce system. Those are transactional records (purchase history, cart activity), user interaction records (clickstreams, browsing behavior), demographic records and product metadata. To provide a more enriching customer profile, it is also possible that the data can be gathered publicly e.g. with the help of social media or third-party APIs. At this point in time, data quality, consistency and completeness is a critical area of consideration in that it dictates the efficiency of further analytical operations.

3.1.2. Data Preprocessing

Raw data is processed to create structured and usable data to be analyzed in data preprocessing. This step involves managing any missing and duplicate values, fixing disjointed agreements and normalizing or standardization of numerical variables. One-hot encoding and label encoding are known as methods of encoding categorical variables. Also, noise minimisation and outlier detection are carried out in order to enhance the reliability of the data. Preprocessing also involves optimization of the dataset to be used with machine learning algorithms hence improving the accuracy and performance of the models used.

3.1.3. Feature Engineering

The process of feature engineering is concerned with the goal of bringing some useful features of the processed data to enhance the process of model learning. This involves the creation of new features like customer lifetime value, frequency of purchase, recency measures and behavior patterns. In order to remove irrelevant/redundant variables, techniques such as dimensionality reduction (e.g., PCA) and feature selection are used. The successful feature engineering is capable of the ability to capture the concealed associations among the data and enhances the predictive potential of the model considerably.

ISSN: 3071-5628
PP: 01-15.

3.1.4. Model Training

Model training consists of a choice of suitable machine learning or deep learning algorithms and aiming them to the ready dataset. Models used can include a decision tree, random forests, support vector machines, or neural networks among others depending on the task being performed. The data is usually divided into training and validation sets in order to assess training performance. The model performance is optimized and overfitting is avoided with the help of hyperparameter tuning and cross-validation. This step provides the system with a chance to identify patterns and relationship in customer data.

3.1.5. Prediction and Evaluation

During this stage, the model is trained and it is applied to predict unknown data, which includes making predictions like customer purchase behavior or product recommendation. Accuracy, precision, recall, F1-score, or RMSE are some of the measures used to evaluate the performance of the model depending on the type of the problem. Evaluation aids in determining the capability of the model in question in terms of effectiveness and generalisation. Depending on the findings, additional improvements can be applied to enhance the performance. The step guarantees that the system provides accurate and practical information on decision-making.

3.2. Data Preprocessing

3.2.1. Handling Missing Values

The treatment of missing values forms a critical process in data preprocessing because missing data may adversely affect the results of the model and consequently give biased outcomes. The absence of values can be associated with typing wrongs or system malfunction and missing user information. The simple ways to solve this problem are deletion (removing records which have missing values), mean/median/mode imputation and more complex methods such as k-nearest neighborhoods (KNN) imputation or regression-based imputation. The method selected is based on the nature and extent of missing data. Good management will guarantee the safety of the data and avoid loss of important information.

3.2.2. Data Normalization

The normalization of data involves altering the numerical characteristics to a common magical power without adulterating the relationship of these numbers. This is specifically so when the features assume varied ranges as it would guarantee that no single feature controls the model with the magnitude that it has. The Min-Max normalization and Z-score normalization techniques are very popular. Optimization has been shown to accelerate convergence of machine learning algorithms and also increases stability of the model and in particular distance based algorithms such as k-means clustering, k-nearest neighbors.

3.2.3. Outlier Detection

Outlier detection refers to defining and controlling the data sets that are substantially different than the rest of the dataset. Such anomalies may result because of mistakes, some special events, or abnormal customer behaviors. Some of the widely used methods to recognize outliers are statistical methods (e.g., Z-score, IQR), clustering based, and visualization (e.g., box plots). The outliers can either be removed or treated once they are identified based on its relevance. Good outlier control is used to increase the models accuracy and to avoid skewed results.

3.2.4. Feature Scaling

The issue of feature scaling is to make all the input variables equally contribute to the model by scaling them to similar extent. In contrast to normalization where the data are rescaled into a fixed range, the scaling techniques like standardization (mean = 0, standard deviation = 1) alter the data depending on its analytical data characteristics. Algorithms that use distances such as support vector machines and gradient descent-based models require feature scaling. It improves computational performance and guarantees the accelerated convergence during model training.

3.3. Feature Engineering

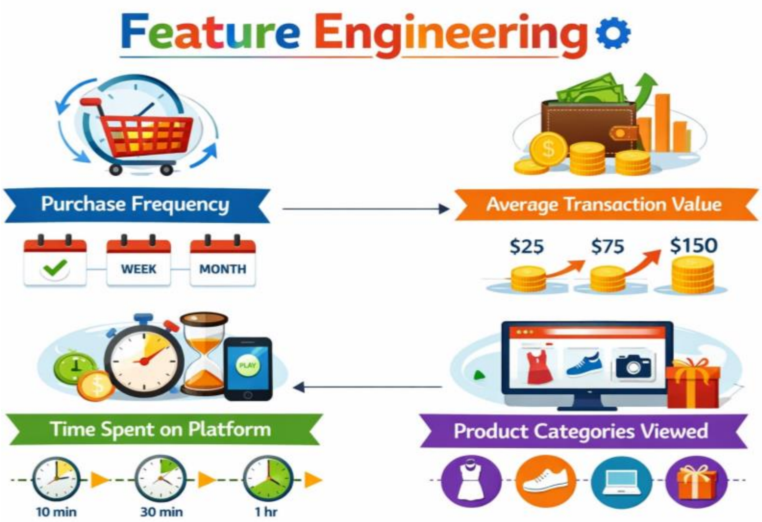

Fig 3 - Feature Engineering

3.3.1. Purchase Frequency

Purchase frequency is the frequency on which a customer conducts transactions in a given time period. It is a critical measure of customer interest and dedication in an online shop. More important customers are those who have increased frequency of purchase since they bring in stable revenue. The feature is used to determine repeat buyers, as well as customer segmentation, and future customers. Mostly it is counted as purchases per week, month or year, according to the analysis needs.

3.3.2. Average Transaction Value

Average transaction value (ATV) is the average purchase price of a customer. It is determined by dividing the overall revenue made by a customer with the transactions they have made. The attribute can offer an understanding of customer tendency to spend and buying power. An increase in the ATV values suggests consumers who usually purchase high quality or several products, and they are the key product of strong personalized marketing efforts. It also assists firms to plan the pricing techniques and offers.

3.3.3. Time Spent on Platform

The time spent is an indicator of how long a user is actively involved with the e-commerce site or application. This involves searching goods, descriptions, and dealing with suggestions. It can be used as a powerful behavioral indicator of customer interest and intention. Shorter sessions are usually associated with the absence of purchase, and longer sessions are usually associated with a

ISSN: 3071-5628
PP: 01-15.

high probability of purchase. This keeps its end users in the platform by analysing their experience and optimising its design to enhance retention and conversion.

3.3.4. Product Categories Viewed

Product categories of interest capture the nature of products that a customer will constantly visit during the process of entering the web pages. It is an aspect that assists the company in appreciating customer preference, interests, and intent. The patterns recognized by the system through the analysis of interaction on category levels include repetition of interest in electronics, clothing, or home appliances. These data are critical when developing on-to-one recommendation systems and user-specific marketing programs. It can also help predict future-buys by relating the recommendations to the previous browsing history of the customer.

3.4. Machine Learning Models

3.4.1. Logistic Regression

Logistic regression is another popular binary classification supervised learning algorithm used to predict a purchase or not by a customer. It is a modeling tool used to predict the probability of a given outcome given the input features such as the browsing behavior, past purchases, and the user demographics. This model is easy to implement, efficient and is easy to use in making predictions; hence applicable in basis forecasting. The interpretability is one of its major strengths because the analyst can know the impact of every feature on the prediction. Nonlinear relationships in customer behavior may however be difficult to capture using it.

3.4.2. Decision Tree

Decision trees are explainable and easy to understand models that categorize the information by subdivision into recurrent divisions based on the feature values. With e-commerce, the patterns that can be identified with the help of decision trees include What factors contribute to a purchase decision. Every internal node would be the choice to make depending on a feature and leaf nodes would be the ultimate outcome. The models are simply visualizable and explainable, hence useful in business enlightenment. Nevertheless, decision trees might be susceptible to overfitting; particularly when the data is large and complex unless they are pruned and/or regularised.

3.4.3. K-Means Clustering

K-means clustering is an unsupervised clustering algorithm that is used in how to segment customers into various groups based on behavioral similarities. It divides the data set into specifically specified number of clusters, in which cluster is considered as a group of customers with same characteristic such as purchasing behavior or Wholesale behavior. The method is especially ideal when used in market segmentation, target marketing, and a personalizing strategy. K-means is accurate, efficient, and computationally, as well as scaling, yet the algorithm relies on the selection of cluster numbers and the initial location of cluster centres.

3.4.5. Neural Networks

Neural networks are effective machine learning models that are aimed at incorporating complex and nonlinear data relationships. They are built up of several layers of connected nodes which process input features and learn hierarchical representations. Neural networks have the ability to process huge amounts of data, such as user interaction data, product features, and time series data in e-commerce applications to produce accurate prediction and suggestions. They are particularly convenient to be used in demand forecasting and customized recommendations. Nevertheless, neural networks are very computationally intensive and substantial amounts of data have to be trained and their inability to be interpreted can be a limitation in practice.

3.5. Proposed Hybrid Model

The hybrid model suggested a combination of both the unsupervised and supervised learning methods, in order to improve the accuracy and efficiency of customer behavior predictions in e-commerce sites. The first phase involves the customer segmentation by the K-means clustering algorithm that categorizes customers according to the similarity of their behavioral pattern in terms of purchasing, visiting and staying at the online shop, spending, and selection of products. This division will enable the system to distinguish between different classes of customers including frequent buyers, occasional customers, and inactivity customers in order to conduct more specific analysis. After the first step, classification tools, e.g. logistic regression, decision trees, or neural networks, are used on each cluster, as opposed to the whole dataset. This local modelling strategy gives better performance on prediction since the models are trained on more homogenous data which enables them to be more effective in segment specific patterns. An example is that factors affecting the purchase decisions where there is a difference between the high-value customers and casual browsers and the hybrid model reflects that fact. The system in the last step generates the individualized forecasts of each user e.g. the probability of a purchase, products, or probability of churn. The model can increase the prediction accuracy not only with clustering and classification but it also becomes more scalable and interpretable. Also, this methodology lets evolve dynamic updates so that customer groups and predictive models can be updated also as new data are shared. In general, the hybrid framework offers a powerful and adaptable model of offering unique customer experience and streamlined business practices in the competitive e-commerce setup.

3.6. Evaluation Metrics

3.6.1. Accuracy

One of the most common evaluation measures when deployed in classification tasks is accuracy. It quantifies the quality of the goodness of the model, by computing the percentage ratio of instances that are correctly predicted (both negative and positive) against total predictions made. Under the e-commerce environment, accuracy means how favorably the model is able to forecast the customer behaviours like purchasing or not purchasing. It is simple to interpret, but accuracy might not necessarily be credible, where the dataset is imbalanced, which is likely to produce deceptive outcomes as it favors a majority category.

3.6.2. Precision

Precision is a basic measure that shows the ratio of accurate positive predictions to the positive predictions that the model makes. That is, it checks the number of the foreseen good results that are in fact accurate. True precision is significant in the e-commerce use case where CFP is very high like the one which suggests irrelevant product to the customer. A highly precise model guarantees that suggestions or estimates by the model are very relevant and credible thus enhancing the satisfaction of the users.

ISSN: 3071-5628
PP: 01-15.

Fig 4 - Evaluation Metrics

3.6.3. Recall

Recall or sensitivity is the ratio of the number of correct responses of the model on exactly positive responses. It dwells on the capability of the model to attract all the relevant cases. Recall plays a key role in customer behavior analysis where it is essential to determine as many potential buyers as possible. When the value of recall is large, the model has managed to identify most of the customers that are likely to take a desired action despite having false positive rates.

3.6.4. F1-Score

The F1-score represents the harmonic mean of the two variables, precision and recall, which is a balanced score used to capture the false positives and false negatives. It is especially handy in the case of lopsided datasets, in which simplicity of accuracy alone might not provide a full understanding of the model performance. The F1-score can also be used to assess the compromise between the model identifying a relevant customer and not making a wrong prediction in cases of e-commerce. When the F1-score is higher, it means that the model is stronger and has more reliability.

4. RESULTS AND DISCUSSION

4.1. Performance Analysis

The performance of the suggested system analysis demonstrates how the AI-based models are much superior to conventional statistical methods of customer behavior analysis in the framework of e-commerce platforms. The traditional approaches, regression and rule-based approaches tend to fail to resolve complex and nonlinear relationships found in the large volumes of customer information. Machine learning and deep learning models, on the other hand, exhibit a superior level of detecting hidden patterns, positively affecting more reliable and accurate predictions. The findings show that AI models are more accurate as well as precise and recall, especially in tasks like purchase prediction, customer segmentation, as well as recommendation generation. Out of all the considered methods, the suggested hybrid model presents the best improvement in performance. The combination of the clustering methods and the classification algorithms makes the model useful in dividing the customers into the homogeneous groups and then further applying the predictive analysis. This segmentation enables the system to model specific behavior of customers and therefore noise is minimized and hence learning is done efficiently. Consequently, the hybrid method has superior generalization and reduced errors in prediction in comparison to individual models. Besides, the hybrid model provides enhanced scalability due to the ability to split the computation load into segments, and it can support large and dynamic data. The assessment also shows the model has a consistent performance across various segments of customers, which implies that it is robust and flexible. Moreover, the presence of feature engineering methods helps to enhance the performance of the models due to the more informative input variables. All in all, the discussion proves that the

ISSN: 3071-5628
PP: 01-15.

suggested AI-powered hybrid framework does not only shine in comparison to the conventional approaches, but it also provides a scalable and efficient way to provide a customer with personalized results and optimise the decision-making process on the e-commerce system nowadays.

4.2. Results Analysis

Table 1: Results Analysis

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	78%	75%	72%	73%
Decision Tree	82%	80%	78%	79%
K-Means + Classifier	85%	83%	81%	82%
Neural Network	89%	87%	86%	86%
Proposed Hybrid Model	93%	91%	90%	90%

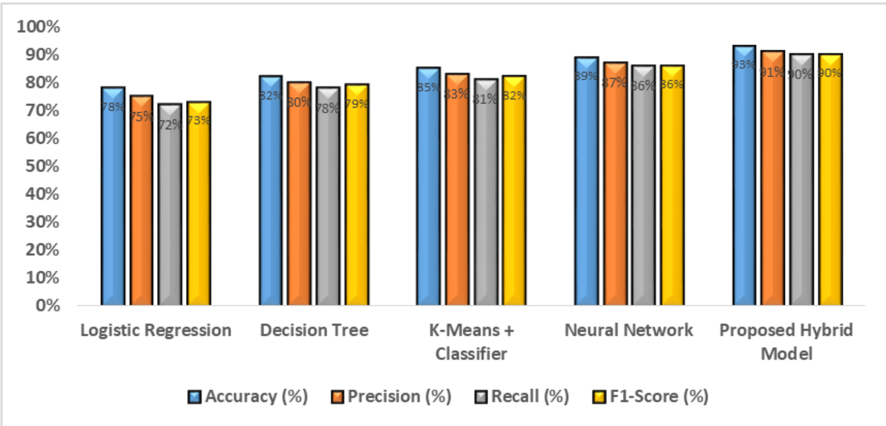

Fig 5 - Results Analysis

4.2.1. Logistic Regression

This logistic regression model had an accuracy of 78% and the values of precision, recall and f1-score were 75, 72, and 73 respectively. These findings imply that the performance of logistic regression is low because it fails to depict complicated nonlinear interactions in the data though it provides a decent baseline on predicting customer behaviors. The recall rate is comparatively low, which implies that the model is not capturing a significant portion of the potential positive cases, including customers who decide to purchase the product. Although logistic regression is easy to interpret and use, it has reduced the ability to work with complex behavioral patterns in the huge e-commerce data collections.

4.2.2. Decision Tree

The decision tree model proved to be more efficient than logistic regression because it had an accuracy of 82% as well as precision, recall, and F1-score values of 80 per cent, 78 per cent and 79 per cent respectively. This can be explained by the fact that the model can be used to estimate nonlinear association and hierarchical decision making that damage splits of features. The decision trees are considered very beneficial because they help understand the path of the decisions made by the customers and hence they are useful in terms of interpretability. Nonetheless, due to overfitting, their work still may be impaired, particularly in terms of dealing with complex and high-dimensional data.

ISSN: 3071-5628
PP: 01-15.

4.2.3. K-Means + Classifier

The hybrid solution that uses K-means clustering and a classification model was found to reach an accuracy of 85 and precision, recall and F1-score of 83, 81, and 82. This model has an advantage of splitting customers into homogenous groups, followed by the classification thus improving accuracy of predictions. The classifier is able to learn patterns of segments better by decreasing the heterogeneity of the data. The perceiving likelihood is improved thus providing a better identification of the existing customers whereas the F1-score and precision were balanced indicating the overall strength. The strategy shows that a combination of unmonitored and supervised learning methods is effective.

4.2.4. Neural Network

The model of the neural network attained a high accuracy of 89% reflected in the values of precision, recall, and F1-score of 87, 86, and 86. This makes it perform superiorly owing to its capability of modeling model complex and nonlinear relationship with customer data. Neural networks have ability to analyze a large area of high-dimensional data and hence suitable in capturing complex behavioral patterns. The balanced measures of evaluation are strong predictive and generalization. Nevertheless, the model is time-consuming in terms of computational resources, and it does not have an interpretable nature, which can become a constraint when applied in practice.

4.2.5. Proposed Hybrid Model

The hybrid proposal was more successful than any other model with the highest accuracy of 93, precision, recall, and F1-score of 91, 90, and 90. This great enhancement can be explained by the incorporation of customer segmentation and classification methods, with the help of which the model is able to examine the behaviour of the more homogeneous groups. Being based on both clustering and the sophisticated predictive modeling, the hybrid method proves to be sufficient to cover intricate patterns and at the same time allows a high level of generalization. The balanced and high evaluation metrics demonstrate that the model is accurate and reliable to make it highly appropriate to the application in the real-world e-commerce, aimed at personalization and decision support.

4.3. Discussion

As shown in the results of the experiment, the developed hybrid AI model showed better results compared to separate machine learning models because it allows the customer segmentation process to be effectively combined with the predictive analytics. The hybrid method is the combination of the methods of clustering and the classification models, which combine the advantages of unsupervised and supervised learning. The processes of segmentation takes the customers and clusters them into more homogeneous groups on the basis of their behavior traits thus minimizing data variability and enabling the models of classification to lean on rather specific and meaningful patterns in each segment. Such targeted learning has a strong improvement in prediction quality and the general model performance as compared to the use of one model prediction on the entire dataset. Although neural networks have also recorded good performance because of their ability to bring about complex nonlinear relationships, they lack interpretability and it is not easy to discern the line of thinking by the prediction made by the neural network by a business. Conversely, models such as decision trees are more transparent and explainable and this is important in strategic decision-making. Nevertheless decision trees might not be sufficient to build high predictive accuracy as complex models. The hybrid model has managed to strike a balance by enhancing accuracy with some extent of interpretability in its segmented analysis. Additionally, the adoption of clustering leads to more favorable management of various customer behaviors, which can be more personalized and conservative. Not only does this enhance model efficiency, but also increases scalability in real life e-commerce scenarios where the data can be observed to be continually varying. On the whole, the results present the significance of the combination of several methods of analysis to address the drawbacks of separate models to come up with a more robust, precise, and practical solution to the analysis of customer behavior.

5. CONCLUSION

In this paper, the author provided an in-depth AI-based model of customer behavior analysis in online shopping websites, highlighting the increasing significance of smart systems in the contemporary digital market. The paper also systematically examined the history of how traditional statistical models evolved to the current methods of machine learning and deep learning, and established how the models are improving consumer behavior by a significant margin in understanding, predicting, and influencing their behavior. Using the opportunities provided by the data on large numbers of customers, the proposed framework can assist businesses in learning more about the type of preferences in users, the pattern of purchasing, as well as the engagement behavior, which can, in turn, be used to implement more efficient approaches to personalization. The results of the experiment proved that machine learning frameworks are more accurate, scaled and adaptable as compared to traditional approaches, and thus can be applied successfully in dynamic environments in e-commerce. One of the main contributions of the given research is the creation of a hybrid framework where, by combining clustering techniques with classification techniques, a better predictive quality is achieved. Through both the clustering algorithm based customer segmentation and the classification models within the segment, the proposed method can effectively decrease the heterogeneity of the data and attain behavioral segments depending on the segment. This combination does not only increase the quality of prediction but also increases the strength and extrapolation capacity of the system. The hybrid model has been noted to be the best in all the evaluation metrics, thus demonstrating the need to integrate several AI methods to overcome the shortcomings of each model. What is more, the structure promotes scalable deployment, which explains its suitability with realistic e-commerce platform with ever-changing datasets. The results of this research support the disruptive nature of artificial intelligence in e-commerce analytics. Artificial intelligence model allows making decisions according to the data because it will give the practical information that can be used to optimize the market strategy and customer retention and make a business more profitable. Nonetheless, interpretability of the model, privacy of the data and the real time processing are major areas that are still at a critical stage and need prior consideration. Future studies can, in this regard, be directed to integrating the concept of real-time analytics to process streaming data, using reinforcement learning to create adaptive and flexible recommendation systems and building explainable AI models to increase transparency and trust. These issues can be solved to make the systems of the future more efficient, reliable, and user-oriented. On the whole, the present work shows that the combination of innovative AI methods is required to create the smart, scalable, and efficient customer behavior analysis systems in the fast-developing e-commerce environment.

REFERENCES

[1] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” *Proc. 20th Int. Conf. Very Large Data Bases (VLDB)*, pp. 487–499, 1994.
[2] D. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining*, Cambridge, MA, USA: MIT Press, 2001.
[3] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, New York, NY, USA: Springer, 2009.
[4] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., San Francisco, CA, USA: Morgan Kaufmann, 2011.
[5] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[6] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

ISSN: 3071-5628
PP: 01-15.

[7] J. B. MacQueen, “Some methods for classification and analysis of multivariate observations,” *Proc. 5th Berkeley Symp. Mathematical Statistics and Probability*, pp. 281–297, 1967.
[8] G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” *IEEE Trans. Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, 2005.
[9] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” *Computer*, vol. 42, no. 8, pp. 30–37, 2009.
[10] X. Su and T. M. Khoshgoftaar, “A survey of collaborative filtering techniques,” *Advances in Artificial Intelligence*, vol. 2009, pp. 1–19, 2009.
[11] [P. Resnick and H. R. Varian, “Recommender systems,” *Communications of the ACM*, vol. 40, no. 3, pp. 56–58, 1997.
[12] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 1097–1105, 2012.
[14] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[15] J. Chen, H. Zhang, X. He, L. Nie, W. Liu, and T. S. Chua, “Attentive collaborative filtering: Multimedia recommendation with item- and component-level attention,” *Proc. 40th Int. ACM SIGIR Conf.*, pp. 335–344, 2017.