JOSTIR Publication 2025 | ISSN: 1118-3896 Special Issue December 2025 | 440-448

Journal of Science, Technology and Innovation Research
Volume 1 Special Issue | December 2025

Development of a Predictive Model for E-commerce Users' Shopping Patterns

[1]Ibam, E. O., [2]Abimbade, A. B., [3]Makinde, A. I.

[1,3]Department of Information Systems, Federal University of Technology, Akure, Nigeria
[2]Department of Computer Science, Federal University of Technology, Akure, Nigeria

Correspondence: eoibam@futa.edu.ng, ababimbade@futa.edu.ng, iamakinde@futa.edu.ng

ABSTRACT

The exponential growth of e-commerce has led to an increased need for understanding users' shopping patterns. This study proposes a predictive model using the Markov Decision Process (MDP) and Recurrent Neural Networks (RNN) to forecast e-commerce users' shopping patterns on e-commerce platforms. The objective is to develop a model that accurately predicts users' shopping patterns, enabling the delivery of personalized recommendations and an enhanced customer experience. A large amount of user-interaction data of Alibaba e-commerce site was gathered and pre-processed to obtain the meaningful features. These features were then used to train the MDP model which was then used to predict future user behaviour. The model takes into consideration the states that represent the user actions, actions indicating transitions between states and rewards indicating the probability of a purchase. These findings show that the hybrid of the models is more effective in predicting the shopping patterns of the users as compared to the traditional predictive models. The Markov model shows an accuracy of 93.3%, the MDP policy shows an accuracy of 89.9%, and a Q-Learning policy accuracy of 94.3%. This study concludes that the models effectively model complex user behaviour on e-commerce platforms. The model can be incorporated into the existing recommendation systems, which will allow business to offer personalized experiences and enhance customer satisfaction and retention. The implications of our findings to the e-commerce industry are great since MDPs have the potential to increase user engagement and sales.

Keywords: E-commerce, Predictive Modeling, Markov Decision Process (MDP), Recurrent Neural Networks (RNN), Shopping Patterns, User Behavior, Recommendation Systems, Q-Learning

Introduction

The limitations of Ziqi (2024) and Rumen *et al.* (2023) on predicting e-commerce users' purchase behaviour are the key motivation for this research work. These include geographical limitations, insufficient data sources, and the capture of only consumers' personality traits. Therefore, there is a need to develop a model that analyzes, forecasts, and accurately predicts personalized recommendations for e-commerce users' shopping patterns. Electronic commerce (e-commerce) is the term that indicates the use of Information and Communication Technology (ICT) facilities in doing business (Ibam *et al.*, 2018). E-commerce has been growing at a rapid rate which has resulted in the emergence of advanced digital marketplaces where businesses must keep up with the changes in technology, consumer behaviour and regulation. Growth and development of e-commerce is an activity of confidence that is built through trust that is present between the players, business and government within the e-commerce setting. (Akinyede, 2017). Online shopping is constantly

doi.org/10.51459/jostir.2025.1.Special-Issue.0187

gaining popularity, as individuals find the platform much more convenient, as they are able to shop at any place. (Syeliya *et al*, 2024).

As the Internet technology has been rapidly growing, the e-commerce industry has slowly formed a significant segment of the global economy. The future of e-commerce is an innovation and expansion with a greater number of new technologies and an emphasis on customer-centric approaches. The long-term observation of the actions of the e-commerce users will give us an idea of their needs and habits (Grunt *et al*, 2017).

The prediction of the will to purchase and demand by the users can help the platform to plan the management of inventory and distribution of logistics and, therefore, optimize operations and cost reduction (Ziqi, 2024). Nevertheless, the evolution of technology has led to the use of the clickstream data as the main source of information that a company uses to research on the preferences of online customers. It consists of the pattern of online activities performed by the users as they browse the web pages and online platforms and reflects a treasure of information on user behavior and interaction and engagement (Deming *et al.*, 2024).

This has resulted in the rising exponential growth of the E-commerce industry that has created a demand on the availability of precise sales forecasts to promote effective business strategies (Rohit *et al.*, 2023).

The user behaviour is vital in businesses in the e commerce environment since personalisation, trust and security, convenience, and availability of information are some of the issues that affect the consumer behaviour. The e-commerce has been transformed with the use of such technological innovations as AI, ML, blockchain, and mobile technologies and has improved personalization, security, and the overall user experience. E-commerce is a driver in the internationalization of doing business because it minimizes barriers to entry as well as maximizing logistics and allowing access to global markets. Through the analysis of the buying behaviour amongst users, not only can the level of sales efficiency be improved, but also the consumer can receive more personalized and precise shopping suggestions, which is a win-win situation to both the business performance and the user experience (Shuang *et al.*,2024).

Predictive analytics is fundamental in improving user experience, personalisation and business strategies in e-commerce through the use of data to predict customer behaviour and optimize business. The behaviour prediction used with e-commerce users implies the forecast of potential consumption demand and habits of the consumers on the basis of their behaviour on the E-commerce websites. (Yuran *et al.*, 2022)

With the help of more sophisticated algorithms, e-commerce companies can predict consumer demand, work with it individually and increase the overall satisfaction of customers. The significant contributions of the predictive analytics in this field are the improvement of the user experience, strategies personalisation, inventory management and proactive decision-making.

The Markov Decision Process (MDP) and Recurrent Neural Networks (RNN) are some advanced modelling techniques that have a great impact on predicting the user shopping preference. MDP offers a solid framework in making a decision in situations of uncertainty, and the modelling of customer behaviour in any shopping situation is possible. RNNs are highly successful in the task of extracting temporal patterns of user behaviour and are therefore an effective tool in the analysis of complex and dynamic e-commerce data.

RNNs are also efficient at identifying patterns with time which is essential in analyzing shopping patterns that change as users interact. They have proved to be very accurate in predicting customer behaviour as compared to the traditional methods with a high of 93.2 accuracy in different studies.

The MDP model of shopping behaviour consists of the state description, dynamics of action and the reward processes. A reward value is attached to every state and thus the decision making process and to determine the best actions to take to maximize expected returns.

Predicting models should be accurate so as to maximize the marketing, inventory management and retention of customers. These models rely on historical information and powerful analytics to predict consumer behaviour so that businesses can make quality decisions that improve operational efficiency and profitability. Marketing optimization, inventory management, and customer retention are some of the important areas of application.

Marketing optimization is a process of analysing consumer data so as to enhance conversion rate and customer retention. Knowing customer preferences and behaviours enables the firms to customize their products to increase satisfaction and loyalty. The inventory management allows the retailers to make good demand predictions, reduce overstocking or stock-outs, and they are able to make correct changes in inventory levels to match the expected demand changes due to seasonality or promotions. Businesses have always required sales forecasting to predict the demand, manage inventory and make forecasts for strategic decisions. Sales forecasting has been an important aspect of business since the beginning of time; it helps businesses to know what to expect, and manage the inventory as well as make reasonable strategic decisions (Harshit *et al.*, 2024).

Customer retention is a process that is conducted to find the at-risk customers and to apply proactive engagement schemes to improve customer relationship management. Predictive insights can be recommended to customers in a personalised way to encourage stronger customer relationships and boost lifetime value.

Nevertheless, there are a number of challenges in capturing and predicting the complex e-commerce user behaviour. Information can never be useful until it is transformed into data. There is a need to process this massive data and find some valuable information in it (Pavani *et al.*, 2020).

The large dimensions and unequal distribution of classes can complicate the process of feature extraction and model training and temporal dynamics demand models that are capable of capturing time-dependent trends. Such methods as Recurrent Neural Networks (RNNs) and time-series attention models have demonstrated potential in dealing with those issues. The encoded time-stamps of events have the potential to improve the insight into the behaviour of the user over the time, a factor that can easily be ignored within the conventional framework.
Personalisation and user modelling are considered to aid in successful personalisation because different factors determine personal behaviour in an individual user, such as demography and transaction history. Session based prediction activities that require real-time personalisation are critical activities whereby user intent may change drastically in one shopping session.

Materials and Methods

The research methodology was achieved using five major stages, including Phase 1: Review of related works, Phase 2: Data Collection, Phase 3: Design of a model predicting the shopping patterns of e-commerce users using MDPs and RNNs, Phase 4: System Implementation, and, Phase 5: System Evaluation, were used to reach the research methodology, based on Accuracy, Precision, Recall, and F1 score.

The systematic search was carried out on the internet resources and academic literature on e-commerce and e-commerce shopping patterns, as well as on machine learning and machine learning models based on MDPs and RNNs.The dataset was obtained through the AliExpress e-commerce website and further enriched with web scraping using Apify.

Data Collection and Preprocessing

Apify was used to scrape additional transaction records from online retail websites. The extracted dataset was merged with the AliExpress-based online retail dataset, which contains: Invoice Numbers (Unique purchase identifiers), Stock Codes and Descriptions (Product details), Quantities (Number of items per transaction), Invoice Dates (Timestamps for transaction sequences), Customer IDs (Anonymous unique customer identifiers), Prices (Product cost in Nigerian Naira)

Data Cleaning and Processing

The dataset was cleaned and formatted using Pandas and NumPy. Rows with incomplete information were removed. Invoice dates were transformed into a datetime format using pd.to_datetime. Transactions were grouped by Customer ID to identify purchase patterns.

Implementation of the Predictive Model

Libraries Used

Python libraries used include Pandas for data manipulation and preprocessing, NumPy for numerical computations, Seaborn for data visualization (bar plots, distributions), and Matplotlib. pyplot for generating loss curves and evaluation metric charts, sklearn.model_selection for splitting the dataset into training/testing, sklearn.metrics for computing accuracy, precision, recall, and F1-score, Pymdptoolbox for implementing Markov Decision Process(MDP), and Tabulate for formatting dataset display in the console

Markov Decision Process (MDP) and Recurrent Neural Network (RNN) were applied to predict the next product a customer is likely to purchase based on previous transactions.

Transition Matrix Calculation: Customer purchase sequences were extracted and analysed. A transition matrix was computed, representing probabilities of moving from one product to another. The matrix was normalized to obtain realistic probabilities.

Prediction Mechanism: Given a customer's current purchase, the model predicts the most probable next product. The pymdp toolbox library was used to implement state transitions based on past purchases.

Results and Discussion

Purchase Likelihood and User Behaviour Cluster

Figure 1 shows the likelihood of purchase at the first event stage and highest likelihood is given to purchase. This means that purchase action users are very likely to make further purchases and initial purchase

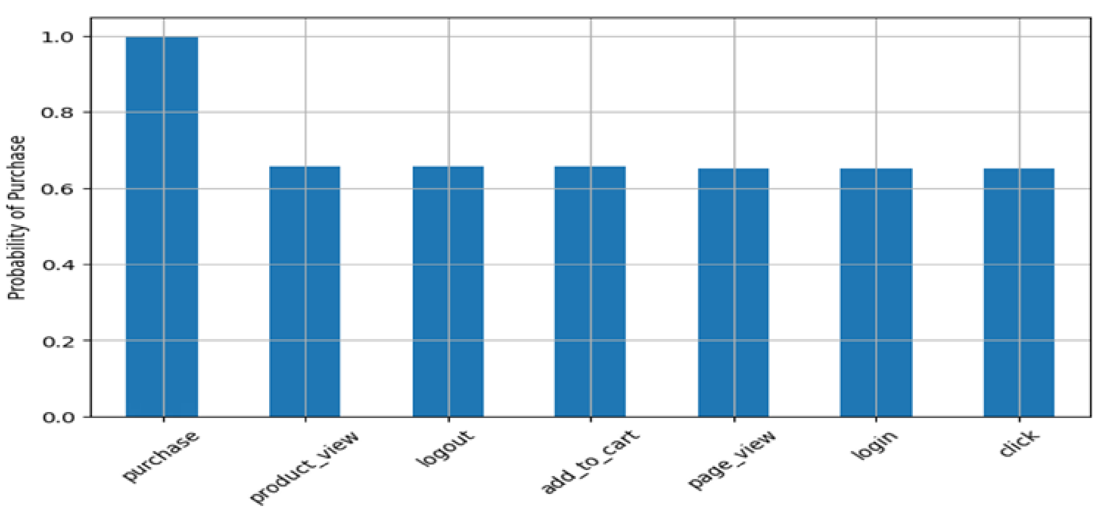

Figure 1. Purchase Likelihood by first Event Stage

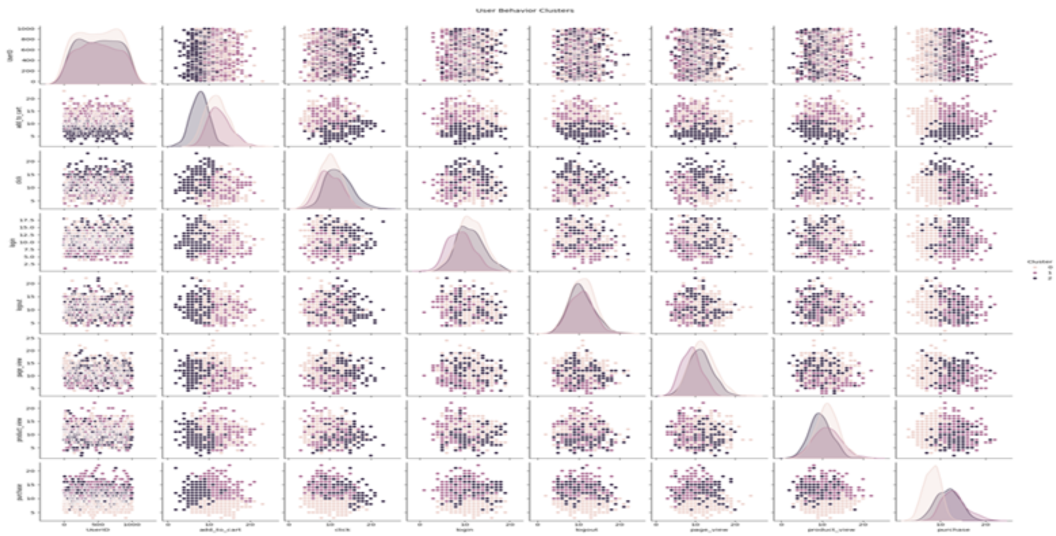

Figure 2: User Behaviour Cluster

or buying indicators are important in modelling customer lifetime value or buying propensity.

The cluster scatter presented in Figure 2 and density plots, represent segmentation in user behavior on the basis of several features (such as user ID, add-to-cart, click, login, logout, etc.). These three clusters give insights into the underlying patterns or the user groups, who may be casual browsers, engaged buyers, and frequent purchasers, with their unique distribution of features. Density plots demonstrate the concentration of behaviors of every cluster and provide information about the common user types and the level of activity they have.In Figure 3, we have provided the Sankey diagram which visually models a simulation of user journeys on an e-commerce platform which depicts the flow of the user through various actions. The diagram shows important transitions of users with a higher volume of users indicated by thicker flows between the actions. The process starts with users accessing pages as indicated by the flow of the node of page views. The next action users take is the login action, and this serves as a point of entry into more active actions. Upon their logging in, they are taken to a product view, indicating an interest in particular products. The key distinction of this diagram is the cyclic nature of the flow between product view to again login as some users may want to revisit the

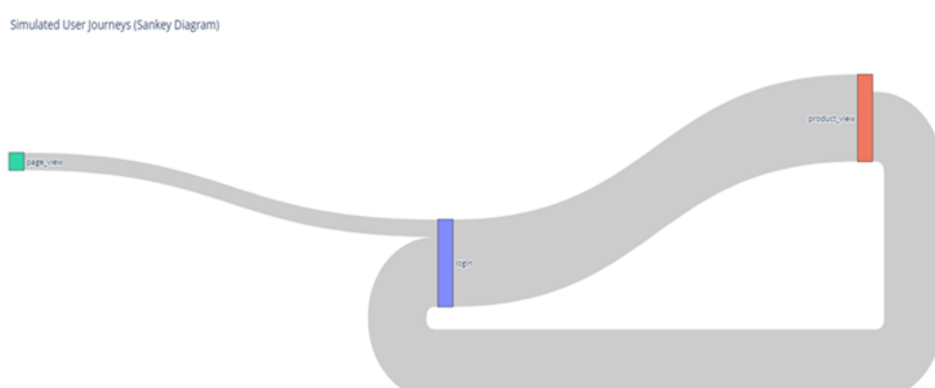

Figure 3. Sankey Diagram showing simulated user journey

product again post the re-logging in process, perhaps to continue with the product further or compare it with other ones.

The Sankey diagram provides the information about the user navigation directions, thus allowing the analysts and marketers to see the major routes and points of the user journey that can be adjusted and optimized to enhance the user experience and conversion rates.

Table 1. Comparative insights of Markov Model, MDP Policy and Q-Learning Policy

Model	Accuracy	Precision	Recall	F1 Score
Markov Model	93.3%	0.93	0.93	0.93
MDP Policy	89.9%	0.89	0.89	0.89
Q-Learning Policy	94.3%	0.94	0.94	0.94

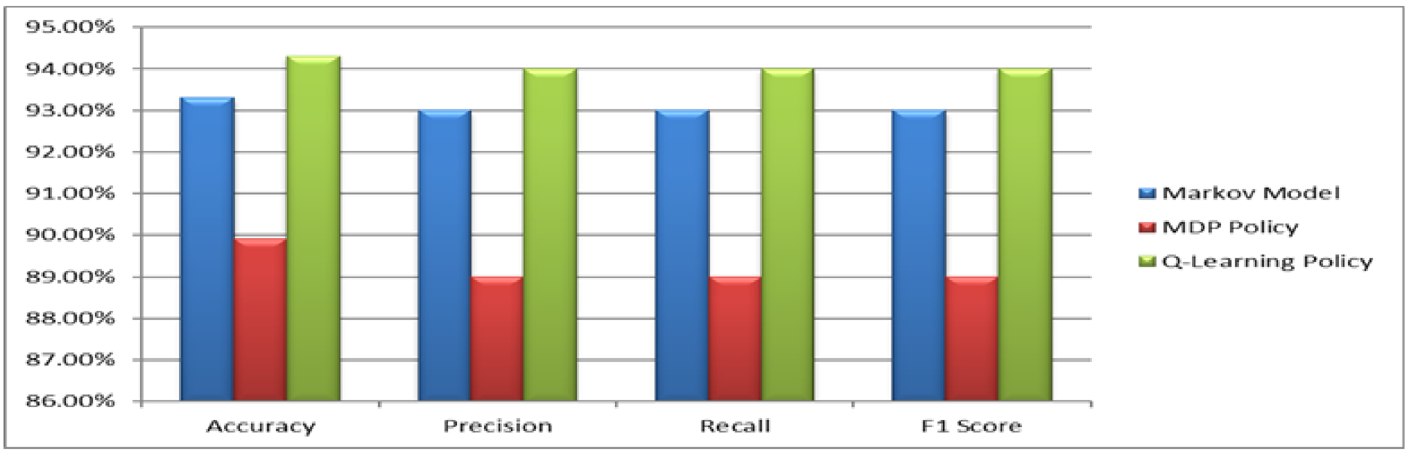

Figure 4. Comparative insights of Markov Model, MDP Policy and Q-Learning Policy

Prediction Accuracy Analysis

Standard measures of classification: Accuracy, Precision, Recall and F1 Score were used to assess the predictive strength of the developed models. Three different modeling strategies that included Markov Model, MDP Policy and Q-Learning Policy were implemented to predict user shopping behaviour given the data of sequential user behaviour. Table 1 and Figure 4 indicate high predictive effectiveness of all models and Q-Learning is more effective than the rest.Markov Model Performance

The Markov Model has a prediction accuracy of 93.3 and the Precision, Recall and F1 Score values were also estimated at 0.93. This model was effective in determining consecutive states of the user and forecasting transition like browsing, adding to cart and purchasing. It demonstrates the fact that it is not necessary to deep-learn policies to model state-to-state transitions. The findings affirm that the Markov Model, though less complex than the deep reinforcement learning models, is strong in representing regularities in the user navigation behaviour in e-commerce websites.

MDP Policy Performance

Markov Decision Process (MDP) Policy model reported that its accuracy is 89.9 percent, Precision, Recall, and F1 Score estimates are 0.89. This is to include the chances of transitioning between one state and the other not only in terms of transitions frequencies but also in a reward based strategy. The findings render MDP appropriate in locating the path of high value which includes those that result in conversions hence helpful in designing the specific marketing or recommendation strategies. This less performance is explained by the fact that specifying rewards is more challenging and that some rewards (purchases vs page views) are sparse.

Q-Learning Policy Performance

Q-Learning Policy model did better at performance with a precision of 94.3 and estimated Precision, Recall and F1 Score of 0.94 each. This algorithm successfully used user action sequences of large scale to learn the behavioural patterns that result in successful outcomes including purchases or sustained interactions. The classification report indicates that there are high values consistently (0.93) in the major user actions such as add to cart, click, login, log out, page view, product view, and purchase. The large support numbers (9,100 to 9,300 per type of action) show that the dataset is well-balanced, which supports the idea of the model being reliable in all cases of interactions between users.

Policy Simulation and Business Impact

The adjusted policy simulation of 0.789 shows that the model can assist users with a more streamlined shopping process, which means that their chances of conversion are increased. This simulated rate indicates that, when implemented, the policy will, in fact, result in a conversion rate of about 79 which is much higher than industry averages which is relevant in influencing strategic decisions in the management of inventory, dynamic pricing, promotional offers as well as personalized recommendation engines.

Markov Decision Processes and the application of reinforcement learning using Q-Learning has been effective in predicting user behaviour within an e-commerce setting. These findings confirm the presence of considerable potential in improving personalization, user experience and business performance to make e-commerce platforms smarter and more adaptive.

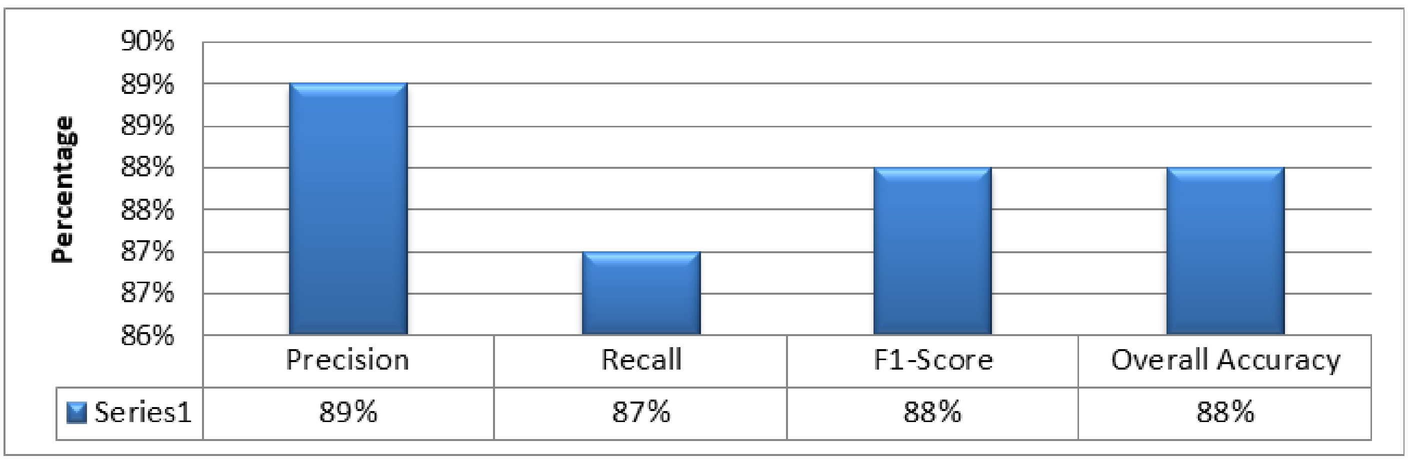

Figure 5. RNN model results

RNN Model

The training results for the RNN model after 10 epochs, as shown in Figure 5, show a very good performance metrics

Training Loss- Slightly decreases from 1.9470 to 1.9451, indicating stable training without overfitting or divergence.

Classification Metrics

Precision: Approximately 89% across individual event classes.
Recall: Around 87%, indicating good coverage of actual events.
F1-Score: About 88%, showing balanced precision and recall.

Overall Accuracy: 88%, reflecting reliable prediction of user actions.

Jaccard Similarity Scores

Figure 6 shows the high average Jaccard similarity scores of 0.92 for the RNN, 0.94 for the MDP, and an overall 0.95, indicating that both models are remarkably effective at generating user paths closely aligned with actual user behaviors.

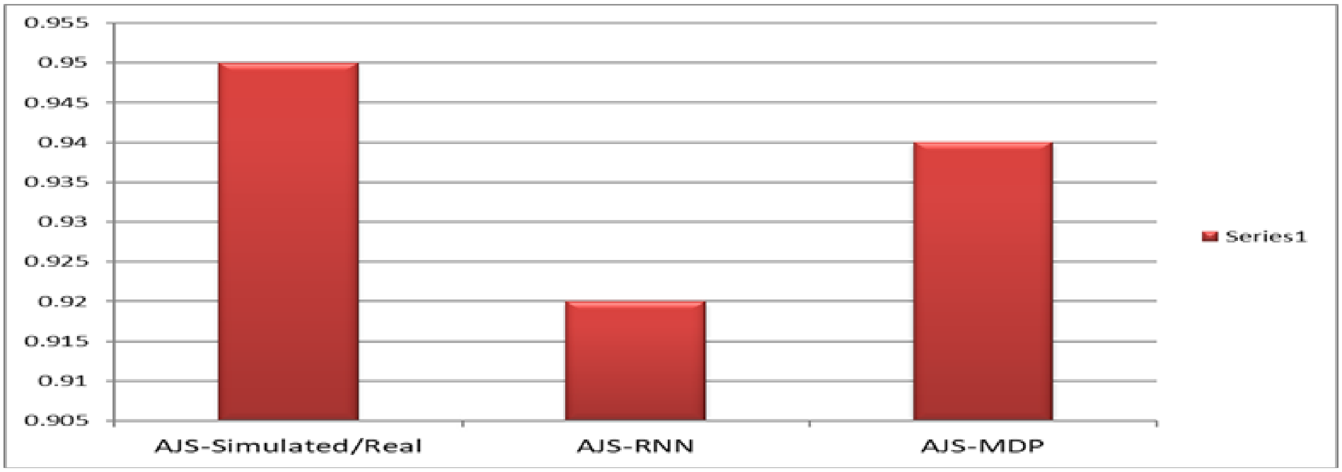

Figure 6. Average Jaccard similarity scores for RNN, MDP & simulated/real

References

Akinyede, R.O. (2017) 'Building an Enterprise-Class Electronic Commerce in Nigeria', International Journal of Applied Information Systems, 15, (7)

Dong *et al.* (2022) 'Integrated Machine Learning Approaches for E-commerce Customer Behaviour Prediction', Proceedings of the 2022 7th International Conference on Financial Innovation and Economic Development, p. 648.

Efor, E. and Chukwu, J. (2020) Application of Markov Analysis to Consumer Preference. International *Journal of Mathematics and Statistics Invention*, 8(6) 14-23.

Grunt *et al.* (2017). An approach to customer behavior modelling using Markov decision process. MENDEL 23 (1)

Ibam *et al.* (2018) 'E-commerce in Africa: The case of Nigeria', EAI Endorsed Transactions on Serious Games, 4 (15)

Ketipov *et al.* (2023) 'Predicting User Behavior in e-Commerce Using Machine Learning', Cybernetics and Information Technologies, 23 (3), 89-101.

Liu *et al.* (2024) 'Enhancing customer behavior prediction in e-commerce: A comparative analysis of machine learning and deep learning models', Applied and Computational Engineering, 55, 181-195.

Liu, Z. (2024) 'Prediction Model of E-commerce Users' Purchase Behavior Based on Deep Learning', Frontiers in Business, Economics and Management, 15 (2), 147-149.

Patil, D. (2025) Artificial Intelligence in Retail and E-Commerce: Enhancing Customer Experience Through Personalization, Predictive Analytics, And Real-Time Engagement.

Pavani *et al.* (2020) 'Analysing Customer Buying Behaviour In Online Shopping Using Random Forest Classifier'.

Rohit *et al.* (2023) 'Unlocking Future Transactions: Predicting Customer's Next Purchase in E-commerce through Machine Learning Analysis', International Journal of Advanced Research and Innovative Ideas in Education, 9, (3), 2395-4396.

Sharma *et al.* (2024) 'Enhancement of Sales Forecasting and Prediction with Machine Learning Methods', International Research Journal of Computer Science, 11 (11), 641-644, doi: 10.26562/irjcs. 2024.v1111.02.

Sulastri, L. (ed.) (2023) The Role of Artificial Intelligence in Enhancing Customer Experience: A Case Study of Global E-commerce Platforms. International Journal of Science and Society Proceedings.

Zaini *et al.* (2024) 'The behaviour of e-commerce users: An empirical investigation of online shopping', Journal of Management World, vol. 2024, no. 2, pp. 50-60.

Zhou, S. and Hudin, N.S. (2024) 'Advancing e-commerce user purchase prediction: Integration of time-series attention with event-based timestamp encoding and Graph Neural Network-Enhanced user profiling', PLoS ONE, 19 (4), e0299087, doi: 10.1371/journal.pone.0299087.