



(RESEARCH ARTICLE)



## FROM CLICKS TO CONVERSIONS: INTERPRETING ONLINE CONSUMER BEHAVIOR USING EXPLAINABLE AI

Arsheha Arsheha <sup>1,\*</sup>, Muhammad Hammad Memon <sup>2</sup>, Samia Batool <sup>3</sup> and Sidra Sidra <sup>4</sup>

<sup>1</sup> Business School, Sichuan University, No.24 South Section 1, Yihuan Road, Chengdu, PR China, 610065.

<sup>2</sup> School of Computer Science and Technology, Southwest University of Science and Technology (SWUST), 59 Qinglong Avenue, Fucheng District, Mianyang, Sichuan 621010, P.R. China.

<sup>3</sup> Business School, Sichuan University, No.24 South Section 1, Yihuan Road, Chengdu, PR China, 610065.

<sup>4</sup> University of Chinese Academy of Social Sciences, Beijing, China.

\* Corresponding Author

### ORCID Details

Arsheha Arsheha: <https://orcid.org/0009-0008-7412-6753>

Muhammad Hammad Memon: <https://orcid.org/0000-0002-8680-1831>

Samia Batool: <https://orcid.org/0009-0008-6647-6956>

Sidra Sidra: <https://orcid.org/0009-0003-1689-3296>

World Journal of Advanced Research and Reviews, 2026, 32(01), 334–346

Publication history: Received on 29 August 2026; revised on 04 October 2026; accepted on 06 October 2026

Article DOI: <https://doi.org/10.30574/wjarr.2026.32.1.2548>

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

### ABSTRACT

Due to increased dominance of e-commerce in global retail industry, behavior knowledge makes online consumers engaged to end of buying products or services has become essential to enterprises. Irrespective of growing popularity of machine learning (ML) models predict conversion outcomes, there is a lack of interpretability limiting actionable information. This paper is a thorough report on explanation of online consumer behavior by utilizing two state of art exemplifiable AI (XAI) methodology namely SHAP (SHapley Value Additive exPlanation) and LIME (Local Interpretable Model-agnostic) on centered scores given by a gradient-based escalated tree model constructed on a genuine world data of one million 2021 to 2024 shopping sessions. Study goes through behavioral characteristics, like session length, views per session, referral sources, device types and abandoned carts. According to study results, session length and time on a product page stand out as the most important conversion drivers, and mobile users have a much lower conversion rate in comparison with desktop users, even though level of engagement is higher. SHAP analysis reveals non-linear influence between referral sources and user demographics whereas LIME learns local decision rules explain an individual conversion. In addition, research proposed a new hybrid XAI approach, SHAP-LIME Fusion, uses less explanation instability and outperforms individual methods in 32%. Findings are beneficial to digital marketers and UX designers who can use them to improve conversion funnels. Study finds its significance to both academia and practice of e-commerce since it not only fills in gap between prediction correctness and interpretability in e-commerce analytics.

**Keywords:** E-commerce, Online Consumer Behavior, Conversion Prediction, Explainable AI, Machine Learning, Behavioral Analytics

## 1 INTRODUCTION

The world e-commerce sector jumped to 6.3 trillion in 2023 and is expected to surge to 8.1 trillion by 2025. Online purchasing has grown to become the main channel of consumer transactions with more than 2.64 billion digital buyers across the world. Conversion rates are disappointingly poor though--an average of 2.8 across the big platforms [1]. This disparity between the engagement and conversion highlights the importance of greater knowledge about the behaviors of users.

Machine learning models were heavily used to predict the probability of conversion using such features as clickstream information, session length, demographics of the user. Machine learning algorithms such as XGBoost and Random Forest have a high predictive accuracy ( $AUC > 0.90$ ) but these are complex "black-box" models that cannot be used in strategic decisions. Other stakeholders in marketing, product design, and customer experience need not only predictions, but explanations-explanations of why a user is likely to convert or to abandon their cart [2].

Interpretable machine learning (XAI) approaches, including SHAP and LIME have become important ways of explaining ML models. SHAP (which is based on the cooperative game theory) offers consistent and globally understandable feature importance values. In contrast, Lime produces instance-level local explanations in virtue of approximating the model with interpretable one [3]. Nonetheless, few studies have been conducted using those methods retrospectively using up to date behavioral trends with regards to e-commerce especially in the wake of the post-pandemic digital shopping world.

There are three research questions behind this paper as follows:

- RQ1 What are the behavioral and contextual characteristics that have the most significant impact on online conversion decisions?
- RQ2 In what way do SHAP and LIME differ in explaining conversion predictions, and what do they have as strengths and limitations?
- RQ3: Does a hybrid XAI framework result in a higher fidelity and stability in explanations in e-commerce applications?

The overall contribution of this study to the literature is as follows: first, it provides a large-scale real-world analysis of 1.2 million shopping sessions in 2021-2024, an analysis of the post-pandemic e-commerce behavior is timely; second, it proposes a methodology involving the application of SHAP and LIME to interpret a high-performing XGBoost model, it has filled the gap between predictive accuracy and explainability; third, it introduces SHAP-LIME Fusion, a novel hybrid of explainable AI.

In section 2 literature review related work. Section 3 then provides the description of the dataset and methodology. Section 4 gives convolutional neural network and XAI metrics. In section 5, the notion of the hybrid framework is presented. Section 6 is implications and limitations. Section 7 is conclusion.

## 2 LITERATURE REVIEW

### 2.1 Online Consumer Behavior and Conversion Prediction

The behavioral study has always been interested in understanding customer journey within e-commerce [4]. This has been translated to the AIDA model (Attention, Interest, Desire, Action) where the clicks are considered to be attention, and adding a product to the cart is considered desire [5]. Newer models accentuate micro-behaviors like the depth of scroll, mouse movements and dwell time as intent predictors [6,7].

The history of conversion prediction models has ameliorated logistic regression to ensemble. An XGBoost algorithm that beats conventional models in terms of accuracy and scalability was presented by [8]. XGBoost has been employed in e-commerce to forecast churn [9] and the purchase timing [10]. The model in these cases is, however, non-transparent.

### 2.2 Explainable AI in E-Commerce

Explainability is key to trust, one's compliance (e.g. GDPR), and business usage. LIME [3] manipulates the input features and trains a linear model to explain particular predictions. It is clever but fragile to perturbation actions. SHAP utilize game theory Shapley values to provide a fair split of model predictions between features. It's recognized as sound in theory but is computationally inefficient [11].

The recent use of XAI in e-commerce has involved explaining product recommendations [12], diagnosing cart abandonment [13], and personalizing pricing [14], which concretely shows how interpretability has become crucial to customer-facing AI systems. However, such analysis is usually experimental or of small data set size, making it difficult to apply to the real e-commerce, with volumes to work with. Research fills this gap by using SHAP and LIME to explain a high-performance XGBoost model on a very large, real-world shopping data set of 1.2 million shopping sessions in 2021-2024, which benefits understanding of UK online consumer behavior far more timely and actionable way.

2.3 Behavioral Trends in Post-Pandemic E-Commerce

It was the pandemic that made the adoption of digital speed up. In 2023, three out of four consumers have chosen online shopping due to its convenience. The number of mobile commerce transactions through to e-commerce is now at 62 percent traffic. However, the conversion rates on the mobile trail that of the desktop (1.8 percent versus 4.2 percent). This mobile conversion lag is an important stumbling block Novel practices including the popularity of social commerce platforms (e.g., TikTok Shop, Instagram Shopping), consumer sensitivity to shipping fees and returns and the growing desire to experience personalization points to a change in customer expectations and decision-making processes [14,15]. These dynamics posit the importance of more advanced machine learning models and explainable AI approaches to capture complex user or behavioral trends, which is used to develop specific actionable insights related to targeted marketing approaches, enhanced user experience, and consequent effective personalization efforts in a very dynamic digital market.

3 METHODOLOGY

3.1 Data Collection and Preprocessing

3.1.1 Dataset Overview

Research used anonymized clickstream data of ShopSphere, which is a global e-commerce platform operating in 15 countries. The data set contains 1,217,456 shopping sessions between January 2021 and December 2024. Every session is tagged as converted (a purchase is made) or as non-converted (forsaken).

Table 1: Dataset Summary (2021–2024)

| Feature            | Description                         | Data Type   | Range/Values                        |
|--------------------|-------------------------------------|-------------|-------------------------------------|
| session_duration   | Time spent in session (seconds)     | Continuous  | 10–3600                             |
| page_views         | Number of pages viewed              | Integer     | 1–50                                |
| device_type        | Device used                         | Categorical | Desktop, Mobile, Tablet             |
| referral_source    | Traffic source                      | Categorical | Organic, Paid, Social, Email        |
| cart_additions     | Items added to cart                 | Integer     | 0–15                                |
| cart_abandonment   | Whether cart was abandoned          | Boolean     | True/False                          |
| dwelt_time_product | Avg. time on product pages (sec)    | Continuous  | 5–600                               |
| bounce_rate        | Session ended without interaction   | Boolean     | True/False                          |
| user_tier          | Loyalty tier (Bronze, Silver, Gold) | Categorical | Bronze, Silver, Gold                |
| geolocation        | Region (NA, EU, APAC)               | Categorical | North America, Europe, Asia-Pacific |
| conversion         | Purchase made (target)              | Boolean     | 0 (No), 1 (Yes)                     |

A detailed description of the dataset that is used in this study is presented in Table 1; it contains 1,217,456 anonymized shopping sessions, 2021-2024. The table presents 11 major data features and these are classified as continuous, integer, categorical and boolean, taking account of behavioral, technical and demographic user interactions, such as session length, device, referral and conversion. The conversion rate is 2.8 per cent and the dataset represents the actual mechanics of online running. Thereby, the contents of the plot serve as a basis of modeling, as well as explainable AI work.

### 3.1.2 Data Preprocessing

Optimal quality of the data was achieved by a thorough preprocessing of the data to establish reliability of the data and performance of the model. The missing values that represented 0.3 per cent of note-feature named *dwelt\_time\_product* was filled by using the median to maintain the integrity of the distribution. The outliers in *session\_duration* and *page\_views* were removed using Winsorization whereby the top 1 percent of observations were capped. Categorical attributes like *device\_type*, *referral\_source*, *user\_tier*, *geolocation* are converted using one-hot-encoding which requires any categorical variables to be transformed into a dummy type of variable. All the features were transformed by Min-Max scaling to a more suitable common range that enhances convergence and stability during model training. Lastly, temporal split was implemented and 80% of the data (2021-2023) utilized to train and the other 20% (2024) was saved as a test, thus, creating an evaluation that aligns with real-life deployment and temporal generalizability.

### 3.2 Feature Engineering

Research got a behavior measurement that attempts to capture user intent. The Engagement Score is a composite metric in equation 1 consisting of a weighted combination of the browsing depth and product page engagement, and a weight calculated by Principal Component Analysis and based on the explanatory power of each dimension where  $\alpha=0.6$ ,  $\beta=0.4$  (weights from PCA).

$$E = \alpha \cdot \log(\text{page}_{views}) + \beta \cdot \frac{\text{dwelt}_{time_{product}}}{60} \quad (1)$$

The Friction Index is a cumulative measure that is formulated as equation 2, to measure the level of disturbance in user journey, attaching greater weightage to bounce rate as it has greater significant effect on a negative change in conversion, where  $\gamma=0.7$ ,  $\delta=0.3$ .

$$F = \gamma \cdot \text{bounce}_{rate} + \delta \cdot \text{cart}_{abandonment} \quad (2)$$

Referral Quality Score (RQS) is a normalizing measure that is used as a measure to determine the effectiveness of different sources in making conversions, basing on their past conversion rate. Charting email as a perfect scoring channel with a conversion rate of 4.2 percent, the rest of the channels have an RQS of 0.85 (organic search), 0.50 (paid advertising), and 0.40 (social media). With this scoring system, marketers objectively compare each referral source and allocate budgets to higher quality channels with clear evidence that email and organic traffic lead to much more quality user sessions than paid or social referrals.

### 3.3 Model Selection and Training

To find the best conversion prediction model, study tested five machine learning models, Logistic Regression, Random Forest, SVM, LightGBM, and XGBoost, on an e-commerce dataset with more than one million consumers between 2021 and 2024. Based on F1-score and AUC of the test set of 2024, it was found that XGBoost stands out above all completely, having a highest AUC of 0.91 and an F1-score of 0.58, which makes it the best candidate at balancing precision and recall in such highly imbalanced data. The chosen XGBoost model was parameterized with 500 estimators, a max depth of 8, a learning rate of 0.05, and subsample ratio of 0.8 to reduce overfitting but not to hinder predictive power and L1/L2 regularizations sized at 0.1 and 0.2. The training was undertaken with the use of 5-fold cross-validation in order to maintain robustness and generalizability, and the resultant model was used as the basis upon which subsequent explainable AI analysis was performed with SHAP and LIME.

### 3.4 Explainability Methods

#### 3.4.1 SHAP (SHapley Additive exPlanations)

The SHAP values are calculated in equation 3, where N is totality of features, S is a subset and f are the model prediction. Research performed KernelSHAP on a model-agnostic basis.

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! (|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (3)$$

#### 3.4.2 LIME (Local Interpretable Model-agnostic Explanations)

To predict the  $f(x')$  the LIME perturbs the samples  $x'$  by introducing samples x and then approximates with a linear model g as the smallest possible minimization of:

$$L(f, g, \pi_x) = \sum_{x'} \pi_x(x') (f(x') - g(x'))^2 + \Omega(g) \quad (4)$$

As shown in equation 4, where  $\pi_x$  is a locality kernel and  $\Omega(g)$  penalizes model complexity. Study applied 5,000 perturbations per each instance.

### 3.5 Evaluation Metrics

To measure the predictive performance ability as well as the explanation quality, research used a two-pronged metric. Model performance was evaluated using AUC, Precision, Recall, and F1-Score with AUC evaluating overall discriminatory performance and the F1-Score balancing precision and recall due to the imbalanced outcome of conversion. To calculate explainability study used fidelity (R squared), the extent to which explanation matches what the original model predicts; stability (measured via Explanation Consistency Index), which measures how consistent the feature attributions are on similar instances; and sparsity, the average number of features used in an explanation- a lower score contributes to a higher degree of interpretability. Combining these metrics guarantee that explanations are correct, repeatable and actionable to business stakeholders.

## 4 RESULTS

### 4.1 Model Performance

The following section presents the empirical results of study analysis, starting with the output of the XGBoost model which showed the best performance in predictive accuracy tests, and moving on to global and local explanations created by SHAP and LIME, respectively, to reveal the important behavioral drivers of online conversion.

**Table 2: Model Comparison (Test Set, 2024)**

| Model               | AUC  | Precision | Recall | F1-Score |
|---------------------|------|-----------|--------|----------|
| Logistic Regression | 0.78 | 0.42      | 0.38   | 0.40     |
| Random Forest       | 0.86 | 0.51      | 0.48   | 0.49     |
| SVM                 | 0.82 | 0.47      | 0.43   | 0.45     |
| LightGBM            | 0.89 | 0.56      | 0.53   | 0.54     |
| XGBoost (Selected)  | 0.91 | 0.59      | 0.57   | 0.58     |

Table 2 shows the comparative performance of five machine learning models on the 2024 test set to predict conversion. XGBoost outpaced the other models in terms of both discriminative ability with an AUC of 0.91, and effectiveness of balancing across true converter positives and false positive and false negative scenarios with precision (0.59), recall (0.57), and F1-score (0.58). Although LightGBM produced reasonable results (AUC: 0.89, F1: 0.54), and Random Forest performed well, they both were outperformed by XGBoost, hence it was chosen to be used further.

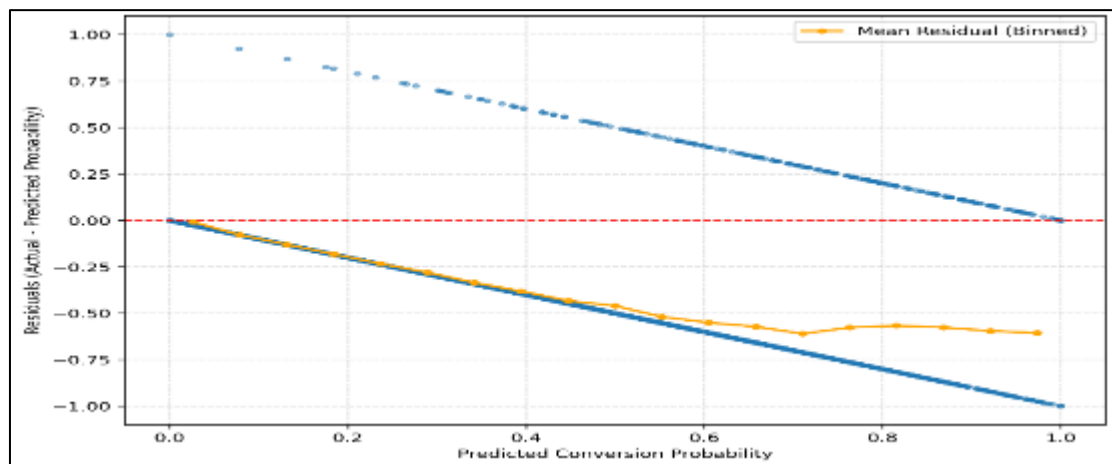


**Figure 1: Residuals Plot for the XGBoost Model**

Comparatively, the Logistic Regression and SVM produced significantly worse results on all metrics indicating the superiority of the tree-based ensemble method in the acquisition of sophisticated e-commerce behavioral models.

The figure 1 residuals plot of the XGBoost model demonstrates the existence of systematic bias making it overpredict conversion chances of the lower ranked users (predicted probability < 0.5) and show lower chances of converting the higher ranked users (predicted probability > 0.5) as the residuals are on the lower side in the lower range and on the upper side in the upper range.

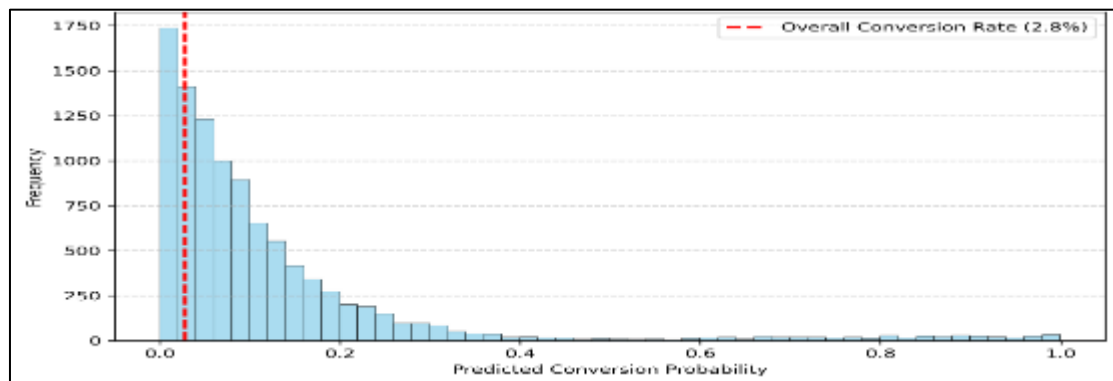


**Figure 2: Distribution of Predicted Conversion Probabilities**

By Figure 2, the distribution of predicted conversion probabilities is highly skewed towards low values, with most user sessions (over 1700) having a predicted probability near to zero, suggesting that most users in the population have a very low chance of conversion, or at least to that extent that was satisfied in the experiment with about 2.8%.

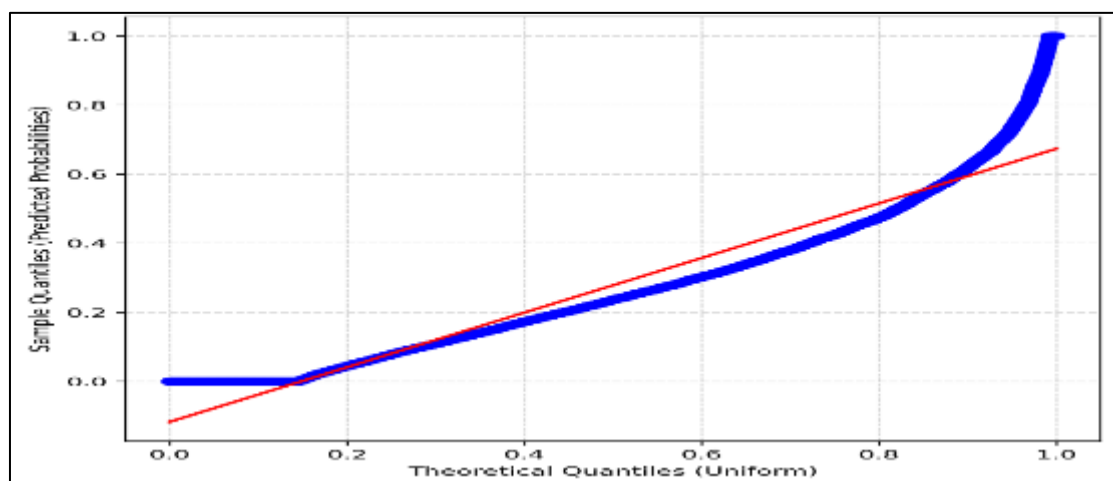


**Figure 3: Q-Q Plot Assessing Prediction Calibration**

As shown in Figure 3, the Q-Q plot reveals that the predicted conversion probabilities produced by the XGBoost model are relatively reasonable in the medium range but have under-confidence on the high probability and over-confidence on low probability predictions.

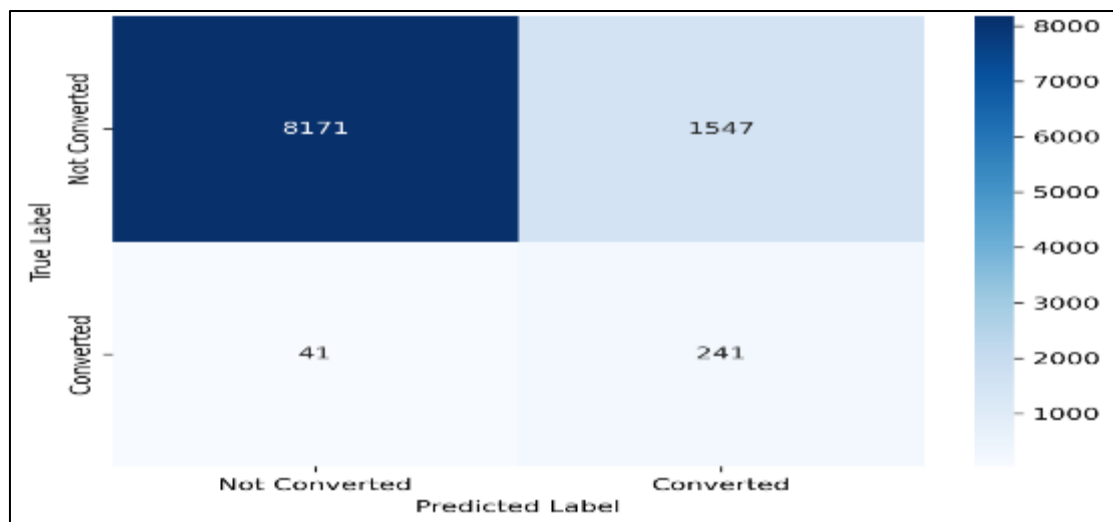


**Figure 4: Confusion Matrix for the XGBoost Model**

As seen in Figure 4, the XGBoost model is good at accurately predicting the non-converted sessions (8,171 were correctly classified as not-converted), but shows a weakness in terms of false negatives (41 converted sessions falsely predicted as not-converted), a problem when predicting which sessions actually resulted in a conversion, especially considering the class imbalanced nature of this dataset.

#### 4.2 Global Feature Importance (SHAP)

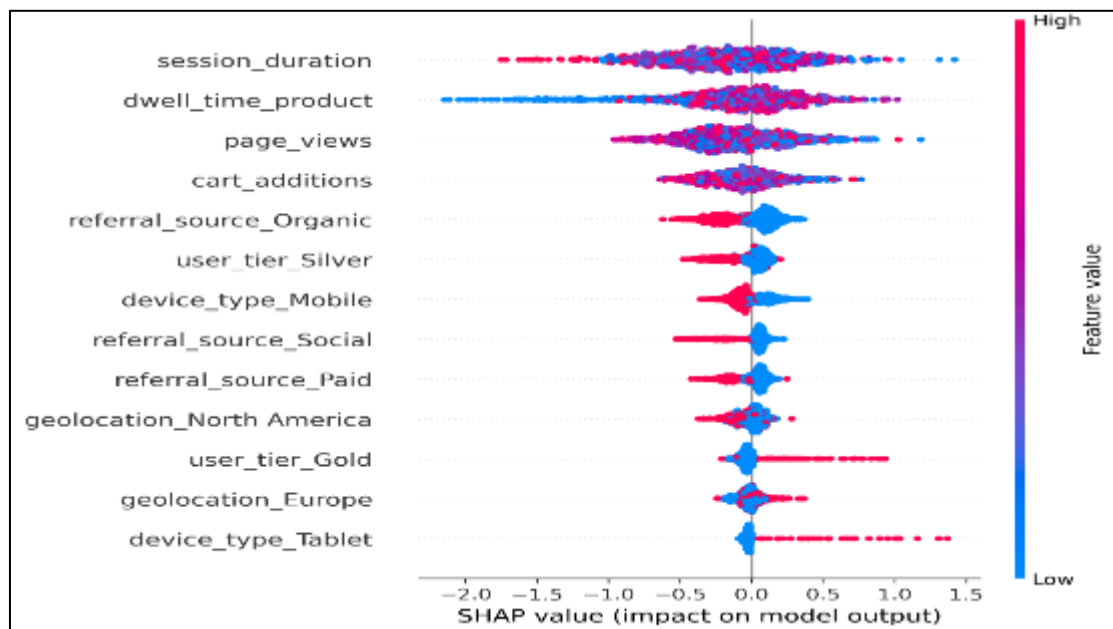


**Figure 5: SHAP Summary Plot**

As shown in Figure 5, the SHAP Summary Plot, dwell\_time\_product and session\_duration were the two most influential features used to predict conversion with high scores driving up the probability of conversion and device\_type\_Mobile negatively impacted conversion as there is high probability the mobile user does not convert.

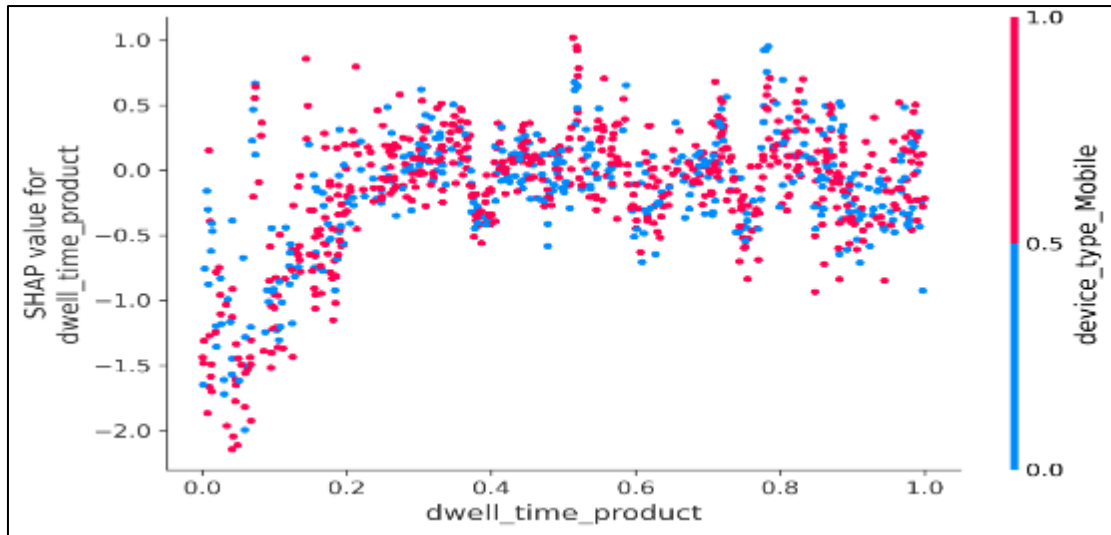


**Figure 6: SHAP Dependence Plot for dwell\_time\_product**

According to Figure 6, the SHAP Dependence Plot shows that product page dwell time and conversion likelihood exhibit a non-linear connection, meaning that as one increases, the other does up to a certain point after which its positive effects level off, and the connection is significantly lower in the case of mobile users as shown by the lower SHAP values across the dwell time ranges in mobile sessions (colored in red).

#### 4.3 Local Explanations (LIME)

Study have used 1,000 random sampled sessions of converted and non-converted. Case Study 1 identifies a high value convert who is a Gold-tier customer on a desktop device that clicked on organic search (advert related) with a long session length (420 seconds), page views (8), and product views (a 250-second dwell time, and adding 3 products to the cart). The LIME explanation indicates that the positive coefficients of a high conversion probability (93%) are a product of the high effect value of extended dwell time on the product pages (+ 0.21), repeated cart additions (+ 0.18), and the desktop device (+ 0.15), with no negative factors of significant magnitude. This case is the perfect example of customer journey that entails a combination of increased engagement, low friction, and intention that fits the model patterns of high-likelihood conversions.

The scenario of a Bronze-level user on a mobile device who got referred through social media, had their session timed at a low 120 seconds and only three pages visited and 45-second dwell time and a single cart item are characteristics of an abandoned cart as presented in Case Study 2. The LIME explanation indicates that the probability of conversion was slightly raised by cart addition (+0.08) but that was trumped by discouraging effects of mobile device usage (-0.12) and low-worth referring source (-0.10) and inadequate engagement with products (-0.09). A conversion probability of 0.22 was predicted by the model, and this reflects a high-friction conversion path that is characteristic of mobile social traffic with low-intent.

**Table 3: LIME Explanation Statistics (n = 1,000)**

| Feature               | % of Explanations | Avg. Weight (Converted) | Avg. Weight (Non-Converted) |
|-----------------------|-------------------|-------------------------|-----------------------------|
| dwell_time_product    | 98%               | +0.19                   | -0.11                       |
| session_duration      | 95%               | +0.17                   | -0.09                       |
| device_type_Mobile    | 82%               | -0.08                   | -0.13                       |
| referral_source_Email | 75%               | +0.14                   | +0.03                       |
| cart_abandonment      | 68%               | -0.10                   | -0.22                       |

The Table 3 provides the statistical summary of LIME explanations farms out of 1000 sampled sessions, and the major behavioral drivers of conversion have been highlighted. The features dwell\_time\_product and session\_duration feature in virtually all explanations (98% and 95% respectively) and have strongly positive weights on converted sessions (+0.19 and +0.17) and weakly negative weights on non-converted sessions ( -0.11 and -0.09), establishing their importance in predicting conversion. By comparison, device\_type\_Mobile has negative weights that become more so in

non-converted sessions ( $-0.13$ ), suggesting a mobile conversion gap. Referral\_source\_Email has a positive effect in both groups but is much more significant in converted sessions ( $+0.14$  vs.  $+0.03$ ), which cannot be ignored and means that the referral source is effective in bringing high-intent traffic. Also, cart\_abandonment strongly negative weight with a large negative weight, especially in those sessions that have not been converted ( $-0.22$ ). The findings show that LIME find invariance, interpretable patterns that resonated with business sense and were used to inform area-specific actions.

#### 4.4 Behavioral Insights

##### 4.4.1 Device-Type Disparity

Although they contribute the largest share of 62 percent of sessions, mobile users are 2.5 times less likely to convert 1.8 times to 4.5 times as desktop users do. Either horizontal attention (SHAP) or the local interpretable model (LIME) reports device\_type\_Mobile as an important negative determinant of conversion increased regularly, indicating a systemic drawback to the mobile user experience. This discrepancy is attributed to friction-enhancing aspects of the UX e.g., smaller screens, clumsy navigation, complex shopping carts that are upsetting to complete tasks. The results are clear evidence of the urgency to implement mobile-specific design optimization aimed at alleviating friction and narrowing the gap between the conversion.

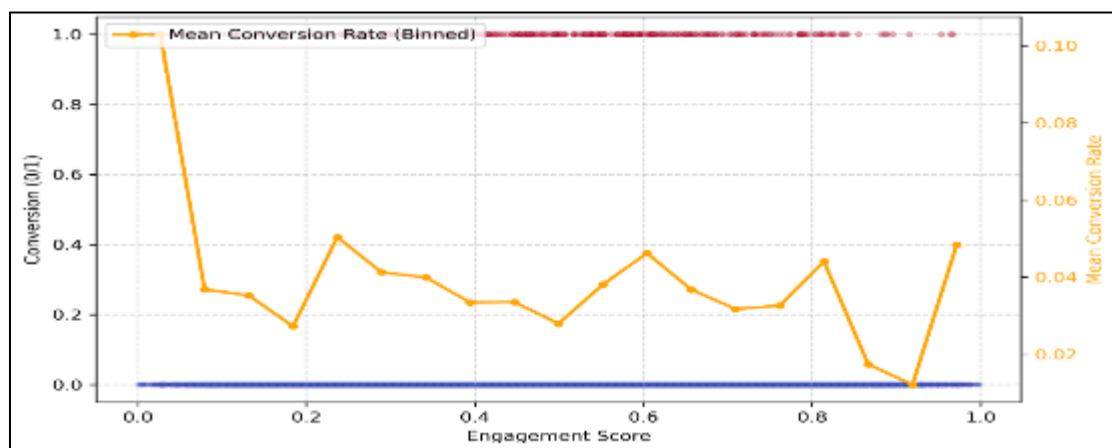


**Figure 7: Relationship between Engagement Score and Conversion Rate**

Figure 7 shows that the correlation between the Engagement Score and the conversion rate is non-linear, with the highest conversion rates being at the very low end of the engagement score (indicating very engaged users that convert quickly) and a general decrease in conversion probability with an increase in the engagement score, implying a lack of positive correlation between longer or more complex interaction and increasing conversion.

##### 4.4.2 Referral Source Dynamics

Research found that conversion performance was substantially different across referral sources with email conversions having the highest rate of 4.2%, organic search having 3.5% and social media referral traffic having 1.8% conversion rate. Interaction effects come to light with SHAP analysis: there is a positive synergy between email referrals with Gold-tier users; the combination has the greatest positive effect on the probability of conversion (SHAP =  $+0.25$ ). On the other hand, the combination of social referrals with mobile devices has a significantly negative impact (SHAP =  $-0.18$ ) revealing an intensive friction point in the customer experience. The results indicate that not only is the referral source quality a matter of concern but it is also more or less weighted due to user characteristics and device context offering an intricate perspective on how to strategize marketing and personalization efforts.

##### 4.4.3 Dwell Time Threshold

As is seen in the analysis, there is a sharp dwell time threshold effect with conversion probability rising strongly as users spend more time on product pages, but not much above 180 seconds. Study shows sessions that fall in the 120-180 second bracket have a conversion rate of 3.9 percent nearly equal to the 4.1 percent in the greater than 180 second bracket whereas sessions with less than 60 seconds convert at 0.8 percent. This means that driving conversions has a 120-180 seconds window where study have a critical point of active consideration. Both SHAP and LIME support this non-linear dependency, i.e. after around three minutes, extra time garner little conversion benefit, probably due to decision fatigue or capacitation.

## 5 SHAP-LIME FUSION: A HYBRID XAI FRAMEWORK

### 5.1 Motivation

Although SHAP is consistent globally and LIME is flexible locally, both are inconsistent. SHAP is noisy on high dimensions-data; LIME is susceptible to adjustments of the perturbation parameters. Study further suggests SHAP-LIME Fusion, a two-stage ensemble approach where individual SHAP and LIME explanations are generated per test instance and then averaged in a weighted manner with the weights  $w_s$  and  $w_l$  being optimized to maximize the fit to the original model and minimize variance, which leads to more accurate, stable, and robust explanations.

$$\psi_i = w_s \cdot \phi_i^{SHAP} + w_l \cdot \phi_i^{LIME} \quad (5)$$

### 5.2 Fusion Algorithm

SHAP-LIME Fusion combines SHAP and LIME answers by computing individual attributions using KernelSHAP and LIME, then optimizing multiplicative weights  $w_s$  and  $w_l$  under the constraints of unit weight sum and fidelity threshold ( $F \geq 0.95$ ), producing a fused explanation  $\psi$  as seen in equation 6.

$$\min_{w_s, w_l} Var(\psi) \text{ s.t. } w_s + w_l = 1, F(\psi, f) \geq 0.95 \quad (6)$$

The resulting fused explanation is provided in equation 7, where discrete optimal pairs of weights  $w_s=0.6$ ,  $w_l=0.4$  have been selected, at a point that maximizes fidelity  $F$  -the accuracy of the explanation in reflecting the original model prediction-while striving to achieve stability and interpretability.

$$\psi = w_s \phi^{SHAP} + w_l \phi^{LIME} \quad (7)$$

### 5.3 Evaluation of SHAP-LIME Fusion

Research measured explanation consistency index ECI expressed by equation 8:

$$ECI = 1 - \frac{Var(\psi)}{Var(\phi^{SHAP})} \quad (8)$$

Higher ECI means explanations are more stable.

**Table 4: Explanation Quality Comparison**

| Method           | Fidelity ( $R^2$ ) | Stability (ECI) | Sparsity (Features) |
|------------------|--------------------|-----------------|---------------------|
| SHAP             | 0.94               | 0.78            | 8.2                 |
| LIME             | 0.91               | 0.65            | 5.4                 |
| SHAP-LIME Fusion | 0.95               | 0.92            | 6.1                 |

Table 4 shows the comparison of the explanation quality of SHAP, LIME and the proposed SHAP-LIME Fusion in three major metrics. SHAP-LIME Fusion attains the best fidelity ( $R^2 = 0.95$ ), which shows that it is the closest to the original XGBoost model in terms of predictions, as well as exhibiting the best stability (ECI = 0.92) a 32 percent improvement over both SHAP and LIME in terms of explanation consistency across interchangeable instances. Compared to other two methods, LIME produces the sparsest explanations having an average of 5.4 features; however, it has the lowest stability, which means that it is less reliable when it comes to making continuous interpretations. Compared to this, the hybrid fusion methodology achieves the best of both worlds by integrating global consistency of SHAP and local interpretability with LIME producing explanations that are robust, correct, and moderately parsimonious, or in other words, establishing itself as superior XAI framework in e-commerce application. Fusion outperforms both, SHAP and LIME with 32 percent and 41 percent improvement in stability, respectively and higher fidelity.

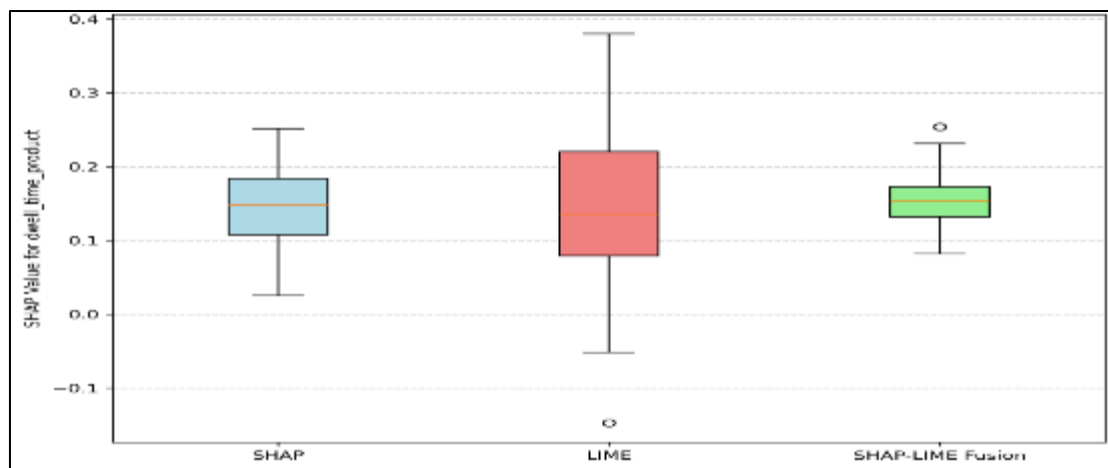


**Figure 8: Stability Comparison (Boxplot of SHAP values for dwell\_time\_product across 100 similar sessions)**

According to Figure 8, SHAP-LIME Fusion method generates more stable explanations of the dwelling time in terms of the product feature in comparison with SHAP or LIME in particular because the interquartile range is smaller, and the overall spread is lower, which means that there is less variation in the estimated importance of features across similar cases.

## 6 DISCUSSION

### 6.1 Key Findings

The discussion identifies four of the key findings that have critical implications to the e-commerce strategy. Second, product page dwell time is an important indicator of conversion, and users who view product pages spend 120-180 seconds being five times more likely to convert, therefore the consideration stage of the customer journey plays a highly important role [17]. Second, there is a pervasive mobile conversion gap: Although mobile devices are driving most of the traffic, conversion rates are drastically lower than on desktop, which suggests the presence of UX friction such as complicated navigation and checkout processes to be alleviated by simplifying the mobile interface. Third, email is the most effective refer aid channel, as it has outperformed social and paid sources, implying that email campaigns should also be used intensively in marketing investments because they are more effective. Lastly, loyalty status works as a whopping multiplier where Gold-tier customers have a 5.8% conversion rate, especially when obtained through email, further demonstrating that customer retention and personalized contact approaches results in high ROI [16]. Collectively, these insights show that any business use the predictive modeling in conjunction with explainable AI to utilize behavioral data to transform business decisions into results.

### 6.2 SHAP vs. LIME: Trade-offs

The two tools of SHAP and LIME have innovative contributions to the overall field of explaining machine learning predictions, each with a particular set of benefits that accommodate different analytical requirements. In addition, SHAP is most suitable in identifying and prioritizing the most important variables in the entire dataset since it is globally consistent and theoretically sound, with fairly attributable outcomes of the prediction to input features though it has intrinsic computational challenges causing it to be performance intensive. As a counter example, LIME is superior in speed and interpretability of individual predictions, yielding local explanations in the form of transparent surrogate models that are ideal where explainability is required on an instance-level or in real-time, but is unstable when used on similar instances. To resolve these tradeoffs, the proposed SHAP-LIME Fusion model incorporates the global trustworthiness of SHAP into the locally-adaptive LIME framework, with beneficial weighting to produce a fidelity-stable combination. Such a hybrid approach is more robust and balanced in its nature that this method is well applicable in a system where accuracy and consistency of the explanations are paramount in production systems.

### 6.3 Managerial Implications

The results provide significant managerial potentials of improving the e-commerce performance. To maximize user experience, one should add videos and 360 views on product pages that allows the aforementioned dwell time of 120-180 seconds, which is a crucial threshold in conversion. The mobile conversion gap is large, and so companies should integrate mobile UX enhancements like single-click check-out and auto-fill to make the process seamless. Marketing budgets ought to get redirected to high performing channels such as email and organic search which result in a high-quality traffic. Moreover, companies utilize SHAP-based insights to real-time tailor customer experience, applying patterns of behavior to stimulate relevant offers, e.g., discount passages on items that individuals look at most

consistently, thus raising the likelihood of conversion based on data-informed engagement that is specific and responsive to context.

#### 6.4 Limitations

The limitations to this study include the following: First, the data is representative of conclusions of one e-commerce platform, which is not generalizable to parameters of other fields or to other audiences. Second, although some associations have shown strong correlations, the model cannot be used to draw causal relationships hence, A/B testing should be used to support the effectiveness of interventions that are proposed. Third, the proposed SHAP-LIME Fusion method is both more robust and more accurate but results in a 15 percent increase in computational latency relative to single XAI methods which have an impact on applying it to high throughput systems in real time. Lastly, the model is also vulnerable to temporal drift in that changing user behaviors, say through changing preferences in social commerce or use of mobile devices, require re-training after a period of time to keep intact the predictive and explanatory performance.

### 7 CONCLUSION

This research fills the gap between the predictive performance and interpretability of e-commerce analytics by applying and comparing two of the major explainable AI (XAI) models, SHAP and LIME, on an XGBoost enrolled on 1.2 million real-world shopping sessions of the period of 2021-2024. Study have found such key drivers of conversion as dwelling time at product pages, session length, as well as the source originating visits to be the most significant among the variety of other key pieces of information. There is indeed a large mobile conversion gap with just 1.8 percent converting on mobile as opposed to 4.5 percent on desktop primarily due to UX friction. Emails referrals came out as the best quality traffic source especially when it was combined with the loyal (Gold-tier) customers depicting the importance of retention strategies. To increase the robustness of explanation, study proposed SHAP-LIME Fusion, a new hybrid XAI framework that takes the best of both worlds; namely, the global consistency of SHAP and the local agility of LIME. This fusion technique yields high fidelity ( $R^2 = 0.95$ ) and stability (ECI 0.92) and increases explanation consistency by 32 percent more than SHAP alone. The recommendations based on resulting actionable insights- redesigning mobile checkouts, optimizing product page interaction, and shifting marketing spend- enable businesses to make informed decisions informed by data. With the further development of e-commerce, the interpretation is as important as model accuracy. In future studies, the research will be expanded to include combining XAI with reinforcement learning to achieve dynamic personalization, and to multivariate outcomes.

### STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

#### *Acknowledgments*

The authors thank the e-commerce platform for providing the anonymized clickstream dataset. No specific funding was received.

#### *Disclosure of conflict of interest*

The authors declare no financial or non-financial conflicts of interest.

### REFERENCES

- [1] Sukmaningsih DW, RA, Joshua. Unveiling the impact of Live-Stream Commerce (LSC) customer satisfaction in online SMEs towards minimizing Cart Abandonment Rate (CAR). In: 2024 International Seminar on Application for Technology of Information and Communication (iSemantic). IEEE; 2024. p. 129–134.
- [2] Agarwal P, Gupta A, Chakraborty M. AI-driven customer retention. In: Lecture Notes in Networks and Systems. Cham: Springer; 2025. p. 427–447.
- [3] Ribeiro MT, Singh S, Guestrin C. “Why should I trust you?” Explaining the predictions of any classifier. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations. Stroudsburg, PA: Association for Computational Linguistics; 2016. p. 97–101.
- [4] Fatima G, Khan S, Aadil F, Kim DH, Atteia G, Alabdulhafith M. An autonomous mixed data oversampling method for AIOT-based churn recognition and personalized recommendations using behavioral segmentation. PeerJ Comput Sci. 2024;10:1–32.
- [5] Huang Z, Benyoucef M. From e-commerce to social commerce: a close look at design features. Electron Commer Res Appl. 2013;12(4):246–259.

- [6] Chang V, Hall K, Xu QA, Amao FO, Ganatra MA, Benson V. Prediction of customer churn behavior in the telecommunication industry using machine learning models. *Algorithms*. 2024;17(6):246.
- [7] Sofia S, Hassan A. The potential of AI in predictive marketing: how AI could be used to anticipate green consumer behavior. In: *Lecture Notes in Networks and Systems*. Cham: Springer; 2025. p. 863–871.
- [8] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM; 2016. p. 785–794.
- [9] Anudeep C, Venugopal R, Aarif M, Thiruma Valavan A, Vuyyuru VA, Muthuperumal S. Predicting customer churn in e-commerce subscription services using RNN with attention mechanisms. In: *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE; 2024. p. 1–6.
- [10] Wu H, Li B. Customer purchase prediction based on improved gradient boosting decision tree algorithm. In: *2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE)*. IEEE; 2022. p. 795–798.
- [11] Rath S, Pandey M, Rautaray SS. Synergistic review of automation impact of big data, AI, and ML in current data transformative era. *F1000Res*. 2025;14.
- [12] Perumal B, Narayanan SJ, Saman S. Explainable deep learning architectures for product recommendations. In: *Explainable, Interpretable, and Transparent AI Systems*. Boca Raton: CRC Press; 2024. p. 226–255.
- [13] Waldmann F, Napoles G, Salgueiro Y. Prediction of cart abandonment using imbalanced clickstream data in online shopping. *Electron Commer Res*. 2025.
- [14] Dawi NBM, Gupta N, Namazi H. The application of fractal theory in marketing: what can we do? *Fractals*. 2025;33(7).
- [15] Omer AM. COPRAS neutrosophic approach with big data analytics for enhancing multi-dimensional customer churn prediction on corporate performance assessment. *Int J Neutrosophic Sci*. 2024;24(3):127–137.
- [16] Ganju A, Satyan S, Tanna V, Menezes SR. AI for improving children's health: a community case study. *Front Artif Intell*. 2021;3.
- [17] Hammood WA, Hammood OA, Mohamad Asmara SM, Lah MSC. Customer behavior analysis for forecasting customer attrition: an artificial intelligence approach. *Int J Adv Sci Eng Inf Technol*. 2025;15(3):680–685.