

Research Article

Exploring Explainability in Multi-Category Electronic Markets: A Comparison of Machine Learning and Deep Learning Approaches

Suleiman Adamu ¹, Aamo Iorliam ^{2*}, and Özcan Asilkan ³

¹ Department of Information Systems, School of Information Technology and Computing, American University of Nigeria, Yola, Nigeria, e-mail: suleiman.adamu@aun.edu.ng

² Data Science Department, School of Information Technology and Computing, American University of Nigeria, Yola, Nigeria, e-mail: aamoorliam@gmail.com

³ Business Analytics Department, Higher Colleges of Technology, United Arab Emirates; e-mail: oasilkan@hct.ac.ae

* Corresponding Author : Aamo Iorliam

Abstract: Artificial intelligence can change many industries as a global phenomenon. Over the years, this transformation has supported Electronic Markets in reengineering the processes and activities that take place in traditional markets, focusing on improving transaction effectiveness and efficiency. While our dependence on intelligent machines continues to grow, the demand for more transparent and interpretable models equally grows. Thus, explanations for machine decisions and predictions are needed to justify their reliability, which requires greater interpretability and often elaborates the need to understand the algorithms' underlying mechanism. This paper, therefore, proposed models based on Decision Tree (DT), Long Short-Term Memory (LSTM), and an ensemble of the two aforementioned models for improving CLV accuracy, interpretability, and explainability of AI-based models in the multi-category electronic market. An open-source e-commerce Behavior Data from a multi-category store, previously used by similar studies on XAI and CLV, was used in this experiment, ensuring the robustness of the product prediction and explanations and fair comparison. From the results, the models from this study demonstrated remarkable performance in terms of minimal error rates of MAE, MSE, and RMSE, with LSTM outperforming the other models. Regarding explainability and interpretation, the `begin_time` is ranked as the most relevant feature in CLV prediction.

Keywords: Artificial Intelligence; Deep Learning; Electronic Markets; Explainable Artificial Intelligence; Information Systems; Machine Learning; Prediction.

Received: December, 24th 2024

Revised: February, 24th 2025

Accepted: February, 28th 2025

Published: March, 8th 2025

Curr. Ver.: March, 8th 2025



Copyright: © 2025 by the authors.
Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>)

1. Introduction

Machine Learning has recorded exponential growth in research and industrial applications, especially with the success of Deep Learning (DL) and Neural Networks (NN). It has become so crucial that its impact and possible after-effects can no longer be taken for granted. In some fields, minute failure, such as a momentary dysfunction, can result in fatality (in autonomous vehicles) or loss of billions (in electronic markets). As a global phenomenon, artificial intelligence can potentially change many industries. Over the years, this transformation has supported Electronic Markets in reengineering the processes and activities that take place in traditional markets with a major focus on improving transaction effectiveness and efficiency. While our dependence on intelligent machines continues to grow, the demand for more transparent and interpretable models equally grows. In addition, the ability to explain the model generally is now the gold standard for building trust and deployment of Artificial Intelligence (AI) systems in critical domains[1]. Some of these AI systems require a high level of accountability and thus transparency[2]. Explanations for machine decisions and predictions are thus needed to justify their reliability. This requires greater interpretability,

which often elaborates the need to understand the mechanism underlying the algorithms. Unfortunately, the black-box nature of DL is still unresolved, and many machine decisions are still poorly understood[2]. The trust issues between humans and machines, in particular, have necessitated authorities such as the European Commission in 2021 to propose the first legal framework on Artificial Intelligence with the obligations of ensuring transparency, lawfulness, fairness, and a standardized process with minimum requirements of explicability in terms of transparency and explainability[3]. Not only from a regulatory point of view, but according to[4], from a methodological point of view also, the inability to provide an explanation that accompanies specific decisions creates high-level problems such as impediments to user acceptance and lack of accountability.

Considering that companies are changing their approach from product-centric to customer-centric [5], and the high mobility of customers in today's digital age, retaining customers in electronic and all forms of businesses has become a crucial success factor for electronic markets. According to [6] in [7], Customer Relationship Management (CRM) efforts are primarily designed to retain current customers, develop long-term relationships, and attract new ones. For any given business, acquiring a new customer is more expensive than retaining an existing customer [8]. Central to this is the Customer Lifetime Value (CLV), which according to [6] is a critical factor that multi-businesses should leverage as a competitive advantage. As one of the most important tasks for enterprises to maintain customer relationships[9], the CLV helps organizations group customers and identify those who generate the most value over time[10]. Customer Session Duration (CSD) prediction is crucial in the broader strategy for enhancing CLV. Its predictive power on various consumer behavioral aspects has helped businesses implement targeted strategies that drive higher engagement and profitability over time.

Understanding and interpreting the decisions made by AI systems, especially in electronic markets, is crucial, not only in Europe but the world over, that can be facilitated by Explainable Artificial Intelligence (XAI). According to [11], XAI studies and designs approaches, algorithms, and tools that produce human-understandable explanations of AI-based systems information and decisions. This (XAI) would not only ensure compliance with regulatory guidelines but also enable the alignment of the Electronic Market objectives with customer expectations, among others. According to [12], XAI can be categorized based on criteria such as time of model implementation (before or after training), scope (local or global), and applicability (model-specific or model-agnostic). Based on this, [12] distinguishes two types of XAI models, intrinsic and extrinsic. The intrinsic models (such as linear regression, decision tree, logistic analysis, and Grad-CAM) obtain knowledge by restricting the rules of machine learning models, while extrinsic models (such as Model-agnostic Explanations (LIME) and SHAP) refer to the application of interpretation after training, such as Local Interpretable [13]. In this effort, several approaches (of XAI) were proposed that fall in either of the two categories. However, according to [14] the trade-off between these two approaches is on prediction (for intrinsic) and explanation (for extrinsic) accuracy. As such, this study combined these two approaches to produce an improved XAI for electronic markets. Long Short Term Memory (LSTM), which is a type of Recurrent Neural Network (RNN) architecture that combines fast training with efficient learning on tasks that require sequential short-term memory storage for many time steps during a trial [15], and Decision Tree (DT) is a powerful and interpretable machine learning model capable of handling complex, nonlinear data structures that can work with both categorical and numerical features[16].

While remarkable success in AI adoption has been achieved in some industries, the penetration pace in other industries has been slow due to barriers [17] such as privacy concerns, security issues, ethical challenges, and trust issues between humans and machines [18] that are propelled by the major role artificial intelligence plays in electronic markets. Despite the contribution of AI systems in e-commerce, its ethical soundness continues to be a contentious issue, especially regarding the concept of explainability [19]. Considering the contributions made by [6] as the most recent study on XAI and CLV in multi-category e-commerce where DNN outperformed DT and RF, there is a gap as the Long Short-term Memory (LSTM) has been reported by [20] and [21] to outperform DNN (which currently is the state-of-the-art in terms of XAI in multi-category e-commerce performance). Furthermore, based on the presented literature on multi-category e-commerce, and to the best of our knowledge and findings, studies are yet to explore the hybridization of DT, LSTM, with XAI technique as suggested by [22] to overcome the trade-off between prediction accuracy and explainability.

LSTM is a Recurrent Neural Network (RNN) architecture that combines fast training with efficient learning on tasks requiring sequential short-term memory storage for many time steps during a trial [15]. Both decision trees and LSTM have distinct strengths that make them suitable algorithms for this study because they offer complementary approaches that are useful for modeling the complex dynamics of customer interactions. Specifically, the DT is well-suited because of its interpretable, nonlinear decision-making process and flexibility [23] in handling customer session duration and CLV, where multiple factors such as customer segments, purchase history, and browsing behavior are involved. Similarly, the LSTM is well suited for this study because it captures sequential patterns inherent in customer behavior [24] and long-term dependencies [25]. Regarding feature interaction, the DT algorithm inherently highlights key interactions between features, allowing for direct interpretation of feature importance. This is particularly useful when understanding the relationships between features and outcomes, even if predictive accuracy is lower. As an explanation technique, SHAP is based on Shapley's values from game theory [26]. It is an explanation technique that offers a robust, theoretically grounded, and versatile interpretation for complex machine learning models. SHAP captures these complex interactions and allows users to understand the marginal contribution of each feature to the overall prediction [27]. Inspired by these strengths, we employ these algorithms (i.e., DT and LSTM combined with SHAP) to implement the proposed model in this study.

The motivation for this study is to also: (i) facilitate the adoption of AI in the electronic markets domain by producing an XAI model for gaining stakeholders' trust, which is essential considering the black-box nature of AI and the sensitivity of digital transactions on electronic markets, lack of explainability behind AI decisions and role which hinders its adoption, and (ii) comply with regulations and authorities i.e. European commission who deemed it necessary to enact policies for electronic markets, which part of such policy stated that AI systems in the electronic markets must be supported with the feature of explainability.

This paper is structured as follows: it begins with the introduction, which is this section, followed by section two, which presents the literature review; section three, which explains the adopted methodology; and sections four and five, which discuss the results and findings, and conclusions and recommendation for future work, respectively.

2. Review of Related Studies

XAI, as described by DARPA, enhances understandability, manageability, and trust in AI, addressing the non-intuitive, opaque, and incomprehensible nature of machine learning. [28]. With XAI, a chain of reasoning based on the AI's knowledge and inference can be provided to the user, demonstrating why the algorithm has made certain decisions and not others [29]. XAI can be grouped into two techniques. The difference between these two techniques is based on the explainability approach; intrinsically explainable models involve explanation techniques during training, while models with explainability after training are extrinsic [30].

According to [31], extrinsic-based explanations often have varying scopes and purposes. Depending on the research purpose, different techniques have been employed to accomplish the desired goals regarding AI model explanation. The Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP) are two widely used XAI methods with tabular data [32]. As explanation techniques, they highlight how important each feature is to a prediction [26], [33]. Other techniques, such as GLocal to loCAL eXplainer (GLocalX), also exist and have been employed by various researchers. While there are several XAI techniques for extrinsic explanations, the most popular based on the reviewed literature on this study and according to [32] are LIME and SHAP.

Intrinsic explanation techniques such as DT are inherently interpretable ML models, sometimes referred to as intrinsic, glass-box, or transparent models [30]. Several studies have been conducted on this technique, such as Decision Trees (and its variants), Bayesian Rule Lists (BRL), and Generalized Additive Models (GAMs) – that are based on linear or logistic regression.

While some researchers proposed CLV prediction approaches based on single models, others, such as [4], proposed an ensemble model that combines multiple models. To accommodate customers' temporal patterns on CLV in periodic purchasing, [4] combined Sequence-to-sequence (S2S) recurrent neural network and Gradient Boosting Machine (GBM) models. Their study shows that while the GBM obtained better results on MAE, the S2S got

better results in terms of RMSE. In general, the proposed combination of the two models produced the best results regarding both MAE and RMSE. Another study by [34] proposed a general system for personalized CLV prediction in online games based on the combination of LSTM and Seq2one stacked on CNN. Results from their study proved efficient and accurate by significantly outperforming other baseline methods.

Several scholars have researched the CLV to solve various problems. According to [35], CLV helps allocate marketing spending effectively, identify and nurture high-value customers, and mitigate loss exposure. This section presents a literature review on CLV and its associating factors. To address the problem of the ever-changing product catalog, [35] proposed an approach for generating embeddings for long-term prediction tasks based on feedforward neural networks. In a related study on the prediction of CLV, [36] proposed a Zero-inflated Lognormal (ZILN) distribution approach. The approach was used with Deep Neural Network (DNN) to deal with the sensitivity problem regarding a significant fraction of zero-value LTV from one-time purchasers.

While CLV prediction can be based on a set of multifactor influences such as the context of engagement and customer segmentation, Customer Session Duration (CSD) is a common factor that cuts across CLV activities because for each CLV activity, it (1) serves as an indicator of engagement [37] (2) helps e-commerce companies with personalization opportunities [38] (3) helps in resource allocation [39], and (4) helps in modeling customers' long-term behavior patterns [40]. Predictive Analytics Models [41] approach that employs statistical techniques and machine learning algorithms based on historical data to predict customer behavior on its three (3) key metrics – session duration, purchase frequency, and churn rates [42]. Based on this, this study significantly impacts CSD in the multi-category electronic market.

Although contributions have been made in XAI and CLV research, no work has been done on XAI in multi-category ecommerce of electronic items, specifically CLV prediction. This is further supported by the recommendation of [22] in their literature review on XAI in IS where they confirmed that studies are yet to explore the hybridization of intrinsic (such as DT) and extrinsic (such as SHAP combined with LSTM model) XAI technique, and suggested so as a future research direction. Furthermore, the research by [6] only employed the Deep Neural Networks (DNN) with SHAP for explanation. Despite the potential exceptional performance of DT and LSTM, these two algorithms have not been implemented for CLV prediction.

3. Proposed Method

This section elaborates on the methodology employed in this study. The conceptual framework, as shown in Figure 1, is structured into four main stages: data acquisition, pre-processing, prediction modeling, and model evaluation. These stages are designed to systematically guide the research toward achieving an accurate and interpretable CLV prediction in multi-category electronic markets.

This study adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) to ensure a structured and systematic approach to data-driven analysis. CRISP-DM is a widely accepted methodology in data mining and machine learning research, providing a well-defined process for discovering patterns and building predictive models [43]. Although CRISP-DM consists of six standard phases, this study focuses on the four most relevant steps, namely:

- **Data understanding:** This is the first step in defining the data used in this research. The definition is based on the aim and objectives of the study, and it also includes the procedure for how the data would be acquired. The defined data for this study is Multi-category Electronic Market Customer Activity data sourced from e-commerce websites.
- **Data preparation:** This stage focuses on activities to prepare the dataset for modeling. The target at this stage is to identify data and ensure its quality by dealing with any associated problems in the dataset.
- **Modeling:** In this stage, the target identifies and predicts the customer's CLV. It is in this stage that algorithms play a key role. These algorithms are rooted in DL, mathematics, and statistics. Therefore, various Python packages that facilitate the modeling activity are employed to accomplish this study's target.
- **Evaluation:** This step evaluates modeling processes, as shown in Figure 2. The evaluation, which is the testing phase, focuses on evaluating the performance of the proposed

model. The testing process continues repeatedly together with modelling until a satisfactory and reliable performance is achieved.

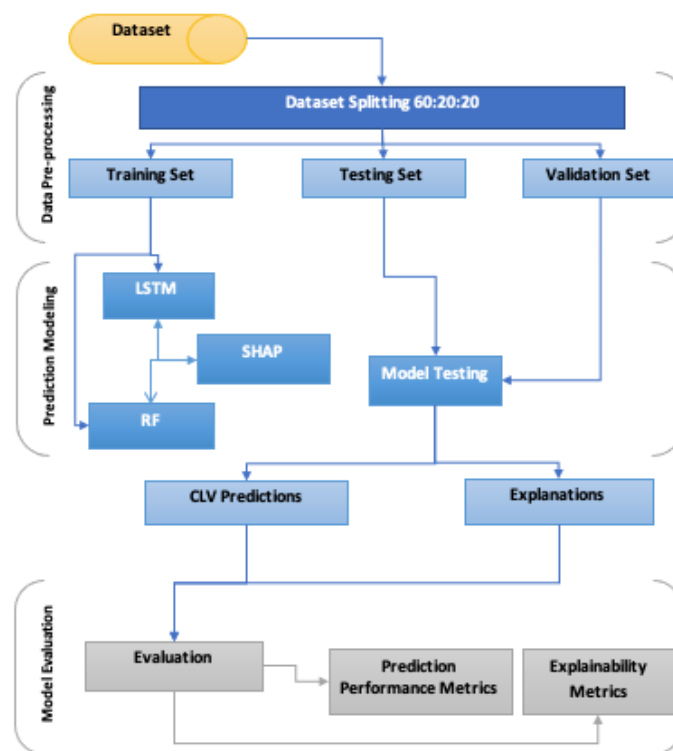


Figure 1. Proposed conceptual approach

3.1. Dataset

To ensure the robustness and reliability of the product CLV prediction and explanation, this study used eCommerce Behavior Data (EBD) from a multi-category store by REES46 (2020) available at <https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store>. The Open-source Customer Data Platform (CPD) dataset, described in Table 1, is publicly available on Kaggle.

Table 1. Tabular description of dataset features.

S/No.	Feature	Description
1	Event_Time	This contains the time of each event by the customer on the website
2	Event_Type	This contains a set of event types that the customer can carry out on the website. The events are four in number. They are: view, cart, remove_from_cart, and purchase.
3	Product_ID	This contains the identification number of the product that the customer performed on an event
4	Category_ID	This contains the category identification number that is unique to each product category.
5	Category_Code	This contains the category code written in English that uniquely identifies each category. An example of category_code in the dataset is a computer notebook.
6	Brand	This contains the brand information for each product. An example of a brand in the dataset is Lenovo.
7	Price	This contains the price for each product.
8	User_ID	This contains a nine-digit unique user identification number for each user on the website.
9	User_Session	This contains an alphanumeric sequence that uniquely identifies each user's session.

Our justification for using this dataset is that it is based on visitor behavior on website and mobile apps, and according to [35] web and app session logs are the most comprehensive dataset available for CLV prediction. Furthermore, using this dataset would enable a fair comparison with previous studies on XIA and CLV. The semantics or interpretation of the dataset for each record is given as "User user_id during session user_session added to shopping cart (property event_type is equal cart) product product_id of brand brand of category category_code with price price at event_time."

3.1.1. Data Pre-processing

The raw EBD dataset, as described in Table 1, requires pre-processing to be fit for the tasks outlined in the objectives of this research. The dataset contains two million five hundred thousand values. Due to our limitation in terms of computation power, we used one hundred thousand (100,000) as a sample for our experimentation in this study. Figure 2 shows the different pre-processing activities performed on the dataset.

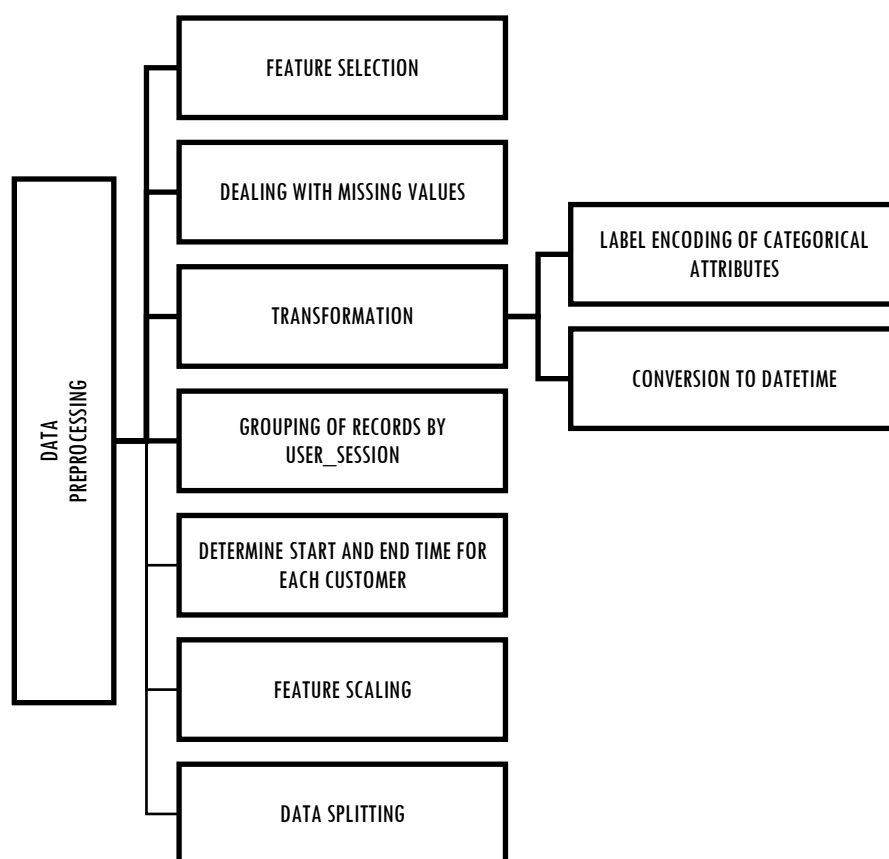


Figure 2. Pre-processing activities on the dataset

3.1.2. Data Splitting

The outcome from the above steps in Figure 2 (from feature selection down to feature scaling) of pre-processing activities makes the data ready for modeling. In this stage, the dataset is split into training, testing, and evaluation, as shown in Figure 3.



Figure 3. Dataset splitting

3.1.3. Input and Target Feature Definition

After the dataset splitting, the input feature for training and the target feature to be predicted by the trained model are all defined. The features are shown in Table 2.

Table 2. Input and target feature specification.

S/N	Feature	Status
1	User_Session	Input Feature
2	Event_Type	Input Feature
3	Category_Code	Input Feature
4	Brand	Input Feature
5	Begin_Seconds	Input Feature
6	End_Seconds	Target Feature

3.2. Modelling

The modeling stage consists of LSTM as a DL approach, DT as an ML (with intrinsic explanation) approach, and SHAP as an extrinsic explanation technique. The justification for the selection is based on each person's reputable performance, as presented in the literature review section.

3.2.1 LSTM

LSTM experimental setup, using the parameters shown in Table 3 to produce the best results for this study.

Table 3. Hyper-parameters configuration of the LSTM model.

Hyper-parameter	Description	Value
Units	The number of LSTM units or memory cells in the LSTM layer. This controls the model's capacity to learn and remember patterns.	50
Activation Function	The activation function is used within the LSTM units. ReLU introduces non-linearity and helps the model learn complex relationships.	ReLU
Epochs	The number of times the entire training dataset is passed through the model during training. More epochs can lead to better learning.	100
Batch Size	The number of samples processed before the model's weights are updated.	32
Loss Function	The function measures the difference between the predicted and actual values during training.	Mean Square Error (MSE)
Optimizer	The algorithm is used to update the model's weights during training.	Adam

3.2.1. Decision Tree

The Decision Tree model experimentation was conducted based on various parameters. The parameters shown in Table 4 yielded the best results for this study and are therefore used.

Table 4. Hyper-parameters configuration of the decision tree model.

Hyperparameters	Description	Values
ccp_alpha	Complexity parameter used for Minimal Cost-Complexity Pruning. The subtree with the largest cost complexity that is smaller than ccp_alpha will be chosen.	0.0
Criterion	The function is to measure the quality of a split.	Squared Error
Max_depth	The maximum depth of the tree. If None, then nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples.	None

Hyperparameters	Description	Values
Max_features	The number of features to consider when looking for the best split	None
max_leaf_nodes	Grow a tree with max_leaf_nodes in the best-first fashion. Best nodes are defined as relative reduction in impurity	None
min_impurity_decrease	A node will be split if this split induces a decrease of the impurity greater than or equal to this value	0.0
min_samples_leaf	The minimum number of samples required to be at a leaf node.	1
min_samples_split	The minimum number of samples required to split an internal node	2
min_weight_fraction_leaf	The minimum weighted fraction of the sum total of weights (of all the input samples) required to be at a leaf node	0.0
random_state	Controls the randomness of the estimator	42
splitter	The strategy used to choose the split at each node. Supported strategies are "best" to choose the best split and "random" to choose the best random split	Best

3.2.3. Ensemble Model

The ensemble model combines the prediction of the LSTM and DT models. As shown in Figure 4, Equation (1), and Algorithm 1, the ensemble model is based on the simple averaging technique, i.e., the combination of the predictions of the LSTM and Decision Tree model divided by 2. Simple averaging is used in the ensemble stage due to its robustness, ease of implementation, and ability to reduce overfitting while improving generalization [44], [45]. The rationale for employing the Simple Averaging technique in our ensemble model is due to the reported competitive performance by simpler ensemble methods against complex ensemble techniques, as reported in [46]. Thus offering a balance between performance and computational efficiency.

Furthermore, unlike complex ensemble techniques that offer additional improvements in accuracy, a well-constructed simple ensemble can often yield significant performance gains compared to the base models without additional computational overhead [44]. Thus yielding competitive performance with far less complexity and equivalence to weighted techniques in ensembles of decision trees [47]. Based on these, the Simple Averaging technique can perform equally or, in some cases, better than more advanced techniques in scenarios where the base models are diverse or complex, which increases the risk of overfitting, as in the current study.

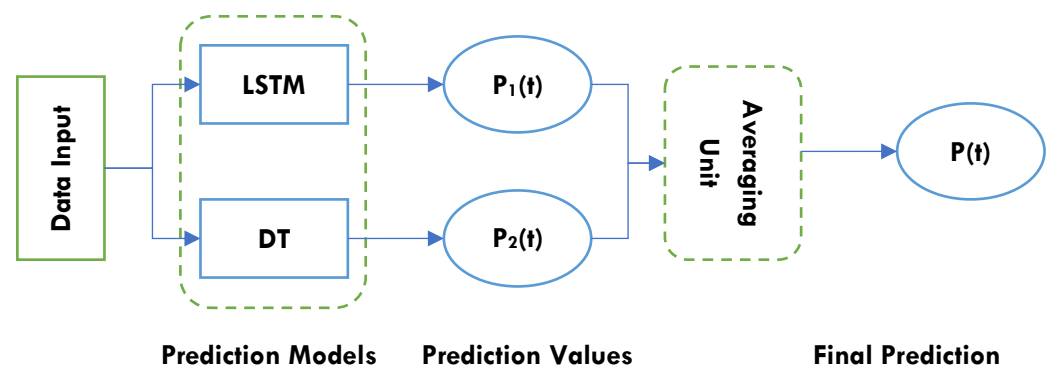


Figure 4. Ensemble Approach.

$$P_t = \frac{1}{N} \sum_{i=1}^N P_i(t) \quad (1)$$

Algorithm 1. Ensemble Prediction

INPUT:

- 1: X_{test} : Test dataset features
- 2: dt_model : Trained Decision Tree model
- 3: $lstm\_model$: Trained LSTM model

OUTPUT:

- 4: $ensemble\_pred$: Ensemble predictions
- 5: BEGIN
- 6: // 1. Decision Tree Prediction
- 7: $dt\_pred = dt\_model.predict(X_{test})$
- 8:
- 9: // 2. LSTM Prediction
- 10: $X_{test\_reshaped} = \text{reshape}(X_{test})$ // Reshape for LSTM input
- 11: $lstm\_pred = lstm\_model.predict(X_{test\_reshaped})$
- 12:
- 13: // 3. Ensemble Prediction (Averaging)
- 14: $ensemble\_pred = (dt\_pred + lstm\_pred) / 2$
- 15:
- 16: RETURN $ensemble\_pred$
- 17: END

3.3. Metrics Evaluation

Model evaluation is performed using three main metrics, namely Mean Absolute Error (MAE), Mean Square Error (MSE), and Root Mean Square Error (RMSE). MAE measures the average absolute difference between actual and predicted values, indicating the level of bias in the model calculated by Equation (2). MSE calculates the average of the squared differences, which is more sensitive to outliers due to a larger penalty for large errors. MSE is calculated by Equation (3). RMSE is the root of MSE, retaining sensitivity to large errors but easier to interpret because its units are the same as the target variable. RMSE is calculated by Equation (4).

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (2)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (4)$$

Where N is the total number of samples in the dataset; y_i is the actual value of sample i ; $\hat{y}_i$ is the predicted value of the model for sample i .

4. Results and Discussion

This section presents the experimental results for predicting duration_seconds, representing customer lifetime value (CLV) based on user interactions (views, cart, purchases). The model performances are evaluated using MSE, RMSE, and MAE, as illustrated in Figures 5, 6, and 7, respectively.

From Figure 5, it can be seen that the LSTM model has the lowest MSE compared to the DT and Ensemble models. On the other hand, the DT model has the highest MSE, followed by the Ensemble model. This result indicates that the best-performing model in terms of MSE rate is the LSTM, followed by the Ensemble and DT models. The difference between the LSTM and the Ensemble model is 0.0007, which is less (in terms of error rate) when compared with that of the DT against the LSTM (0.0030) and ensemble (0.0023). Similarly,

as shown in 6, the LSTM model has the lowest RMSE compared to the other models (DT and Ensemble). Likewise, just like the MSE rate, the DT model recorded the highest RMSE, followed by the Ensemble model. This record implies that the LSTM model in terms of MSE rate, outperforms the DT and Ensemble models. However, based on this result, the LSTM model outperforms the DT much more than it does on the Ensemble model. The specific difference in terms of the RMSE rate between the LSTM and the Ensemble model is 0.0061, which is less when (the LSTM model is) compared with the DT (0.0023).

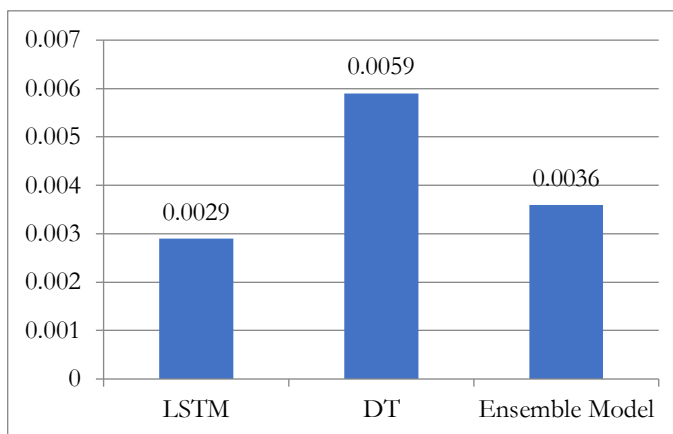


Figure 5. MSE Prediction Results

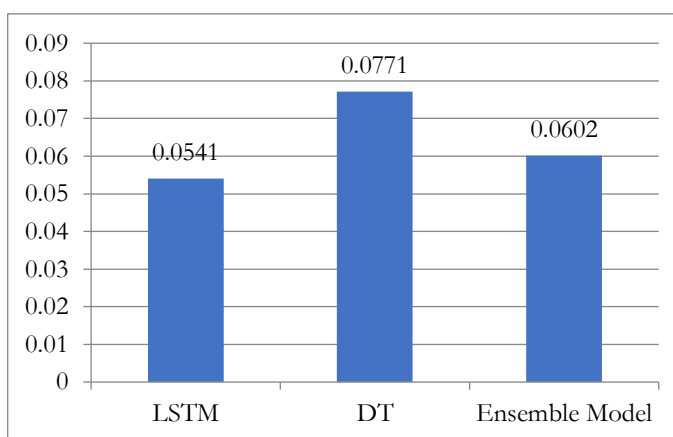


Figure 6. RMSE Prediction Results.

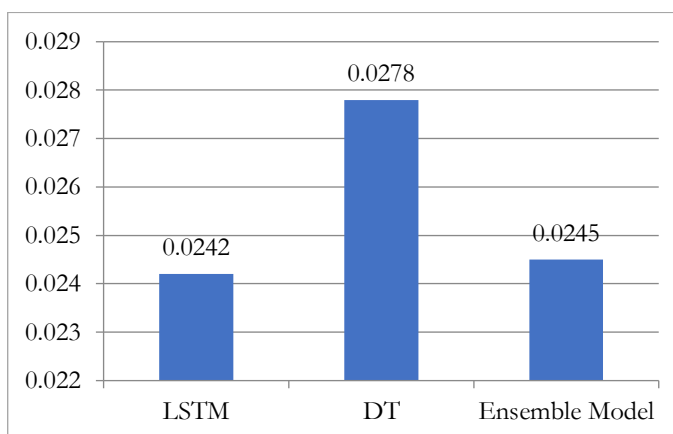


Figure 7 MAE Prediction Results

As shown in Figure 6, the results show that the highest MAE rate was recorded by the DT mode, followed by the ensemble, and lastly, the LSTM model with the lowest MAE rate.

This result indicates that the best-performing model in terms of MAE rate is the LSTM, followed by the Ensemble model and the DT model (just like the previous results on MSE and RMSE). The specific difference between the best performing model (i.e., LSTM) and the others is 0.0003 (for the Ensemble model) and 0.0036 (for the DT model).

Based on the obtained results, the LSTM model outperformed our experiment on both DT and ensemble models. This can be attributed to the LSTM's ability to handle sequential and temporal dependencies, making them excellent in scenarios where such patterns are significant. The poor performance of the DT model, as shown in the results obtained, has significantly affected the ability of the ensemble model to outperform the LSTM. While it is clear that the ensemble benefits from combining models with diverse strengths, when one model (DT) introduces high bias and large errors (in this case, the DT model), it counteracts the strengths of the better-performing model (LSTM). As a result, the ensemble's performance is dragged down, making it less effective than the LSTM alone.

While the predictive performance of DT in terms of MSE, RMSE, and MAE may be lower than the LSTM, it excels in interpretability. The DT creates a rule-based structure where the prediction for our target predictors is traceable back through the decision paths of the tree. This makes it easy to understand why certain predictions are based on specific input features. However, the limitation, as evident from the presented result, is its sacrifice on performance for interpretability, especially when dealing with complex (such as the electronic market dataset used in this study), nonlinear relationships, or high-dimensional data like time series, where the LSTM outperforms in terms of predictive accuracy. While the LSTM produces more accurate predictions by recording lower MSE, RMSE, and MAE than the DT model, it cannot inherently explain predictions, thus requiring explanation techniques such as SHAP. By combining LSTM and SHAP, we can discover several factors, such as feature importance and interaction. This has helped our explanation of the model by identifying the features that are globally important across all predictions and how their level of influence on the CLS predictions.

4.1. Comparison of Actual vs Predicted CLV

The prediction results from each model compared against the actual values is presented in this section. Figure 8 and Table 5 show the graphical comparison of all the predictions from this study against the actual CLV. In order to have clear visibility of the results, this comparison was done on the first 100 evaluation records to avoid making the graph overpopulated with thousands of dense data points, making the graphs poorly visible and inflexible for analytical conclusions.

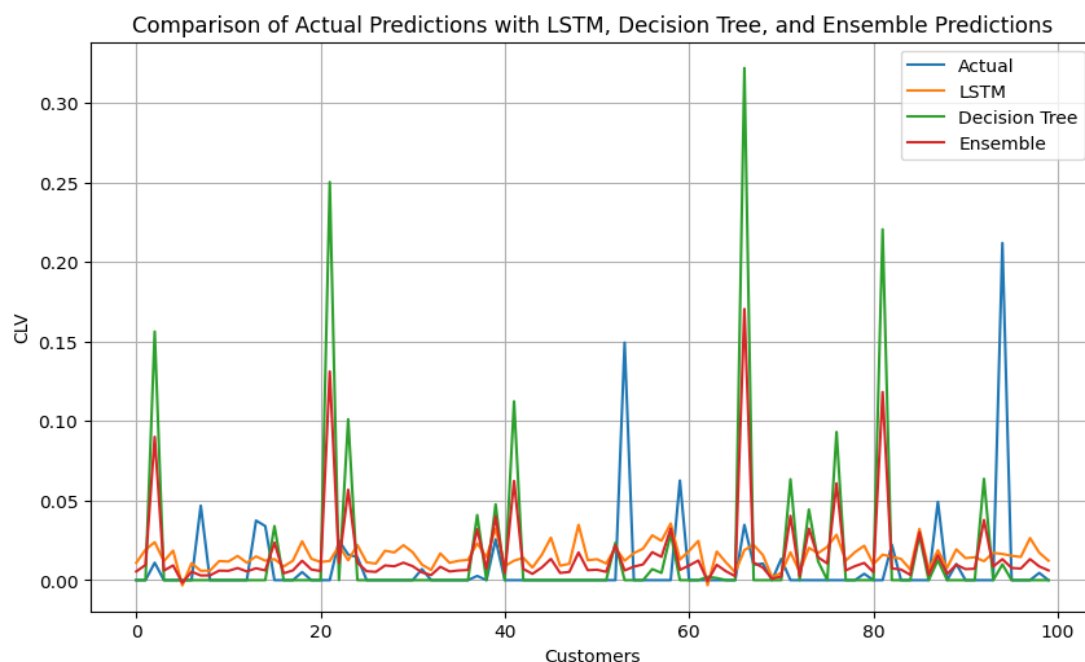


Figure 8. Comparison of actual and all models predicted CLV.

Table 5. Prediction result comparison.

S/No.	Actual	LSTM	DT	Ensemble
0	0.000000	0.007175	0.000000	0.003587
1	0.000000	0.018067	0.000000	0.009034
2	0.010974	0.027040	0.156317	0.091678
3	0.000000	0.010291	0.000000	0.005146
4	0.000000	0.017041	0.000000	0.008521
5	0.000000	-0.000429	0.000000	-0.000215
6	0.000000	0.007297	0.000000	0.003648
7	0.046842	-0.000486	0.000000	-0.000243
8	0.000000	-0.000177	0.000000	-0.000088
9	0.000000	0.007145	0.000000	0.003573
10	0.000000	0.019270	0.000000	0.009635
11	0.000000	0.010429	0.000000	0.005214
12	0.000000	0.004137	0.000000	0.002068
13	0.037473	0.014916	0.000000	0.007458
14	0.033994	0.011565	0.000000	0.005782
15	0.000000	0.012273	0.033994	0.023083
16	0.000000	0.010462	0.000000	0.005231
17	0.000000	0.012514	0.000000	0.006257
18	0.005086	0.025091	0.000000	0.012545
19	0.000000	0.010723	0.000000	0.005362
20	0.000000	0.005683	0.000000	0.002842
21	0.000000	0.010807	0.250268	0.130537
22	0.024893	0.022441	0.000000	0.011221
23	0.016060	0.010388	0.101178	0.055783
24	0.014454	0.023822	0.000000	0.011911
25	0.000000	0.006993	0.000000	0.003496

4.2. Comparison with Previous Study

The results obtained in this study are compared to previous studies, as shown in Table 6, based on MAE, MSE, and RMSE.

Table 6. Results comparison with previous studies.

S/No.	Models	MAE	MSE	RMSE
1	Deep Neural Network (DNN)[6]	0.4800	0.3900	0.6200
2	Decision Tree (DT)[6]	0.5600	0.6900	0.7900
3	Random Forest (RF) [6]	0.5200	0.7500	0.6700
4	Long Short-Term Memory (LSTM) This study	0.0242	0.0029	0.0541
5	Decision Tree (DT) This study	0.0278	0.0059	0.0771
6	Ensemble This study	0.0245	0.0036	0.0602

As shown in Table 6, the results from this study (in particular) show that LSTM outperforms the DNN, DT, and RF models by [6]. Specifically, the DT model from this study, despite being the least performing model, outperformed the DT model from [6] and the RF and DNN models. In this regard, the models from this study have demonstrated remarkable performance in terms of minimal error rates of MAE, MSE, and RMSE.

4.2. Result of Explainability

This section presents the explainability analysis of the CLV prediction using Shapley Additive Explanations (SHAP). Among the three models evaluated (LSTM, DT, and Ensemble), SHAP was applied to the best-performing model, LSTM, to interpret the most influential

features in predicting CLV. Decision Trees offer the advantage of fast training and inference, making them highly interpretable and suitable for immediate decision-making. However, LSTMs provide higher accuracy and better handling of complex, sequential data, albeit at the cost of interpretability. To address this limitation, SHAP was employed to enhance the interpretability of the LSTM model.

Figures 8 and 9 illustrate the contribution of different features to the CLV prediction. Figure 8 (SHAP Summary Plot) visualizes the impact of each feature across all individual predictions, while Figure 9 (SHAP Feature Importance) ranks features based on their average absolute SHAP values. The results indicate that `begin_seconds` is the most influential feature, while `event_type` has the least impact.

The high influence of `begin_seconds` suggests its strong correlation with CLV, as timing plays a critical role in user engagement and purchasing behavior. This aligns with e-commerce and social media insights, where interaction timeframes (e.g., morning sessions) can indicate varying user activity levels, enabling targeted marketing strategies.

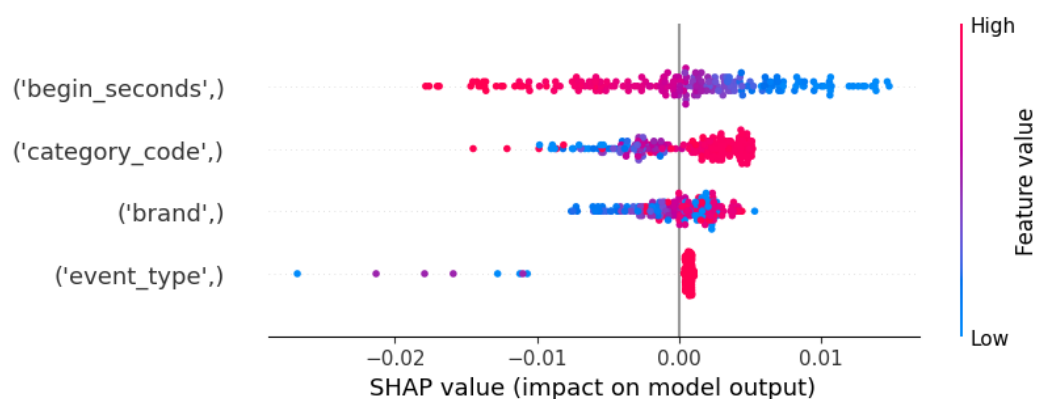


Figure 8. SHAP Summary plot of feature values.

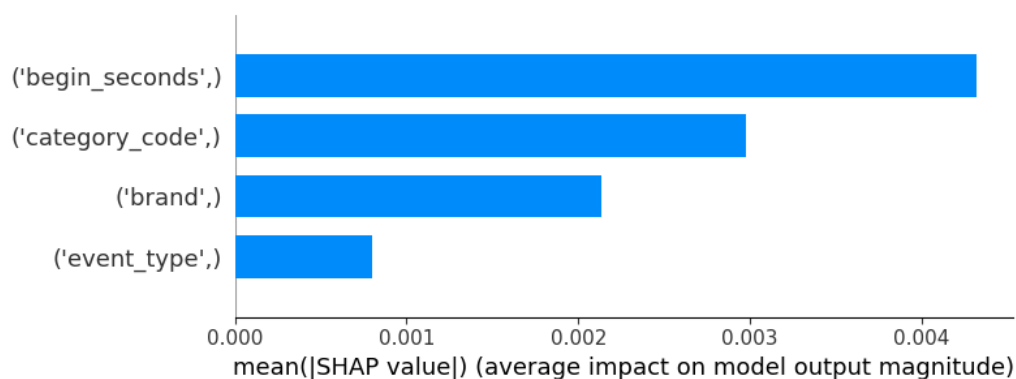


Figure 9 SHAP feature importance

5. Conclusions

In conclusion, this paper has implemented an improved CLV prediction model employing LSTM, DT, and ensemble approaches. It overcomes the limitation of the Deep Neural Network (DNN) implemented by previous studies by leveraging LSTM with optimal parameters for CLV prediction. As reported in the paper, out of all three (i.e., LSTM, DT, and Ensemble) CLV predictors implemented, the LSTM outperformed. This is evident due to the inherent architectural advancement of the LSTM model over the DNN architecture. Furthermore, regarding explainability and interpretation, the `begin_time` (when a user begins browsing/shopping activity on the ecommerce website) ranked the most relevant in CLV prediction. The item category, brand, and event type followed this. In terms of future research directions, possible improvements that can be done are (I) Expanding the interpretability by specifying the exact hours of the day/night that play the most important role in determining the CLV; (II) Expanding the scope of the study to include shopping recommendations based

on the predicted CLV to achieve improved customer experience; (III) Providing interaction and feedback in a deployed environment for interaction with users and improve trust in AI-enabled Information Systems, and (IV) considering other explainability techniques and their practical impacts in e-commerce businesses.

Author Contributions: Conceptualization: S.A., A.I. and O.A.; Methodology: Conceptualization: S.A., A.I. and O.A.; Modelling: S.A. and A.I.; Validation: S.A., A.I. and O.A.; Resources: S.A., A.I. and O.A.; Writing—original draft preparation: S.A., A.I. and O.A.; Writing—review and editing: S.A., A.I. and O.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used in this paper is publicly available on Kaggle.

Acknowledgments: The authors sincerely thank the American University of Nigeria, Yola, for providing a conducive research environment.

Conflicts of Interest: The authors declare no conflict of interest.

References

- [1] R. Dwivedi *et al.*, “Explainable AI (XAI): Core Ideas, Techniques, and Solutions,” *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1–33, Sep. 2023, doi: 10.1145/3561048.
- [2] E. Tjoa and C. Guan, “A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 32, no. 11, pp. 4793–4813, Nov. 2021, doi: 10.1109/TNNLS.2020.3027314.
- [3] F. Sovrano, S. Sapienza, M. Palmirani, and F. Vitali, “Metrics, Explainability and the European AI Act Proposal,” *J.*, vol. 5, no. 1, pp. 126–138, Feb. 2022, doi: 10.3390/j5010010.
- [4] J. Bauer and D. Jannach, “Improved Customer Lifetime Value Prediction With Sequence-To-Sequence Learning and Feature-Based Models,” *ACM Trans. Knowl. Discov. Data*, vol. 15, no. 5, pp. 1–37, Oct. 2021, doi: 10.1145/3441444.
- [5] H. Aeron, A. Kumar, and J. Moorthy, “Data mining framework for customer lifetime value-based segmentation,” *J. Database Mark. Cust. Strateg. Manag.*, vol. 19, no. 1, pp. 17–30, Mar. 2012, doi: 10.1057/dbm.2012.1.
- [6] G. Yilmaz Benk, B. Badur, and S. Mardikyan, “A New 360° Framework to Predict Customer Lifetime Value for Multi-Category E-Commerce Companies Using a Multi-Output Deep Neural Network and Explainable Artificial Intelligence,” *Information*, vol. 13, no. 8, p. 373, Aug. 2022, doi: 10.3390/info13080373.
- [7] L. Ryals and S. Knox, “Measuring and managing customer relationship risk in business markets,” *Ind. Mark. Manag.*, vol. 36, no. 6, pp. 823–833, Aug. 2007, doi: 10.1016/j.indmarman.2006.06.017.
- [8] T. T. Win and K. S. Bo, “Predicting Customer Class using Customer Lifetime Value with Random Forest Algorithm,” in *2020 International Conference on Advanced Information Technologies (ICAIT)*, Nov. 2020, pp. 236–241. doi: 10.1109/ICAIT51105.2020.9261792.
- [9] Y. Sun, H. Liu, and Y. Gao, “Research on customer lifetime value based on machine learning algorithms and customer relationship management analysis model,” *Heliyon*, vol. 9, no. 2, p. e13384, Feb. 2023, doi: 10.1016/j.heliyon.2023.e13384.
- [10] S.-Y. Kim, T.-S. Jung, E.-H. Suh, and H.-S. Hwang, “Customer segmentation and strategy development based on customer lifetime value: A case study,” *Expert Syst. Appl.*, vol. 31, no. 1, pp. 101–107, Jul. 2006, doi: 10.1016/j.eswa.2005.09.004.
- [11] I. Ahmed, G. Jeon, and F. Piccialli, “From Artificial Intelligence to Explainable Artificial Intelligence in Industry 4.0: A Survey on What, How, and Where,” *IEEE Trans. Ind. Informatics*, vol. 18, no. 8, pp. 5031–5042, Aug. 2022, doi: 10.1109/TII.2022.3146552.
- [12] N. Radulovic, “Post-hoc Explainable AI for Black Box Models on Tabular Data,” Institut Polytechnique de Paris, 2023. Accessed: Mar. 17, 2024. [Online]. Available: https://theses.hal.science/tel-04362470/file/107534_RADULOVIC_2023_archivage.pdf
- [13] P. Jiang, H. Suzuki, and T. Obi, “XAI-based cross-ensemble feature ranking methodology for machine learning models,” *Int. J. Inf. Technol.*, vol. 15, no. 4, pp. 1759–1768, Apr. 2023, doi: 10.1007/s41870-023-01270-2.
- [14] K. Bauer, O. Hinz, W. van der Aalst, and C. Weinhardt, “Expl(AI)n It to Me – Explainable AI and Information Systems Research,” *Bus. Inf. Syst. Eng.*, vol. 63, no. 2, pp. 79–82, Apr. 2021, doi: 10.1007/s12599-021-00683-2.
- [15] D. Monner and J. A. Reggia, “A generalized LSTM-like training algorithm for second-order recurrent neural networks,” *Neural Networks*, vol. 25, pp. 70–83, Jan. 2012, doi: 10.1016/j.neunet.2011.07.003.
- [16] O. Günlük, J. Kalagnanam, M. Li, M. Menickelly, and K. Scheinberg, “Optimal decision trees for categorical data via integer programming,” *J. Glob. Optim.*, vol. 81, no. 1, pp. 233–260, Sep. 2021, doi: 10.1007/s10898-021-01009-y.
- [17] H. G. Lee and T. H. Clark, “Market Process Reengineering through Electronic Market Systems: Opportunities and Challenges,” *J. Manag. Inf. Syst.*, vol. 13, no. 3, pp. 113–136, Dec. 1996, doi: 10.1080/07421222.1996.11518136.
- [18] Y. Xing, L. Yu, J. Z. Zhang, and L. J. Zheng, “Uncovering the Dark Side of Artificial Intelligence in Electronic Markets,” *J. Organ. End User Comput.*, vol. 35, no. 1, pp. 1–25, Aug. 2023, doi: 10.4018/JOEUC.327278.
- [19] L. T. Khrais, “Role of Artificial Intelligence in Shaping Consumer Demand in E-Commerce,” *Futur. Internet*, vol. 12, no. 12, p. 226, Dec. 2020, doi: 10.3390/fi12120226.
- [20] S. Shin, J.-U. Won, and M. Kim, “Comparative research on DNN and LSTM algorithms for soot emission prediction under transient conditions in a diesel engine,” *J. Mech. Sci. Technol.*, vol. 37, no. 6, pp. 3141–3150, Jun. 2023, doi: 10.1007/s12206-023-0538-y.

- [21] H.-R. Lou, X. Wang, Y. Gao, and Q. Zeng, "Comparison of ARIMA model, DNN model and LSTM model in predicting disease burden of occupational pneumoconiosis in Tianjin, China," *BMC Public Health*, vol. 22, no. 1, p. 2167, Nov. 2022, doi: 10.1186/s12889-022-14642-3.
- [22] J. Brasse, H. R. Broder, M. Förster, M. Klier, and I. Sigler, "Explainable artificial intelligence in information systems: A review of the status quo and future research directions," *Electron. Mark.*, vol. 33, no. 1, p. 26, Dec. 2023, doi: 10.1007/s12525-023-00644-5.
- [23] B. Lantz, *Machine learning with r: Expert techniques for predictive modeling*. Packt, 2019. [Online]. Available: <https://books.google.com.ng/books?id=IiWewwEACAAJ>
- [24] M. Ranjith Kumar, P. S. J. Srinivasan Anusha, V. Chatiyode, J. Santiago, and D. Chaudhary, "Enhancing Telecommunications Customer Retention: A Deep Learning Approach Using LSTM for Predictive Churn Analysis," in *2024 International Conference on Data Science and Network Security (ICDSNS)*, Jul. 2024, pp. 01–07. doi: 10.1109/ICDSNS62112.2024.10691038.
- [25] N. Sahoo, P. V. Singh, and T. Mukhopadhyay, "A Hidden Markov Model for Collaborative Filtering," *MIS Q.*, vol. 36, no. 4, pp. 1329–1356, 2012, [Online]. Available: <https://dl.acm.org/doi/abs/10.5555/2481674.2481689>
- [26] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017.
- [27] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. LeanPub, 2020. [Online]. Available: <https://christophm.github.io/interpretable-ml-book/>
- [28] D. Gunning, E. Vorm, J. Y. Wang, and M. Turek, "DARPA's explainable AI (XAI) program: A retrospective," *Appl. AI Lett.*, vol. 2, no. 4, Dec. 2021, doi: 10.1002/ail2.61.
- [29] M. Van Lent, W. Fisher, and M. Mancuso, "An Explainable Artificial Intelligence System for Small-unit Tactical Behavior," in *LAAT'04: Proceedings of the 16th conference on Innovative applications of artificial intelligence*, 2004, pp. 900–907. Accessed: Mar. 16, 2024. [Online]. Available: <https://dl.acm.org/doi/10.5555/1597321.1597342>
- [30] A. Barredo Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012.
- [31] S. Maksymiuk, A. Gosiewska, and P. Biecek, "Landscape of R packages for eXplainable Artificial Intelligence," *arXiv*. Cornell University, Sep. 24, 2020. doi: 10.48550/arxiv.2009.13248.
- [32] A. M. Salih *et al.*, "A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME," *Adv. Intell. Syst.*, vol. 7, no. 1, Jan. 2025, doi: 10.1002/aisy.202400304.
- [33] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You?," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 1135–1144. doi: 10.1145/2939672.2939778.
- [34] S. Zhao *et al.*, "perCLTV: A General System for Personalized Customer Lifetime Value Prediction in Online Games," *ACM Trans. Inf. Syst.*, vol. 41, no. 1, pp. 1–29, Jan. 2023, doi: 10.1145/3530012.
- [35] B. P. Chamberlain, A. Cardoso, C. H. B. Liu, R. Pagliari, and M. P. Deisenroth, "Customer Lifetime Value Prediction Using Embeddings," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2017, pp. 1753–1762. doi: 10.1145/3097983.3098123.
- [36] X. Wang, T. Liu, and J. Miao, "A Deep Probabilistic Model for Customer Lifetime Value Prediction," *arXiv*. ARXIV, 2019. [Online]. Available: <https://arxiv.org/pdf/1912.07753>
- [37] H.-H. Chang and T.-S. Yang, "Consumer rights or unethical behaviors: Exploring the impacts of retailer return policies," *J. Retail. Consum. Serv.*, vol. 64, p. 102779, Jan. 2022, doi: 10.1016/j.jretconser.2021.102779.
- [38] A. Madhuri, "Exploring the Role of Personalization in E-commerce: Impacts on Consumer Trust and Purchase Intentions," *Eur. Econ. Lett.*, vol. 14, no. 3, pp. 907–919, 2024, doi: 10.52783/eel.v14i3.1845.
- [39] D. P. Sakas, D. P. Reklitis, N. T. Giannakopoulos, and P. Trivellas, "The influence of websites user engagement on the development of digital competitive advantage and digital brand name in logistics startups," *Eur. Res. Manag. Bus. Econ.*, vol. 29, no. 2, p. 100221, May 2023, doi: 10.1016/j.jiedeen.2023.100221.
- [40] X. Wang, Y. Wang, D. Liu, Y. Wang, and Z. Wang, "Automated recognition of epilepsy from EEG signals using a combining space–time algorithm of CNN-LSTM," *Sci. Rep.*, vol. 13, no. 1, p. 14876, Sep. 2023, doi: 10.1038/s41598-023-41537-z.
- [41] V. Kumar and W. Reinartz, "Creating Enduring Customer Value," *J. Mark.*, vol. 80, no. 6, pp. 36–68, Nov. 2016, doi: 10.1509/jm.15.0414.
- [42] R. E. Ako *et al.*, "Effects of Data Resampling on Predicting Customer Churn via a Comparative Tree-based Random Forest and XGBoost," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 86–101, Jun. 2024, doi: 10.62411/jcta.10562.
- [43] C. E. Durango Vanegas, J. C. Giraldo Mejía, F. A. Vargas Agudelo, and D. E. Soto Duran, "A Representation Based on Essence for the CRISP-DM Methodology," *Comput. y Sist.*, vol. 27, no. 3, Sep. 2023, doi: 10.13053/cys-27-3-3446.
- [44] T. G. Dietterich, "Ensemble methods in machine learning," in *MCS '00: Proceedings of the First International Workshop on Multiple Classifier Systems*, 2000, pp. 1–15. [Online]. Available: <https://dl.acm.org/doi/10.5555/648054.743935>
- [45] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 4, pp. 1–18, 2018, doi: 10.1002/widm.1249.
- [46] D. Opitz and R. Maclin, "Popular Ensemble Methods: An Empirical Study," *J. Artif. Intell. Res.*, vol. 11, pp. 169–198, Aug. 1999, doi: 10.1613/jair.614.
- [47] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning (still) requires rethinking generalization," *Commun. ACM*, vol. 64, no. 3, pp. 107–115, Mar. 2021, doi: 10.1145/3446776.