# Explainable machine learning models applied to predicting customer churn for e-commerce

**Ikhlass Boukrouh, Abdellah Azmani**
Laboratory of Intelligent Automation and BioMedGenomics, Facuty of Sciences and Technologies of Tangier,
Abdelmalek Essaâdi University, Tetouan, Morocco

**Article Info**

***Article history:***

Received Feb 19, 2024
Revised Jun 23, 2024
Accepted Jul 26, 2024

***Keywords:***

Customer churn
E-commerce
Explainable artificial intelligence
Local interpretable model
Machine learning
Shapley additive explanations
Supervised learning

**ABSTRACT**

Precise identification of customer churn is crucial for e-commerce companies due to the high costs associated with acquiring new customers. In this sector, where revenues are affected by customer churn, the challenge is intensified by the diversity of product choices offered on various marketplaces. Customers can easily switch from one platform to another, emphasizing the need for accurate churn classification to anticipate revenue fluctuations in e-commerce. In this context, this study proposes seven machine learning classification models to predict customer churn, including decision tree (DT), random forest (RF), support vector machine (SVM), logistic regression (LR), naïve Bayes (NB), k-nearest neighbors (K-NN), and artificial neural network (ANN). The performances of the models were evaluated using confusion matrix, accuracy, precision, recall, and F1-score. The results indicated that the ANN model achieves the highest accuracy at 92.09%, closely followed by RF at 91.21%. In contrast, the NB model performed the least favorably with an accuracy of 75.04%. Two explainable artificial intelligence (XAI) methods, shapley additive explanations (SHAP) and local interpretable model-agnostic explanations (LIME), were used to explain the models. SHAP provided global explanations for both ANN and RF models through Kernel SHAP and Tree SHAP. LIME, offering local explanations, was applied only to the ANN model which gave better accuracy.



***Corresponding Author:***

Ikhlass Boukrouh
Laboratory of Intelligent Automation and BioMedGenomics, Facuty of Sciences and Technologies of Tangier
Abdelmalek Essaâdi University
Old Airport Road, Km. 10-Tetouan, Morocco
Email: ikhlass.boukrouh@etu.uae.ac.ma

## 1. INTRODUCTION

Electronic commerce is a rapidly growing field in which business transactions are conducted on the internet [1]. This evolution has transformed the way companies operate and interact with customers. In the 21st century, successful companies actively engage with their customers [2], and recognize the important role of customer loyalty in achieving economic success [3], [4]. However, the increasing competition in the online markets presents various challenges for companies, with customer churn emerging as one of the most critical.

Customer churn is when a customer ceases to use a company's product or service [5], posing a significant challenge across industries due to missed profit opportunities [6]. Furthermore, acquiring new customers often results in higher costs for companies in contrast to maintaining existing ones by satisfying their actual needs [7], resulting in companies spending six times more to acquire customers compared to retain them [8], [9]. This underlines the importance of customer loyalty, ensuring businesses a continuous generation of revenue covering costs in the short, medium, and long term [3], [4], [7]. The success of a company relies on the proactive understanding and management of customer relations [10], achieved through the implementation of customer relationship strategies that fit their business [11]. However, businesses, irrespective of their industry, cannot completely avoid the phenomenon of customer churn, which refers to the transition of a customer from one service provider to another [12]. This is where the significance of predicting customer churn becomes apparent, aiming to evaluate the proportion of customers terminating or potentially terminating their use or subscription to a product or service provided by an organization or company [13].

Artificial intelligence (AI) is the development of computer systems that replicate human intelligence, encompassing natural language processing, visual perception, speech recognition, and decision-making [14]. In e-commerce, AI techniques enhance network marketing, electronic payments, and logistics management, while also allowing platforms to understand the factors influencing the purchasing behaviors of current and potential clients, enabling the development of innovative strategies to address the evolving needs and preferences of consumers [15]. Although AI is powerful in solving a variety of tasks, its complexity and lack of transparency can sometimes make it inadequate. AI models are often viewed as 'black boxes', which complicates the understanding of their internal decision-making processes and the foundations of their conclusions. The key question is how we can bridge this gap between AI's capabilities and our understanding of its operational mechanisms. This challenge has led to the emergence of explainable artificial intelligence (XAI), which focuses on interpreting complex models and emphasizing the understanding of AI models and their predictions. XAI aims to provide clarity on the processes behind these predictions, thereby fostering trustworthiness, ensuring causal relationships, enhancing transferability, building confidence, promoting fairness, facilitating accessibility, and encouraging interactivity [16].

## 2. RESEARCH CONTEXT

### 2.1. Problem statement and proposed solution

While e-commerce platforms apply various strategies to retain customers and reduce churn, the lack of transparency in understanding why customers leave remains a critical issue. Traditional black-box machine learning models used for churn prediction offer limited insights into the factors that contribute to churn, which poses a challenge to the development of effective retention strategies. This problem necessitates the application of XAI to enhance the interpretability of churn prediction models. By addressing this problem, e-commerce platforms can proactively identify at-risk customers, understand the reasons behind potential churn, and implement targeted retention efforts, ultimately improving customer satisfaction and revenue stability or increase. This study aims to explore the application of XAI in the context of predicting customer churn in e-commerce, with the goal of bridging the gap between machine learning accuracy and model interpretability. The study employs supervised learning techniques, including decision trees (DT), random forests (RF), naïve Bayes (NB), logistic regression (LR), support vector machines (SVM), k-nearest neighbors (K-NN), and artificial neural networks (ANN) for customer churn prediction. Additionally, it provides explanations for two models with the highest accuracy using shapley additive explanations (SHAP) and local interpretable model-agnostic explanations (LIME).

### 2.2. Study overview

The current study evaluates different machine learning models for predicting customer churn in the e-commerce industry and utilizes XAI to identify the factors contributing to it. It is organized into five main sections with this introduction offering an overview of the research, encompassing e-commerce, customer churn, AI techniques, and XAI, while also addressing the problem statement. The second section examines an extensive review of existing literature on the utilization of AI techniques for addressing customer churn across various domains, including e-commerce, telecommunications, and banking. It also explores existing XAI models in the context of customer churn. The third section explores the research methodology, covering the methodological steps, machine learning algorithms utilized, performance metrics for classification models, and approaches to XAI models. The fourth section presents the study's results and discusses its findings, which include the accuracy of the applied machine learning models and the identification of the most effective model, along with the key factors influencing customer churn. Finally, the fifth section offers a brief summary of the overall study findings, discusses study limitations, and proposes directions for future research.

## 3. RELATED WORKS

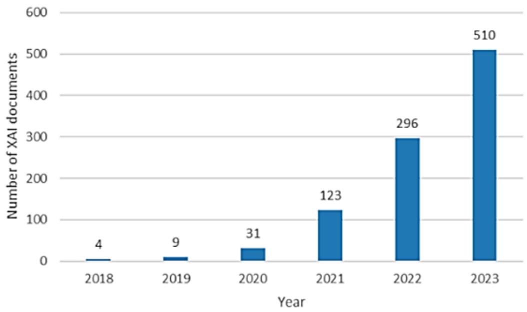
### 3.1. Customer churn classification models

Numerous studies have focused on the classification of customer churn across varied sectors such as e-commerce, telecommunications, banking, and other services and e-services. Baghla and Gupta [5] focuses on predicting customer churn in the e-commerce sector, employing five machine learning techniques, including neural network, SVM, NB, RF, and the deep learning technique Adam. The results reveal that the RF classifier achieves the highest prediction accuracy at 99.35%, surpassing other methods in the study. According to Xiahou and Harada [17], a predictive model for customer churn in business-to-consumer (B2C) e-commerce combines customer segmentation using k-means and prediction with SVM. This study demonstrates a significant improvement in prediction accuracy after customer segmentation, emphasizing the necessity of k-means clustering. The prediction accuracy of the SVM model is found to be superior to LR. Shi *et al.* [18] proposes and assesses a classification model employing machine learning algorithms for predicting customer churn using e-commerce customer data. Notably, the DT algorithm emerges as the most effective model.

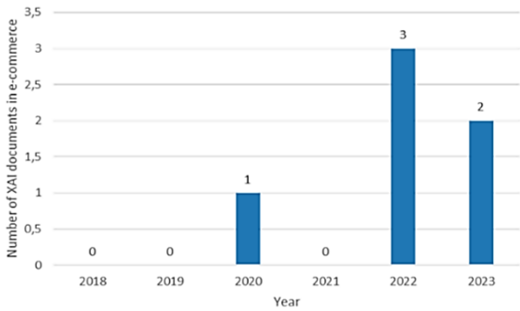
Larasati *et al.* [19] examine customer churn in the Indonesian telecommunications company PT. XYZ, presenting an optimized deep learning algorithm ANN with an accuracy rate of 76.35%. Two influential variables, contract type, and service type are identified. Momin *et al.* [20] evaluates various supervised classification methods to predict customer churn rates using international business machines (IBM) Telco data. The examined models include K-NN, NB, RF, DT, LR, and ANN. The ANN model achieves an accuracy of 82.83% on validation data, surpassing K-NN which achieves 78% accuracy. Panjasuchat and Limpiyakorn [21] focuses on applying the deep Q network (DQN) model in reinforcement learning, comparing it to three supervised classification methods, namely XGBoost, RF, and K-NN. The results highlight the excellence of DQN in terms of accuracy. Kanwal *et al.* [22] compares different classification algorithms, namely gradient boosted tree (GBT), DT, K-NN, and NB. These methods demonstrate significant performance in terms of accuracy, reaching 93%, 90%, 89%, and 89%, respectively. Amin *et al.* [23] proposes an innovative learning approach for predicting customer churn using the NB classifier. This approach integrates a feature weighting method based on a genetic algorithm, part of a broader set of evolutionary algorithms. The evaluation of the effectiveness of this approach is done on public datasets such as BigML Telco, IBM Telco, and Cell2Cell, demonstrating a significant improvement in predictive performance compared to several reference classifiers (Deep-backpropagation (BP)-ANN, convolutional neural network (CNN), neural network, linear regression, NB, XGBoost, K-NN, LogitBoost, SVM, and principal component analysis with linear boosting (PCALB)). With respective accuracies of 0.95, 0.97, and 0.98 on the considered datasets, this adaptive approach represents a notable contribution to predicting customer churn in the telecommunications sector. Najjar *et al.* [24] aims to compare models predicting customer churn for credit cards to anticipate their behavior towards banks, including Bayesian network, C5 tree, chi-squared automatic interaction detector (CHAID) tree, classification and regression (CR) tree, and neural network. The results demonstrate the effectiveness of all models, with the C5 model outperforming the others in performance, influenced by key variables such as the total number of transactions and the total renewable balance on the credit card.

### 3.2. Explainable artificial intelligence models

XAI is a research field focused on interpreting complex models and emphasizing the understanding of AI models and their predictions. Various approaches have been proposed to achieve this objective. The trend towards referencing XAI in the scientific literature is clearly illustrated in Figure 1(a). The first mention of XAI in Scopus, either in titles, abstracts, or keywords, was in 2018 and was associated with four studies [25]-[28]. There was a notable growth in the adoption of XAI, reaching its peak in 2023 with 510 studies, accounting for over half of the total studies in this area. The first research article listed on Scopus that explores the application of XAI in the e-commerce field, illustrated in Figure 1(b), was published in 2020 [29]. This paper demonstrates the integration of XAI into the DeepLimeSeg model to interpret results for customer segmentation.




(a)




(b)

Figure 1. Yearly number of Scopus publications from 2018 to 2023 containing the terms (a) XAI and (b) XAI and e-commerce in titles, abstracts, or keywords. Research conducted on 02-02-2024

Saitoh [30] utilizes XAI with the XGBoost model to identify potential online customers from a group of offline customers. Mandeep *et al.* [31] applies XAI to forecast stock market trends and elucidates the predictions through the utilization of SHAP and LIME. Lee *et al.* [32] applies SHAP to comprehend the factors that affect customer behavior. Xuehan [33] employs SHAP models to comprehend the feature effects on personal credit risk in the e-commerce sector.

## 4. METHODOLOGY

The current study aims to compare seven machine learning models for predicting customer churn in a marketplace. The two best-performing models are further analyzed using XAI techniques to explain the outcomes. To accomplish this, the research methodology is divided into six phases, as illustrated in Figure 2.

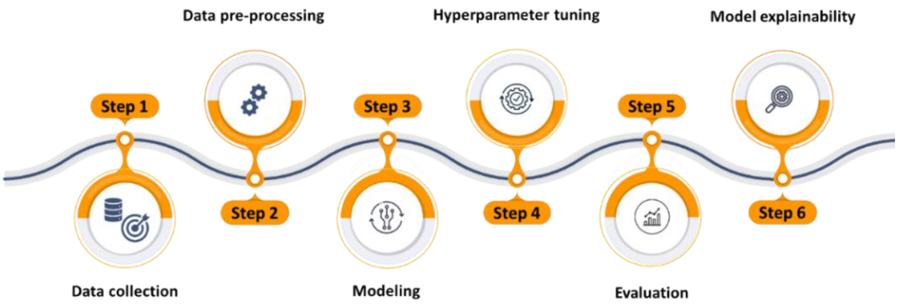



Figure 2. Research methodology

### 4.1. Data collection and pre-processing

The data used for this study were obtained from kaggle [34], consisting of 2,841 customers, each defined by 16 features, including gender, marital status, city tier, tenure, preferred order category, preferred payment mode, preferred login device, coupon usage, order amount hike from last year, order count, days since the last order, hours spent on the app, warehouse-to-home distance, number of registered devices, complaints and satisfaction score. The target variable, 'churn,' indicates whether a customer has canceled their subscription and is coded as either 'yes' or 'no.' Out of the 2,841 instances, 2,362 are labeled 'no' and 479 are labeled 'yes'.

To prepare these data for machine learning prediction, cleaning steps were conducted, addressing missing data, removing duplicates, and converting categorical variables to an appropriate format. Following the cleaning process, the dataset was divided into two subsets: 80% of the data was allocated to the training set, and the remaining 20% was set aside for the test set. The training set, comprising the majority of the data, was utilized to train the machine learning model, while the test set, representing a smaller portion, was reserved for evaluating the model's performance on previously unseen data instances.

### 4.2. Machine learning models

Since the utilized database contains a target variable, 'churn', the selected machine learning models belong to the category of supervised learning. These models DT, RF, NB, LR, SVM, K-NN, and ANN. Each of these models was chosen to evaluate their effectiveness in predicting customer churn.

#### 4.2.1. Decision tree

DT employs an inductive method to learn from known data classes [35]. It takes the form of a tree-like structure where each path from the root to a leaf is defined by a sequence of data separation leading to an outcome. This tree represents a hierarchy of knowledge relationships comprising nodes and connections [36]. It subdivides data into smaller groups through simple decision-making steps, promoting similarity between values within each group. As an algorithm, it is both easy to interpret, integrate into databases, and reliable, making it widely preferred in classification [35].

#### 4.2.2. Random forests

RF adopts a splitting strategy for model construction. It generates various types of DT, with each tree trained by randomly selecting an attribute from the complete set of predictive attributes. It grows to its maximum depth based on a specific subset of features [37]. The RFs performance is influenced by the number of DT it contains, such that an increase in this number is proportional to an improvement in the probability of accuracy [14]. This approach combines the strengths of different trees to enhance model reliability, thereby contributing to improved predictive performance.

#### 4.2.3. Naïve bayes

NB, introduced by Reverend Thomas Bayes in 1,760, relies on two types of probabilities: i) $\mathcal{P}(\mathcal{C}_i)$ the a priori probability of each class $\mathcal{C}_i$ and ii) $\mathcal{P}(\mathcal{A}_j/\mathcal{C}_i)$ the conditional probability of each class label given the value x of the variable [23]. The steps of the NB algorithm in classifying datasets are outlined in the study by Afifah and Much [38]. This method can provide high accuracy and fast performance when analyzing large datasets [39].

#### 4.2.4. Logistic regression

LR serves as an alternative to the least squares method, particularly when dealing with multivariate models that involve a distinction between dependent and independent variables [35]. This method is capable of handling both real and/or categorical inputs. The prediction rule states that if the predicted probability is above 0.5, the output is assigned to class 0; otherwise, it is assigned to class 1 [40].

#### 4.2.5. Support vector machines

SVM, developed by Cortes and Vapnik in 1995, is particularly effective for tasks involving binary classification [35]. The algorithm maps each data point into an n-dimensional space, where n represents the number of features, assigning a value to each feature based on its respective coordinate. Its objective is to define the optimal hyperplane for a perfect separation of classes [41]. New unlabeled samples are then assigned to one of the two categories based on their position relative to the hyperplane.

#### 4.2.6. K-nearest neighbors

K-NN, introduced by Fix and Hodges in 1951, is based on the fundamental principle that the most similar data points belong to the same class [35]. When a new observation is presented, the algorithm assesses its proximity to all other observations in the training dataset. It then identifies the K most similar observations. The most frequent class among these K neighbors is assigned to the new observation, categorizing it into the class most represented among its close neighbors. A notable feature of this method is that it does not require prior knowledge of the data distribution for classification [42]. Additionally, it demonstrates the ability to provide accurate predictions even with large datasets [43].

#### 4.2.7. Artificial neural networks

ANN consists of interconnected layers of neurons, functioning as basic processing units similar to biological neurons [44]. These units integrate weighted inputs, representing influences from synaptic connections and simulating the roles of dendrites and axons. Signal transmission between neurons occurs through activation functions.

### 4.3. Hyperparameters configuration

To select optimal hyperparameters, a grid search approach is employed. This method explores predefined combinations of hyperparameters and evaluates the performance of each combination based on accuracy as the optimization criterion. The combination of hyperparameters that demonstrates the best average results is chosen as the final model for forecasting customer churn.

### 4.4. Performance evaluation metrics

To evaluate the performance of classification models, various metrics are applicable. Details of 16 measures are provided in [45]. This study specifically utilizes five metrics: confusion matrix, accuracy, precision, recall, and F1-score. The confusion matrix involves marking actual values as true or false and predicting outcomes as positive or negative. Accuracy measures the ratio of correct predictions to total predictions. Precision measures the accuracy of positive predictions. Recall or sensitivity is the proportion of correct positive predictions to the total number of actual positives. The F1-score is the harmonic mean of recall and precision.

### 4.5. Explainable artificial intelligence

The objective of interpretable machine learning is to comprehend the prediction-making process of models. It seeks to address questions about the relationships between input and output, as well as identify the most influential features driving predictions [46]. This study utilizes two different XAI techniques: SHAP and LIME. SHAP employs game theory principles to explain the functioning of a machine learning model. Using an additive feature attribution method, SHAP seeks to make the model interpretable [47]. SHAP provides several specialized versions for distinct model types, such as Kernel SHAP, Tree SHAP, Deep SHAP, Gradient SHAP, Linear SHAP, and Partition SHAP. More details about these techniques can be found in the documentation at the SHAP tools. The SHAP documentation is available at: https://shap.readthedocs.io. LIME falls into the category of Post-Hoc and Model-Agnostic methods, offering localized explanations for specific predictions [48].

## 5. RESULTS AND DISCUSSIONS

### 5.1. Classification results

The performance of the seven machine learning models was evaluated using confusion matrix, accuracy, precision, recall, and F1-score. The results are summarized in Table 1 and Figure 3, The DT model achieved an accuracy of 89%. For each class, it showed a precision of 95%, a recall of 93%, and an F1-score of 94% for class 0. Additionally, it demonstrated a precision of 69%, a recall of 75%, and an F1-score of 72% for class 1. The RF model attained a 91% accuracy. It displayed a precision of 92% and a recall of 98%, resulting in an F1-score of 95% for class 0. For class 1, it showcased a precision of 88%, a recall of 58%, and an F1-score of 70%. The NB model achieved a 75% accuracy. It demonstrated a precision of 89% and a recall of 80%, leading to an F1-score of 84% for class 0. In contrast, for class 1, it presented a 36% precision, a 54% recall, and a 43% F1-score. The accuracy of the LR model was 86%, with precision and recall values of 86% and 98%, resulting in an F1-score of 92% for class 0. Moreover, for class 1, the model showcased a precision of 76%, a recall of 26%, and an F1-score of 39%. The SVM model delivered an accuracy of 83%, showcasing precision scores of 83% and 100%, recall rates of 100% and 5%, and F1-scores of 91% and 10% for class 0 and class 1, respectively. The K-NN model achieved an accuracy of 84%, with precision and recall rates of 86% and 96% resulting in an F1 score of 91% for class 0. In the case of class 1, it demonstrated precision, recall, and F1 score values of 71%, 27%, and 38%, respectively. The ANN model achieved a higher accuracy at 92%, with precision, recall, and F1 score metrics of 93%, 98%, and 95% for class 0, and 87%, 65%, and 74% for class 1.

Table 1. The performance of various models: confusion matrix, accuracy, precision, recall and F1-score

| Models | Values of the confusion matrix | | | | Accuracy | Class | Evaluation metrics | | |
|---|---|---|---|---|---|---|---|---|---|
| | TN | FP | FN | TP | | | Precision | Recall | F1-score |
| DT | 436 | 33 | 25 | 75 | 0.8981 | 0 | 0.95 | 0.93 | 0.94 |
| | | | | | | 1 | 0.69 | 0.75 | 0.72 |
| FR | 461 | 8 | 42 | 58 | 0.9121 | 0 | 0.92 | 0.98 | 0.95 |
| | | | | | | 1 | 0.88 | 0.58 | 0.70 |
| NB | 373 | 96 | 46 | 54 | 0.7504 | 0 | 0.89 | 0.80 | 0.84 |
| | | | | | | 1 | 0.36 | 0.54 | 0.43 |
| LR | 461 | 8 | 74 | 26 | 0.8559 | 0 | 0.86 | 0.98 | 0.92 |
| | | | | | | 1 | 0.76 | 0.26 | 0.39 |
| SVM | 469 | 0 | 95 | 5 | 0.8330 | 0 | 0.83 | 1.00 | 0.91 |
| | | | | | | 1 | 1.00 | 0.05 | 0.10 |
| K-NN | 452 | 17 | 73 | 27 | 0.8418 | 0 | 0.86 | 0.96 | 0.91 |
| | | | | | | 1 | 0.61 | 0.27 | 0.38 |
| ANN | 459 | 10 | 35 | 65 | 0.9209 | 0 | 0.93 | 0.98 | 0.95 |
| | | | | | | 1 | 0.87 | 0.65 | 0.74 |

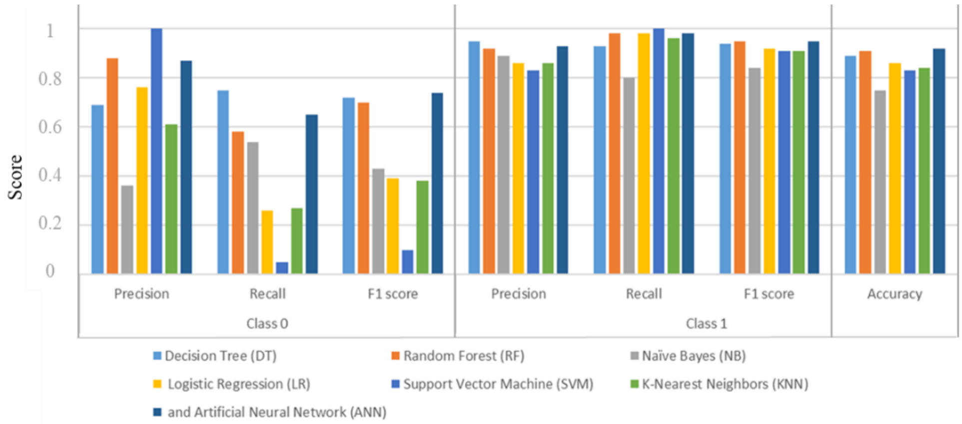


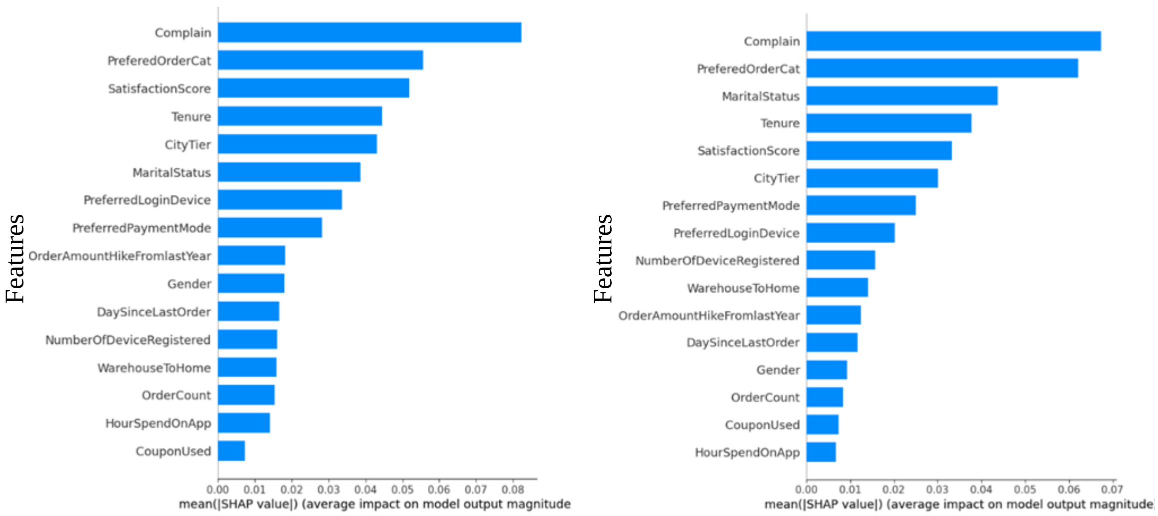

Figure 3. Evaluation model performance

Models DT, LR, K-NN, and SVM demonstrate good performance with accuracies of 89%, 86%, 84%, and 83%, respectively. In contrast, the NB model showed the lowest accuracy, standing at 75%, with precision, recall, and F1 scores of 89%, 80%, and 84% for class 0, and 36%, 54%, and 43% for class 1. To identify non-churn customers, DT achieved the highest precision with a score of 95%, followed closely by ANN with a precision of 93% and RF with 92%. However, the lowest precision was observed with K-NN at 86%. Conversely, to detect churn customers, SVM achieved the highest precision with a score of 100%, followed by RF with a precision of 88% and ANN with 87%. In contrast, the NB model showed the lowest precision at 36%. When identifying non-churn customers, SVM led with the highest recall at 100%, followed by RF, LR, and ANN, each achieving a recall of 98%, while K-NN demonstrated a recall at 96%. Conversely, NB showed the lowest precision at 80%. On the other hand, in the case of identifying churn customers, DT attained the highest precision at 75%, followed by ANN at 65%, RF at 58%, NB at 54%, and K-NN and LR at 27% and 26%, respectively. In contrast, the SVM model showed the lowest precision at 5%. To detect non-churn customers, ANN and RF led with the highest F1-score at 95%, followed by DT at 94%, LR at 92% while K-NN and SVM demonstrated each at 96%. Conversely, NB showed the lowest precision at 84%. On the other hand, when identifying churn customers, ANN attained the highest precision at 74%, followed by DT at 72%, RF at 70%, NB at 43%, and LR and K-NN at 39% and 38%, respectively. In contrast, the SVM model showed the lowest precision at 10%.

### 5.2. Explainable artificial intelligence results

The application of XAI has facilitated the identification of dominant features influencing the prediction of customer churn. In order to gain deeper insights into the reasons behind customer departure, SHAP and LIME methods were applied to two high-performing models, which are the ANN and RF. This strategic decision aims to establish a balance between precision and recall, underscoring the importance of reaching a harmonious state. The SHAP values quantify the disparity between the actual prediction result and the model's average prediction. Kernel SHAP was used for the ANN model, while Tree SHAP was utilized for the RF model.

Figures 4(a) and 4(b) illustrate the global importance for the ANN and RF models, respectively. The x-axis represents the average change in the model output when a feature is excluded. Features are arranged based on the absolute sum value of their effect magnitudes on the model. In both models, 'Complain' emerges as the most influential feature. This suggests that the presence or absence of complaints has a significant impact on the model predictions.

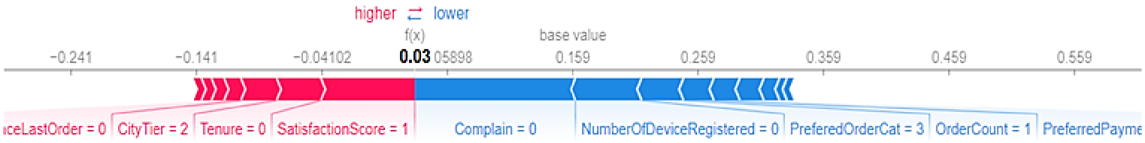


(a) (b)

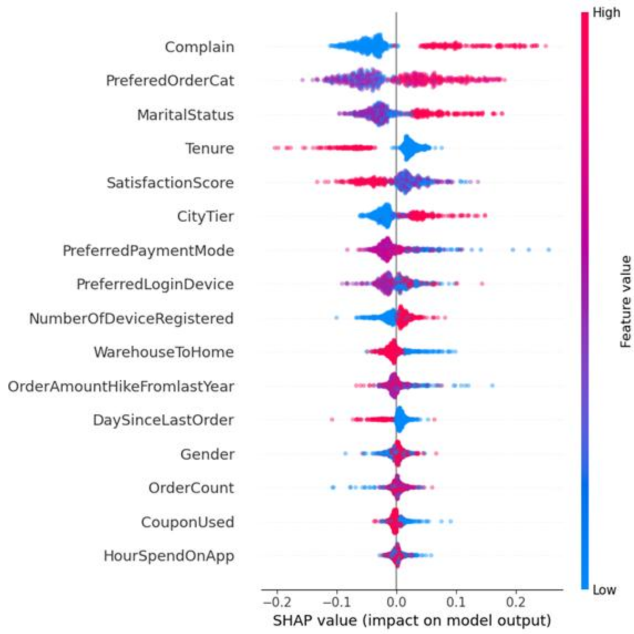
Figure 4. Global feature importance (a) for the ANN model and (b) for the RF model

As shown in Figure 5(a), which illustrates the SHAP values for the ANN model, the placement on the x-axis signifies the influence of that value on the model's output. The factors contributing to an increased likelihood of customer engagement with the company become apparent. Customers with a preferred order category (PreferredOrderCat=3), those who have placed multiple orders (OrderCount=1), and who haven't raised any complaints (Complain=0) show positive influences. Coupled with their preferred payment method, these elements collectively create a significant drive toward engagement, surpassing the negative factors.

Instead of Figure 5(b), which displays the SHAP values of the RF model, a detailed exploration of the contributions of all attributes across both classes is offered. Notably, the 'Complain' feature exhibits both positive and negative SHAP values in both models, indicating a non-linear relationship with the target variable. A positive SHAP value for 'Complain' suggests that, in certain instances, having a complaint is associated with an increase in the predicted outcome, potentially indicating a higher likelihood of the target variable belonging to the 'churn' class.




(a)



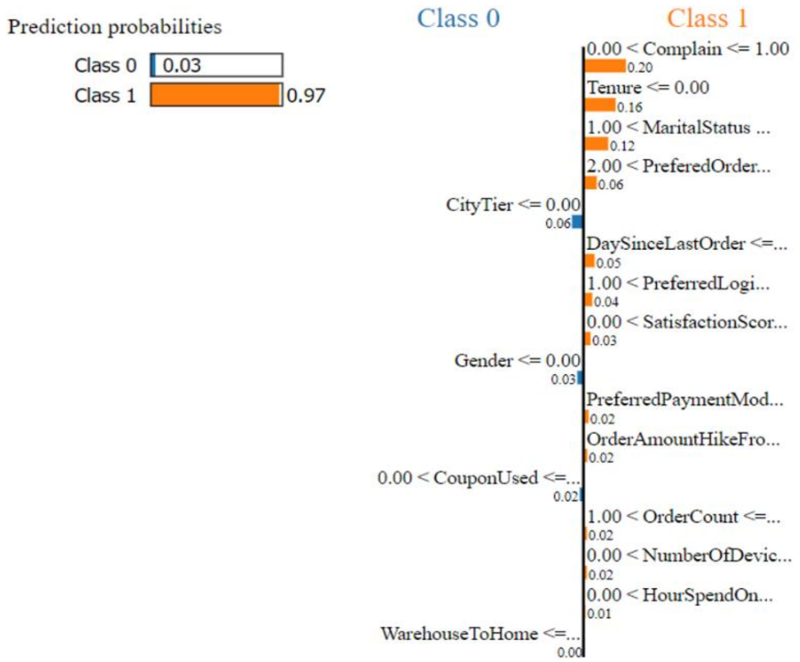

(b)

Figure 5. SHAP values for (a) ANN and (b) RF models

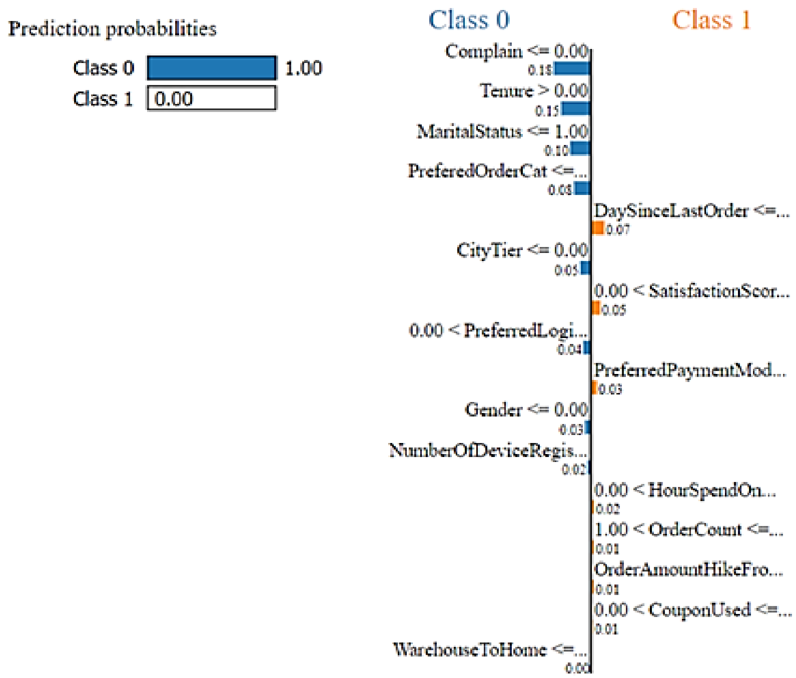
The LIME method was applied to the ANN model due to its superior performance, aiming to identify key features influencing the classification of data points into churn or non-churn categories. Figures 6(a) and 6(b) illustrate the local feature importance generated by LIME for the ANN model, showcasing the contributions for two specific classes: class 1 (churn) and class 0 (non-churn). As illustrated in Figure 6(a), according to LIME's analysis, the features indicating complain, tenure, marital status and preferred order category have a significant positive impact on the classification of this data point as a churn customer. Conversely, features such as city tier, gender, coupon used, and warehouse to home have a negative effect on the classification. The probability values associated with each feature in LIME's interpretation indicate the individual impact that these features have on a specific instance for the model.

A probability value of 0.20 for the 'Complain' feature suggests that, in the context of this specific data point, the presence or absence of a complaint significantly contributes to the model predicting the customer as likely to churn. The higher the probability, the more influential the feature is in driving the prediction toward the churn class. Similarly, other features with associated probabilities provide insights into their respective impacts to understand the decision-making process of the model for classifying instances as churn customers.

Figure 6(b) illustrates how features impact the classification of non-churn customers. The complaint probability stands at 0.20. Regarding the 'tenure' feature, a probability value of 0.15 suggests its substantial contribution to predicting the customer as likely to stay non-churn for this specific data point. Additionally, the probability of 'marital status' is 0.10, and the probability for preferred category is 0.08. Most features show an opposite effect compared to churn customers, contributing significantly to the model's prediction. However, some features have unique contributions that don't follow this pattern.




(a)




(b)

Figure 6. LIME interpretability graph using the ANN model for (a) churn and (b) non-churn prediction

These findings underscore the potential of integrating machine learning and XAI in e-commerce to develop more effective customer retention strategies. By understanding the key factors driving customer churn, e-commerce platforms can design personalized interventions to address specific issues and improve customer satisfaction, thereby reducing churn rates and increasing customer loyalty. This study suggests that proactive customer service, customized marketing, specific engagement strategies, and predictive analytics are crucial for enhancing customer retention in e-commerce. Addressing complaints promptly, utilizing tenure and order preference data for targeted campaigns, developing strategies based on customer segments, and using methods to identify at-risk customers early can significantly boost customer loyalty and grow revenue.

## 6. CONCLUSION

This research provides a comparative analysis of classification machine learning models for predicting customer churn and explores the application of XAI within the e-commerce sector. Starting with an exhaustive review of prior research on customer churn, it underscores the limited exploration of XAI methods for elucidating churn models in e-commerce despite notable progress in addressing this issue. The methodology adopted in this study encompasses five phases: data collection and preparation, model construction, model evaluation, and model explainability. The selected models include DT, RF, LR, SVM, NB, and ANN. The performance of these models was evaluated using five quality indicators: confusion matrix, accuracy, precision, recall, and F1-score. Hyperparameter configuration for each model involves a grid search method aimed at maximizing accuracy. Results indicate that the ANN model achieves the highest accuracy at 92.09%, closely followed by RF at 91.21%. Conversely, the NB model performed the least favorably with an accuracy of 75%. To enhance model interpretability, two explainability techniques were applied: A global SHAP model and a local LIME model. SHAP was implemented on both ANN and RF, utilizing Kernel SHAP for ANN and Tree SHAP for RF. Findings reveal that, in both models, the 'complaint' feature is the most influential in predictions. The local LIME model, applied exclusively to the ANN model, uncovered that 'complaint,' 'tenure,' 'marital status,' and 'preferred order category' positively impact churn classification, while 'city tier,' 'gender,' 'coupon used,' and 'warehouse to home' negatively impact churn classification. For the non-churn class, most features exhibit an opposite effect compared to churn customers. This study explored a comprehensive set of machine learning models and XAI techniques with a dataset from Kaggle. However, further and in-depth studies are needed to confirm these findings across different datasets and e-commerce platforms, especially regarding the generalizability of the identified key features influencing churn. Future studies may explore the integration of other XAI methods with different machine learning models and datasets to validate and expand on these findings, with feasible ways of producing even more refined and actionable insights for e-commerce platforms.

## ACKNOWLEDGEMENTS

The support for this research is provided by the Ministry of Higher Education, Scientific Research, and Innovation, as well as the digital development agency (DDA) and the national center for scientific and technical research (CNRST) of Morocco, under the Smart DLSP Project - AL KHAWARIZMI AI-PROGRAM.

## REFERENCES

[1] A. Setiawan, A. N. Muna, E. R. Arumi, and P. Sukmasetya, "The growth electronic commerce technology and user interface in Indonesia," *Test Engineering and Management*, vol. 83, 2020.

[2] V. Kumar and D. Shah, "Building and sustaining profitable customer loyalty for the 21st century," *Journal of Retailing*, vol. 80, no. 4, pp. 317–329, Jan. 2004, doi: 10.1016/j.jretai.2004.10.007.

[3] M. Bhargava, S. Singh, J. Sharma, and D. F. Vinod, "Telecom customer churn prediction," *Proceedings of International Conference on Wireless Communication*, 2022, pp. 325–333, doi: 10.1007/978-981-16-6601-8_30.

[4] H. Jain, A. Khunteta, and S. Srivastava, "Telecom churn prediction and used techniques, datasets and performance measures: a review," *Telecommunication Systems*, vol. 76, no. 4, pp. 613–630, Apr. 2021, doi: 10.1007/s11235-020-00727-0.

[5] S. Baghla and G. Gupta, "Performance evaluation of various classification techniques for customer churn prediction in e-commerce," *Microprocessors and Microsystems*, vol. 94, Oct. 2022, doi: 10.1016/j.micpro.2022.104680.

[6] M. Pondel *et al.*, "Deep learning for customer churn prediction in e-commerce decision support," *Business Information Systems*, pp. 3–12, Jul. 2021, doi: 10.52825/bis.v1i.42.

[7] A. Amin, F. Al-Obeidat, B. Shah, A. Adnan, J. Loo, and S. Anwar, "Customer churn prediction in telecommunication industry using data certainty," *Journal of Business Research*, vol. 94, pp. 290–301, Jan. 2019, doi: 10.1016/j.jbusres.2018.03.003.

[8] A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," *Journal of Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0191-6.

[9] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector," *IEEE Access*, vol. 7, pp. 60134–60149, 2019, doi: 10.1109/ACCESS.2019.2914999.

[10] A. De Caigny, K. Coussement, K. W. De Bock, and S. Lessmann, "Incorporating textual information in customer churn prediction models based on a convolutional neural network," *International Journal of Forecasting*, vol. 36, no. 4, pp. 1563–1578, Oct. 2020,

doi: 10.1016/j.ijforecast.2019.03.029.
[11] K. Coussement, D. F. Benoit, and D. V. D. Poel, "Preventing customers from running away! Exploring generalized additive models for customer churn prediction," *The Sustainable Global Marketplace*, 2015, pp. 238–238, doi: 10.1007/978-3-319-10873-5_134.
[12] S. Kumar and M. Kumar, "Predicting customer churn using artificial neural network," *Engineering Applications of Neural Networks (EANN 2019)*, 2019, pp. 299–306, doi: 10.1007/978-3-030-20257-6_25.
[13] A. M. Oluwatoyin, S. Misra, J. Wejin, A. Gautam, R. K. Behera, and R. Ahuja, "Customer churn prediction in banking industry using power Bi," *Proceedings of Third International Conference on Computing, Communications, and Cyber-Security*, 2023, pp. 767–774, doi: 10.1007/978-981-19-1142-2_60.
[14] H. Pallathadka, E. H. Ramirez-Asis, T. P. Loli-Poma, K. Kaliyaperumal, R. J. M. Ventayen, and M. Naved, "Applications of artificial intelligence in business management, e-commerce and finance," *Materials Today: Proceedings*, vol. 80, pp. 2610–2613, 2023, doi: 10.1016/j.matpr.2021.06.419.
[15] L. T. Khrais, "Role of artificial intelligence in shaping consumer demand in e-commerce," *Future Internet*, vol. 12, no. 12, Dec. 2020, doi: 10.3390/fi12120226.
[16] A. Chaddad, J. Peng, J. Xu, and A. Bouridane, "Survey of explainable AI techniques in healthcare," *Sensors*, vol. 23, no. 2, Jan. 2023, doi: 10.3390/s23020634.
[17] X. Xiahou and Y. Harada, "B2C e-commerce customer churn prediction based on k-means and SVM," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 17, no. 2, pp. 458–475, Apr. 2022, doi: 10.3390/jtaer17020024.
[18] A. Shi, C. Y. Lim, and S. L. Ang, "Customer churn analysis for live stream e-commerce platforms by using decision tree method," *International Conference on Advanced Communication and Intelligent Systems.* Cham: Springer Nature Switzerland, 2023, pp. 168–178, doi: 10.1007/978-3-031-45124-9_13.
[19] A. Larasati, D. Ramadhanti, Y. W. Chen, and A. Muid, "Optimizing deep learning ANN model to predict customer churn," in *2021 7th International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*, Oct. 2021, pp. 1–5, doi: 10.1109/ICEEIE52663.2021.9616714.
[20] S. Momin, T. Bohra, and P. Raut, "Prediction of customer churn using machine learning," *EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing: BDCC 2018.* Springer International Publishing, 2020, pp. 203–212, doi: 10.1007/978-3-030-19562-5_20.
[21] M. Panjasuchat and Y. Limpiyakorn, "Applying reinforcement learning for customer churn prediction," *Journal of Physics: Conference Series*, vol. 1619, no. 1, Aug. 2020, doi: 10.1088/1742-6596/1619/1/012016.
[22] S. Kanwal *et al.*, "An attribute weight estimation using particle swarm optimization and machine learning approaches for customer churn prediction," in *2021 International Conference on Innovative Computing (ICIC)*, Nov. 2021, pp. 1–6, doi: 10.1109/ICIC53490.2021.9693040.
[23] A. Amin, A. Adnan, and S. Anwar, "An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and naïve Bayes," *Applied Soft Computing*, vol. 137, Apr. 2023, doi: 10.1016/j.asoc.2023.110103.
[24] D. A.-Najjar, N. Al-Rousan, and H. AL-Najjar, "Machine learning to develop credit card customer churn prediction," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 17, no. 4, pp. 1529–1542, Nov. 2022, doi: 10.3390/jtaer17040077.
[25] T. Kistan, A. Gardi, and R. Sabatini, "Machine learning and cognitive ergonomics in air traffic management: recent developments and considerations for certification," *Aerospace*, vol. 5, no. 4, Oct. 2018, doi: 10.3390/aerospace5040103.
[26] H. Hagras, "Toward human-understandable, explainable AI," *Computer*, vol. 51, no. 9, pp. 28–36, Sep. 2018, doi: 10.1109/MC.2018.3620965.
[27] A. Adadi and M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018, doi: 10.1109/ACCESS.2018.2870052.
[28] R. Sheh and I. Monteath, "Defining explainable AI for requirements analysis," *KI - Künstliche Intelligenz*, vol. 32, no. 4, pp. 261–266, Nov. 2018, doi: 10.1007/s13218-018-0559-3.
[29] F. M. Talaat, A. Aljadani, B. Alharthi, M. A. Farsi, M. Badawy, and M. Elhosseini, "A mathematical model for customer segmentation leveraging deep learning, explainable AI, and RFM analysis in targeted marketing," *Mathematics*, vol. 11, no. 18, Sep. 2023, doi: 10.3390/math11183930.
[30] F. Saitoh, "A lead scoring system and it's interpretation of online purchasing customers by XAI," *IEEJ Transactions on Electronics, Information and Systems*, vol. 143, no. 12, pp. 1203–1210, Dec. 2023, doi: 10.1541/ieejeiss.143.1203.
[31] Mandeep, A. Agarwal, A. Bhatia, A. Malhi, P. Kaler, and H. S. Pannu, "Machine learning based explainable financial forecasting," in *2022 4th International Conference on Computer Communication and the Internet (ICCCI)*, Jul. 2022, pp. 34–38, doi: 10.1109/ICCCI55554.2022.9850272.
[32] J. Lee, I. T. Jeong, D. H. Lim, K.-Y. Kwahk, and H. Ahn, "Understanding customer purchasing behavior in e-commerce using explainable artificial intelligence techniques," *Proceedings of the Korean Society of Computer Information Conference*, 2021, pp. 387-390.
[33] P. Xuehan, "Research on personal credit risk assessment based on combination weight and shap interpretable machine learning," *Academic Journal of Computing & Information Science*, vol. 5, no. 10, 2022, doi: 10.25236/AJCIS.2022.051009.
[34] A. Verma, "Ecommerce customer churn analysis and prediction," *Kaggle*, 2020. Accessed: Sep. 23, 2022. [Online]. Available: https://www.kaggle.com/datasets/ankitverma2010/ecommerce-customer-churn-analysis-and-prediction
[35] O. Celik and U. O. Osmanoglu, "Comparing to techniques used in customer churn analysis," *Journal of Multidisciplinary Developments*, vol. 4, no. 1, pp. 30–38, 2019.
[36] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 1, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.
[37] I. Ullah, H. Hussain, I. Ali, and A. Liaquat, "Churn prediction in banking system using k-means, LOF, and CBLOF," in *2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, Jul. 2019, pp. 1–6, doi: 10.1109/ICECCE47252.2019.8940667.
[38] S. R. Afifah and M. A. Much, "Improved accuracy of naive bayes classifier for determination of customer churn uses SMOTE and genetic algorithms," *Journal of Soft Computing Exploration*, vol. 1, no. 1, Sep. 2020, doi: 10.52465/joscex.v1i1.5.
[39] Y. Yulianti and A. Saifudin, "Sequential feature selection in customer churn prediction based on naive Bayes," *IOP Conference Series: Materials Science and Engineering*, vol. 879, no. 1, Jul. 2020, doi: 10.1088/1757-899X/879/1/012090.
[40] H. Jain, A. Khunteta, and S. Srivastava, "Churn prediction in telecommunication using logistic regression and logit boost," *Procedia Computer Science*, vol. 167, pp. 101–112, 2020, doi: 10.1016/j.procs.2020.03.187.
[41] S. R. Labhsetwar, "Predictive analysis of customer churn in telecom industry using supervised learning," *ICTACT Journal on Soft Computing*, vol. 10, no. 2, pp. 2054–2060, Jan. 2020, doi: 10.21917/ijsc.2020.0291.
[42] N. N. A. Sjarif, M. R. M. Yusof, D. H. Ten Wong, S. Ya'akob, R. Ibrahim, and M. Z. Osman, "A customer churn prediction using

pearson correlation function and K nearest neighbor algorithm for telecommunication industry," *International Journal of Advances in Soft Computing and its Applications*, vol. 11, no. 2, pp. 46–59, 2019.
[43] M. A. Imron and B. Prasetyo, "Improving algorithm accuracy k-nearest neighbor using z-score normalization and particle swarm optimization to predict customer churn," *Journal of Soft Computing Exploration*, vol. 1, no. 1, Sep. 2020, doi: 10.52465/joscex.v1i1.7.
[44] S. Zheng, L. Qian, P. Li, C. He, X. Qin, and X. Li, "An introductory review of spiking neural network and artificial neural network: from biological intelligence to artificial intelligence," pp. 12–145, 2022, doi: 10.5121/csit.2022.121010.
[45] Ž. Đ. Vujovic, "Classification model evaluation metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, 2021, doi: 10.14569/IJACSA.2021.0120670.
[46] Z. Li, "Extracting spatial effects from machine learning model using local interpretation method: an example of SHAP and XGBoost," *Computers, Environment and Urban Systems*, vol. 96, Sep. 2022, doi: 10.1016/j.compenvurbsys.2022.101845.
[47] Scott M. Lundberg and S. -I. Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
[48] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should i trust you?' explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144, doi: 10.18653/v1/n16-3020.

## BIOGRAPHIES OF AUTHORS

**Ikhlass Boukrouh** is a statistical engineer who graduated from the Faculty of Sciences and Technologies of Tangier, Abdelmalek Essaâdi University in 2021. With a strong interest in using data to find useful answers for business problems, she is currently pursuing a Ph.D. in artificial intelligence and data science, applied to the e-commerce sector. Her research interests include machine learning, predictive analytics, customer behavior modeling, and risk assessment. She can be contacted at email: ikhlass.boukrouh@etu.uae.ac.ma.

**Abdellah Azmani** obtained his doctorate in industrial computing, dynamic systems modeling and artificial intelligence at the University of Sciences and Technologies of Lille in 1991. He worked as a professor at the Ecole Centrale de Lille, France and at the Institute of Computer and Industrial Engineering from Lens, France. He is a professor in the computer science department of the Faculty of Sciences and Technologies of Tangier, Morocco. He is a founding member and director of the Intelligent Automation and BioMedGenomics Laboratory. He has directed numerous doctoral theses and scientific research projects. Its work focuses particularly on the development of generic and disruptive concepts promoting the integration of artificial intelligence and digital transformation, through the creation of decision-making IT solutions for public and private organizations. He can be contacted at: a.azmani@uae.ac.ma.