A I S ARTIFICIAL INTELLIGENCE STUDIES

Customer Churn Prediction: Performance and XAI (SHAP-LIME) Analysis

Gamze Dursun*[a], Nursal Arıcı [b]

ABSTRACT

With the advancement of digital transformation, the increasing volume of data and intensifying competition have made customer churn a significant risk in the telecommunications sector. Therefore, data-driven approaches for predicting customer churn have become increasingly important. However, evaluating these approaches solely on model performance is insufficient; it is also essential that the models' decision-making processes be understandable. This study aims to evaluate churn prediction models in terms of both performance and interpretability. Within this scope, a dataset obtained from Kaggle was analyzed for missing values and outliers, and data transformations were applied, followed by the SMOTE method to address class imbalance. Models were developed using Logistic Regression, Random Forests, Decision Trees, Naive Bayes, and K-Nearest Neighbors algorithms, and their performance was evaluated using performance metrics. To understand the model's behavior and the effects of variables, the SHAP method was applied at the global level, while SHAP and LIME were used at the local level to explain individual predictions; the results from SHAP and LIME were then compared. Therefore, explainable artificial intelligence methods can enhance the transparency and reliability of model outputs, thereby making significant contributions to decision support processes.

[a,*] Gazi University,
Institute of Informatics,
Management Information Systems
06560 - Ankara, Türkiye
ORCID: 0009-0006-9020-3558

[b] Gazi University,
Faculty of Applied Sciences,
Management Information Systems
06560 - Ankara, Türkiye
ORCID: 0000-0002-4505-1341

* Corresponding author.
e-mail: arslnngamze00@gmail.com

Keywords: explainable artificial intelligence (XAI), SHAP, LIME Machine learning, customer churn analysis

Anahtar Kelimeler: Açıklanabilir yapay zekâ (XAI), SHAP, LIME, makine öğrenmesi, müşteri kaybı (churn) analizi
Submitted: 18.05.2026
Revised: 29.06.2026
Accepted: 30.06.2026

doi: 10.30855/ais.2026.09.01.01

Müşteri Kaybı Tahmini: Performans ve XAI (SHAP-LIME) Analizi

ÖZ

Dijital dönüşümle birlikte artan veri hacmi ve rekabet, telekomünikasyon sektöründe müşteri kaybını önemli bir risk haline getirmiştir. Bu sebeple, müşteri kaybını önceden tahmin etmeye yönelik veri odaklı yaklaşımlar önem kazanmaktadır. Ancak, bu yaklaşımların yalnızca model performansı açısından değerlendirilmesi yeterli görülmemekte, modellerin nasıl karar verdiğinin de anlaşılabilir olması önem kazanmaktadır. Bu çalışma, müşteri kaybı tahmin modellerini performans ve açıklanabilirlik açısından değerlendirmeyi amaçlamaktadır. Kaggle'dan alınan veri seti üzerinde eksik veri analizi, aykırı değer tespiti, veri dönüşümü gerçekleştirilmiş ve SMOTE yöntemi uygulanmıştır. Lojistik Regresyon, Rastgele Orman, Karar Ağaçları, Naive Bayes ve K-En Yakın Komşuluk algoritmaları kullanılarak modeller oluşturulmuş ve performansları metrikleri ile değerlendirilmiştir. Modelin davranışının ve değişkenlerin etkilerinin anlaşılabilmesi amacıyla SHAP yöntemi küresel düzeyde uygulanmış, bireysel tahminlerin açıklanması için ise SHAP ve LIME yöntemleri yerel düzeyde kullanılmış ve SHAP ve LIME sonuçları karşılaştırılmıştır. Elde edilen bulgular, açıklanabilir yapay zekâ yöntemlerinin model çıktılarının şeffaflığını ve güvenilirliğini artırarak karar destek süreçlerine önemli katkı sağlayabileceğini sunmaktadır.

To cite this article: G. Dursun and N. Arıcı, "Customer Churn Prediction: Performance and XAI (SHAP-LIME) Analysis," *Artificial Intelligence Studies,* vol.9, no.1, pp. 01-13, 2026. doi:10.30855/ais.2026.09.01.01

1. Introduction

The telecommunications sector is considered a fundamental pillar of digitalization and plays a crucial role in enabling communication among individuals, businesses, and societies. With advances in technology and rising customer expectations in the telecommunications sector, intense competition among operators has driven greater customer mobility [1]. This mobility is referred to as customer churn. Companies aim to minimize the costs associated with customer churn and therefore seek to retain their customers. Identifying the factors that lead customers to leave the service and detecting customers with a high likelihood of churn in advance are of great importance. There is a large amount of data and numerous variables related to customer behavior in the telecommunications sector. Analyses conducted using machine learning methods on these data can help address the churn problem. However, the large volume and imbalance of the data affect the outcomes of these analyses. The application of machine learning methods highlights the importance of the data preprocessing stage. It is not only important for the results to be accurate, but also for them to be interpretable and applicable to business decision-making processes.

Customer churn is the percentage of customers a company loses over a specific period. In other words, it represents the rate at which customers leave the company. Identifying the underlying reasons for customer departure enables intervention before customers are lost. Therefore, churn prediction has a positive impact on business operations [2].

This study, conducted in Python using relevant libraries, involves obtaining a telecommunications dataset, organizing it, performing data preprocessing, applying machine learning classification models, and using model outputs for decision support via explainable artificial intelligence methods. Within this scope, the study aims to present the impact of explainable AI methods on model interpretation at both global and local levels.

2. Literature Review

In the study by Kök, customer churn prediction in the telecommunications sector was addressed, and an interpretable decision framework was provided using explainable methods. The IBM Telco customer dataset obtained from Kaggle was used. The dataset consists of 7,043 customer records and 20 variables. The dependent variable represents the customer's churn status, while the independent variables include demographic, service usage, contract, and financial information such as "Gender," "SeniorCitizen," "InternetService," "Contract," "PaymentMethod," and "TotalCharges." K-Nearest Neighbors, Decision Tree, Random Forest, Support Vector Machine, and Naive Bayes algorithms were applied to this dataset. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics, and the Random Forest model achieved the best performance. To interpret the models, explainable artificial intelligence techniques, including LIME, SHAP, counterfactual explanations, and ELI5, were utilized. These techniques provided both global and local interpretations of customer churn predictions. As a result, the study demonstrated that the use of explainable AI techniques in churn prediction yields reliable, interpretable outcomes [3].

The study conducted by Boukrouh and Azmani is a research article. In their work, machine learning models were applied for customer churn prediction in the e-commerce sector, and their interpretability was examined. The study used an e-commerce customer churn dataset obtained from the Kaggle platform. It consists of 2,841 observations and 16 independent variables. The dependent variable represents customer subscription cancellation, while the independent variables include customer-related information and behavioral attributes such as "Gender," "Tenure," "Order count," "Complaints," and "Satisfaction score." Decision Tree, Random Forest, Support Vector Machines, Logistic Regression, Naive Bayes, K-Nearest Neighbors, and Artificial Neural Networks algorithms were applied to the dataset. After evaluating using performance metrics, the Random Forest model achieved the highest accuracy. To enhance model interpretability, explainable AI methods such as SHAP and LIME were employed. The study showed that explainable artificial intelligence methods significantly improve the interpretability of customer churn prediction models in the e-commerce sector [4].

In the thesis study by Arslanbay, the interpretability of machine learning models for customer churn prediction was investigated. A dataset obtained from Kaggle, consisting of 10,000 observations and 13 independent variables, was used. The dependent variable represents whether the customer leaves the

bank, while the independent variables include demographic, financial, and behavioral attributes. Logistic Regression, Decision Tree, Random Forest, and XGBoost classification algorithms were applied to predict customer churn, with the aim of enhancing model interpretability. The models were evaluated using accuracy, recall, specificity, and ROC–AUC metrics. The highest performance was achieved by the XGBoost model, which was further explained using explainable AI techniques. SHAP and LIME methods were employed to interpret the models at both global and local levels. Based on the findings, it was concluded that using high-performing models alongside explainability methods yields more reliable customer churn predictions [5].

3. Theoretical And Conceptual Foundations

Machine learning is a subfield of artificial intelligence that enables computers to learn from data [6]. Extracting meaningful patterns from data can make predictions or generate algorithms that classify new information.

A machine learning process involves not only model development but also several preparatory steps such as data collection, cleaning, and transformation [7]. After data collection, the presence of variables with different scales, as well as missing or outlier values, necessitates the data preprocessing stage. Once the data is cleaned and transformed, it is analyzed using machine learning algorithms. By also employing explainable artificial intelligence methods, model outputs are interpretable, and the model's decision-making process becomes more tangible. The key factors underlying the model's results are identified using both local and global explainable AI techniques, providing decision-makers with meaningful insights.

3.1. Data Preprocessing

One of the key steps in successfully applying machine learning algorithms is data preparation [1]. Datasets obtained from real-world environments often contain noisy or dirty data. In particular, datasets in the telecommunications sector may include various issues such as missing values, non-numeric values, and outliers [8]. These issues make it difficult for the model to function correctly and may lead to misleading outputs. Therefore, the data to be used must undergo preprocessing before applying a machine learning algorithm [8]. The data preprocessing stage includes several steps, such as handling missing values, converting categorical variables into numerical form, removing duplicate or irrelevant data, balancing the dataset, and scaling features. In addition, normalization and standardization techniques can be applied to address differences in feature scales.

3.2. Classification Algorithms

Machine learning is a subfield of artificial intelligence that enables computers to learn from data [6]. Techniques in machine learning continue to evolve, advancing scientific progress. Some of these techniques include supervised, unsupervised, reinforcement, and semi-supervised learning [9]. In supervised learning, the labels in the training data are known, and the objective of the model is to learn to predict these labels [6].

This study focuses on supervised learning methods and examines Naive Bayes (NB), Decision Trees, Random Forests, Logistic Regression, and K-Nearest Neighbors (KNN).
Logistic Regression is a supervised machine learning method used when the dependent variable has two possible outcomes [10]. Due to its ease of implementation, it is one of the most widely used classification algorithms (Raschka, 2015). Logistic Regression is derived by taking the logarithm of the ratio $P_i/(1-P_i)$. Here, $P_i$represents the probability that a customer will churn or not churn. The value of $P_i$always lies between 0 and 1.

Random Forest is a machine learning algorithm used for classification. It is a flexible algorithm that can perform both regression and classification tasks. It works by constructing multiple decision trees to find the best possible solution [11]. The output of the Random Forest algorithm is obtained by averaging the predictions of individual trees or by taking the majority vote [12]. From a data science perspective, the strong performance of the Random Forest model stems from the fact that multiple independent models operate as an ensemble. The low correlation among these models is the key to its success [12].

Decision Trees, similar to the Random Forest algorithm, are supervised learning algorithms that can perform both regression and classification tasks [13]. They resemble a tree structure consisting of roots, nodes, branches, and leaves. A decision tree is constructed from nodes representing decision points. The structures connecting the nodes are called branches. Each internal node splits into two or more branches [14]. The node at the top of the tree is called the root node. A decision tree starts from the root and moves downward. The terminal nodes are called leaf nodes. In a decision tree, the evaluation process begins at the root node and proceeds downward through the nodes corresponding to the conditions until a leaf node containing the prediction or output is reached.

The Naive Bayes algorithm is a probabilistic classification algorithm based on Bayes' theorem [15]. It relies on the independence assumption, which allows the model to produce results even when some data are missing [11]. It is easy to construct without requiring parameter estimation, making it practical and efficient for large datasets [16].

The K-Nearest Neighbors (KNN) algorithm is a machine learning algorithm used for both classification and regression tasks [11]. It is also known as a non-parametric machine learning algorithm [17]. Due to its simplicity and fast training process, it is widely used in many applications. In classification problems, KNN identifies the K nearest training points for a new instance. Then, the class labels of these neighbors are examined, and the majority class is assigned to the new instance. This approach assumes that similar instances tend to belong to the same class. In this process, distances between data points are measured using methods such as the Euclidean, Manhattan, and Minkowski distances [7].

3.3. Classification Algorithms

Explainable Artificial Intelligence (XAI) is a research field that focuses on interpreting complex models and enabling artificial intelligence models and their predictions to be understood by humans [4]. It is used to make the functioning of machine learning and artificial intelligence models more transparent. It not only ensures that the decisions made by a model are interpretable but also evaluates whether the model's purpose can be clearly explained [18]. Various methods are used to perform these interpretation and explanation processes. Among these, SHAP and LIME are considered within the scope of this study. SHAP is based on Shapley values from game theory [5]. The Shapley value measures the average marginal contribution of each feature to the model output by considering all possible combinations of feature values [3]. Using these values, SHAP provides a framework for understanding how a model generates its predictions and how each feature influences the model [19]. It is a tool for interpreting black-box models and provides explanations at both global and local levels.

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

$\emptyset_i$: Shapley value for feature i
N: set of all features
S: subset of features
$f(S)$: model prediction obtained using the subset S

LIME is a widely used, model-agnostic approach that provides interpretable and explainable insights into model predictions [19]. This method generates small, interpretable models that approximate the behavior of the original model. Each interpretable model is valid only for a specific instance (x), meaning that within the local neighborhood of this instance, it produces predictions that are very close to those of the original model [20]. This property provides a locally interpretable explanation of individual predictions from machine learning models [3]. LIME provides explanations only at the local level.

$$Explanation\ (x) = arg\ \min_{g \epsilon G} L(f, g, \pi_x) + \Omega(g) \quad (4)$$

explanation (x) : local explanation for instance x
L(f, g, π_x) : loss function measuring how well the explanation model $g$approximates the original model$f$
π_x: proximity function defining locality around x
Ω(g): complexity function

4. Methodology

4.1. Data Preprocessing and Modeling

In this study, the Telco Customer Churn (11.1.3+) dataset, publicly available on Kaggle (https://www.kaggle.com/datasets/alfathterry/telco-customer-churn-11-1-3), was used. The dataset consists of 7,043 customer records and 35 variables. The distribution of the target variable (Churn) includes 5,174 non-churn ("No") instances (73.46%) and 1,869 churn ("Yes") instances (26.54%).

In the initial stage of data preprocessing, the dataset's variable types were analyzed. Incorrectly defined data types were identified and converted into appropriate formats. The variables converted into appropriate formats are presented in Figure 1.

Gender	object
Age	int64
Under 30	object
Senior Citizen	object
Married	object
Dependents	object
Number of Dependents	int64
Country	object
State	object
City	object
Referred a Friend	object
Number of Referrals	int64
Tenure in Months	int64
Offer	object
Phone Service	object
Avg Monthly Long Distance Charges	float64
Multiple Lines	object
Internet Service	object
Internet Type	object
Avg Monthly GB Download	int64
Contract	object
Paperless Billing	object
Payment Method	object
Monthly Charge	float64
Total Charges	float64
Total Refunds	float64
Total Long Distance Charges	float64
Total Revenue	float64
Satisfaction Score	int64
Customer Status	object
Churn Label	object
Churn Score	int64
CLTV	int64
Churn Category	object
Churn Reason	object

Figure 1. Variable Types

In the second step, missing values were analyzed. As shown in Figure 2, these values, determined to be incomplete, were imputed with values appropriate to the data structure. To preserve data integrity and obtain more accurate results after modeling, variables with more than 25% missing values were removed from the dataset. For variables containing less than 25% missing observations and retained in the dataset, the appropriate imputation method was determined. For these variables, measures of central tendency, such as the mean and median, and distribution characteristics, such as minimum and maximum values, were calculated. Based on these findings, missing values were imputed using either the mean or median values.

Gender	0
Age	0
Under 30	0
Senior Citizen	0
Married	0
Dependents	0
Number of Dependents	0
Country	0
State	0
City	0
Referred a Friend	0
Number of Referrals	0
Tenure in Months	0
Offer	3877
Phone Service	0
Avg Monthly Long Distance Charges	2049
Multiple Lines	0
Internet Service	0
Internet Type	1526
Avg Monthly GB Download	0
Contract	0
Paperless Billing	0
Payment Method	0
Monthly Charge	838
Total Charges	107
Total Refunds	147
Total Long Distance Charges	281
Total Revenue	23
Satisfaction Score	0
Customer Status	0
Churn Label	0
Churn Score	0
CLTV	0
Churn Category	5174
Churn Reason	5174

Figure 2. Missing Data Analysis

As another step in the data preprocessing process, the dataset was examined for outliers. The dataset was analyzed, and it was observed that most variables did not follow a normal distribution. Since normality was not met, the Interquartile Range (IQR) method was used to detect outliers. Variables identified as containing outliers using the IQR method were further examined. Following this analysis, it was determined that some variables contained many outliers. It was considered that a high number of outliers could negatively affect the analysis results. Therefore, variables with many outliers were removed from the dataset. However, in some cases, although certain data points appeared to be statistical outliers, they were retained in the dataset as they were structurally meaningful.

Since most machine learning algorithms operate on numerical data, data transformation was performed in the next stage. For categorical variables with two classes, the label encoding method was applied. Ordinal encoding was used for variables with an inherent order. For nominal variables with more than two categories, one-hot encoding was applied to better represent the effect of different category values. The details of these processes are presented in Table 1.

Table 1. Variable Transformation

Variable Name	Variable Value	Operation
Gender	“Male”, “Female”	Label Encoding
Under 30	“No”, “Yes”	Label Encoding
Senior Citizen	“Yes”, “No”	Label Encoding
Married	“No”, “Yes”	Label Encoding
Dependents	“No”, “Yes”	Label Encoding
Country	“United States”	It was removed because it contained only a single variable.
State	“California”	It was removed because it contained only a single variable.
City	“Los Angeles”, “Inglewood”, “Whitter”, “Topaz”…..	It was removed because grouping could not be performed.
Referred a Friend	“No”, “Yes”	Label Encoding
Offer	“Offer E”, “Offer D”, “Offer C”…	One Hot Encoding

Phone Service	"No", "Yes"	Label Encoding
Multiple Lines	"No", "Yes"	Label Encoding
Internet Service	"No", "Yes"	Label Encoding
Internet Type	"DSL", "Fiber optic", "Cable", "No Service"	One Hot Encoding
Contract	"Month-to-Month", "One Year", "Two Year"	Ordinal Encoding
Paperless Billing	"Yes", "No"	Label Encoding
Payment Method	"Bank Withdrawal", "Credit Card", "Mailed Check"	One Hot Encoding
Customer Status	"Churned", "Stayed", "Joined"	It was removed because it corresponds to the variables in the Churn Label.
Churn Label	"Yes", "No"	Label Encoding
Churn Category	"Competitor", "Dissatisfaction", "Price", "Other", "Attitude", "No Churn"	It was removed because it corresponds to the variables in the Churn Label.
Churn Reason	"Competitor offered more data", "Competitor made better offer", "Limited range of services", "Extra data charges", "Product dissatisfaction"…	It was removed because grouping could not be performed.

After the data preprocessing stage, the modeling phase was initiated. The dataset was divided into training and test sets, with 80% of the data used for training and 20% for testing. The random seed was set to 42 to ensure reproducibility. Since the dataset was imbalanced, the SMOTE method was applied to mitigate the imbalance. The SMOTE method generates synthetic samples for the minority class in classification problems, thereby creating a more balanced dataset. SMOTE was applied only to the training data after the train–test split, ensuring that sufficient minority-class samples were obtained for model training.

In the final step, standardization was applied to the dataset to eliminate differences in feature scales and to enable the algorithms to produce more accurate results. To prevent overfitting and data leakage, standardization was first applied to the training data and then to the test data. This approach ensured that the model relied solely on the training data, while the test data was evaluated independently. As a result of all these preprocessing steps, the dataset was prepared appropriately for analysis and modeling.

In the modeling phase, models were developed using Logistic Regression, Random Forest, Decision Tree, Gaussian Naive Bayes (NB), and K-Nearest Neighbors (KNN) algorithms. For the Logistic Regression model, the penalty parameter was set to L2, the C value was set to 0.001, and the maximum number of iterations was set to 1000. In addition, 10-fold cross-validation was applied. For the Random Forest model, the n_estimators value was set to 100, the max_depth value was set to 10, and the random_state value was set to 42. For the Decision Tree model, the criterion parameter was set to gini, the max_depth value was set to 10, the min_samples_split value was set to 10, and the random_state value was set to 42. For the Gaussian Naive Bayes model, GridSearchCV was applied for the var_smoothing parameter; however, the best performance was achieved with the default value. For the K-Nearest Neighbors model, the n_neighbors value was set to 5. The performance of the models was evaluated using classification performance metrics such as accuracy, precision, recall, F1-score, and the ROC–AUC curve. The obtained results are presented in Table 2 below.

According to the results presented in Table 2, the Random Forest algorithm achieved the highest performance in terms of accuracy, precision, and F1-score metrics. The precision metric was considered an important classification criterion in this study because it indicates the proportion of customers predicted as churn by the model who actually churned.

Table 2. Machine Learning Model Results

Algorithms	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0,9347	0,8711	**0,8850**	0,8780	**0,9830**
Random Forest	**0,9560**	**0,9515**	0,8690	**0,9129**	0,9808
Decision Trees	0,9248	0,8472	0,8743	0,8605	0,9087
Naive Bayes	0,8481	0,6951	0,7629	0,7270	0,9172
K-Nearest Neighbors	0,9368	0,8841	0,8770	0,8805	0,9570

4.2. Explainable AI

Explainable artificial intelligence (XAI) techniques were applied to the Random Forest algorithm, which achieved the highest accuracy among the models used in the study and is also interpretable. To explain how the implemented machine learning model makes decisions, SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) methods were employed. In this study, SHAP was used for global explanations, while both SHAP and LIME were utilized together for local comparison. Using SHAP, a global bar plot and a summary plot were preferred to illustrate the importance of variables in the model. Since the Random Forest model is a tree-based algorithm, the Tree SHAP method was applied.

To better understand customer churn predictions at the individual level, local analysis was also conducted. The customer with the highest churn probability was selected, and local interpretability techniques were applied. SHAP and LIME methods were used locally, and the explanations generated by both were compared for consistency and interpretability. Using these explainable AI techniques, both the model's overall behavior and its individual-level predictions were analyzed.

4.2.1. Global explainability with SHAP

The variables influencing the model output were analyzed using the Tree SHAP method on the Random Forest algorithm.

Figure 3. SHAP Global Bar Plot

Figure 3. The SHAP Global Bar Plot illustrates which variables have the greatest influence on the model. As shown in Figure 3, the Satisfaction Score is the most influential variable. Other important variables include the customer's contract type (Contract_Encoded), payment method (Payment Method_Credit Card), the number of referrals made by the customer (Number of Referrals), and the customer's tenure with the company (Tenure in Months). These results indicate that the model is particularly influenced by customer satisfaction and contract structure. Since the Satisfaction Score appears to be highly dominant compared to other variables, it may be necessary to remove this variable from the dataset and reassess the results.

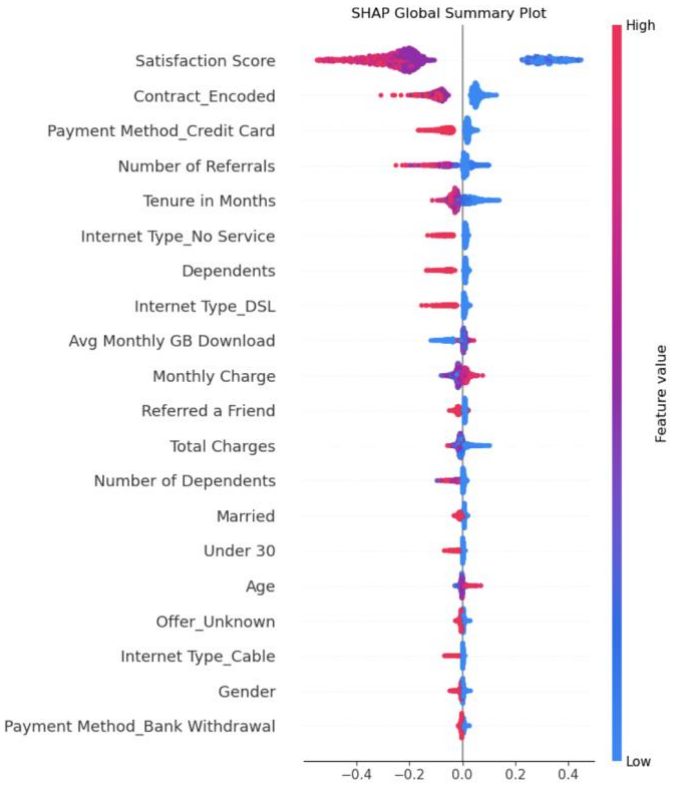

Figure 4. SHAP Global Summary Plot

Figure 4. The SHAP Global Summary Plot shows the distribution of SHAP values for each feature. The horizontal axis shows SHAP values, and the vertical axis shows the variables. The right side indicates an increase in the likelihood of customer churn, whereas the left side indicates a decrease. The color of the points indicates the feature value, with red indicating high values and blue indicating low values. Variables positioned at the top of the y-axis are more important for the model.

As can be observed from Figure 4., the Satisfaction Score is the strongest determinant. High satisfaction (i.e., higher Satisfaction Score values) is mostly located on the left side and reduces the likelihood of churn, whereas low satisfaction is located on the right and increases the risk of churn. When the customer's contract type (Contract_Encoded) and tenure (Tenure in Months) have higher values, the points are again concentrated on the left side. This indicates that longer contracts and greater customer tenure reduce churn risk. Another variable, the number of referrals made by the customer (Number of Referrals), shows that as the number increases, the risk of churn decreases. In contrast, higher values of the Monthly Charge variable tend to shift to the right, indicating an increased risk of churn. Thus, this plot clearly presents both the importance of the variables and the direction of their effects.

4.2.2. Local Explainability with SHAP and LIME

Local SHAP Analysis

Figure 5. The local SHAP Waterfall plot explains why a customer who received a Random Forest model prediction canceled the subscription. Blue bars represent variables that decrease the probability of customer churn, while red bars represent variables that increase it. As shown in Figure 5, the churn probability for customer 38 (idx: 38) is 89%.

On the left side of the plot, the variables that influence the model's decision are presented. The Satisfaction Score is identified as the most influential variable in the model's decision. The customer's contract type (Contract_Encoded) and payment method via credit card (Payment Method_Credit Card) are other variables that significantly affect the model's decision. On the other hand, variables indicated by blue arrows, such as credit card payments, reduce the risk of churn. However, the combined effect of these variables is not sufficient to offset the strong influence of the Satisfaction Score, and the customer remains in the high-risk group.

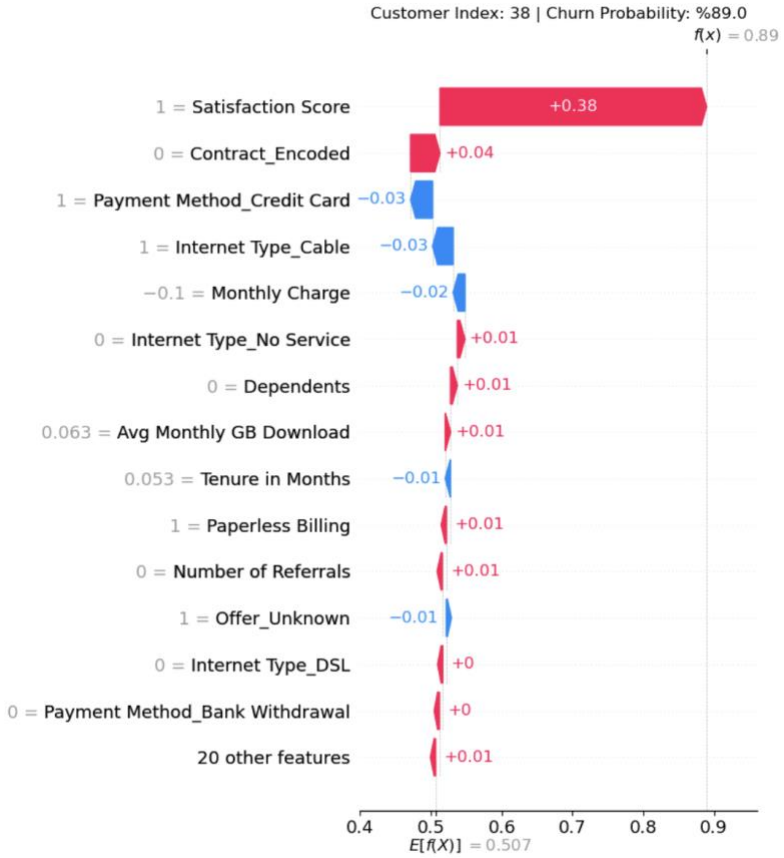

Figure 5. Local SHAP Waterfall Plot

Local LIME Analysis

Figure 6. presents the local LIME plot explaining why a customer churned. In the plot, green bars represent features that increase the probability of customer churn, while red bars represent features that decrease it. The left side of the plot shows the variables influencing the model's decision.

The model's decision is most influenced by the Satisfaction Score. In other words, this variable has the strongest impact on customer churn. Other influential variables include the customer's contract type (Contract_Encoded), whether the customer has dependents (Dependents), whether the internet service is unavailable (Internet Type_No Service), and whether the internet service is DSL (Internet Type_DSL). The red bars in the plot represent variables that do not change the churn outcome but reduce the probability of churn.

The R^2 value reported in the LIME analysis indicates how accurately the LIME explanation reflects the model's decision. The R^2 value ranges between 0 and 1, and as it approaches 1, the success of the explanation in representing the model increases. The value of 0.489 obtained in this study indicates that the LIME explanation reflects the model decision at a limited level.

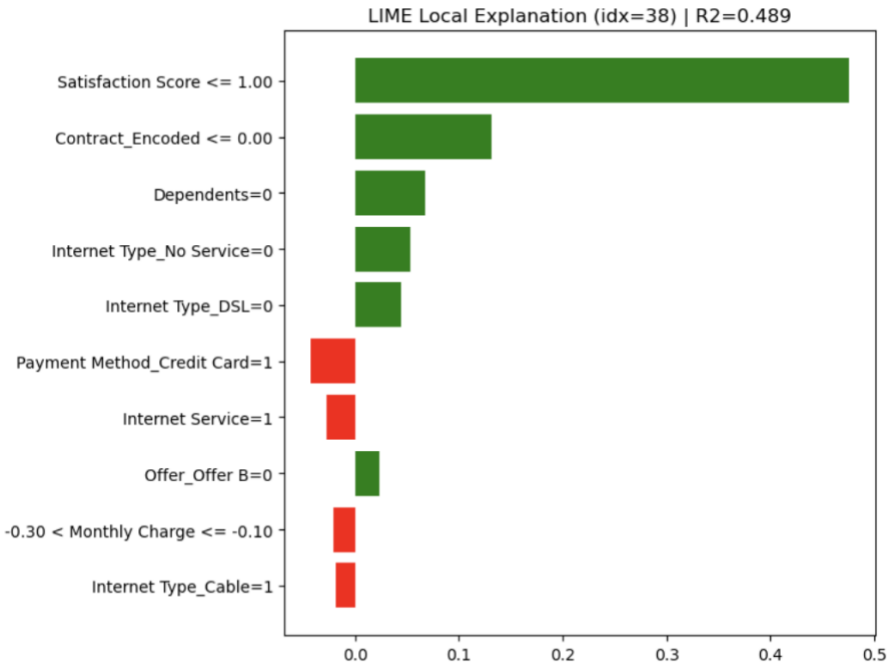

Figure 6. LIME Local Explainability

When SHAP and LIME methods were applied to the selected customer, it was observed that many variables were consistent across the plots presented in Figures 5 and 6. In both methods, the Satisfaction Score emerged as the most influential variable. Although some differences were observed in certain variables, the variable importance rankings and the directions of effects provided by SHAP and LIME are generally consistent. While the LIME method provides an approximation of the model's behavior, the SHAP method computes the actual contribution values of the model.

5. Conclusion

In this study, various machine learning algorithms were applied to predict customer churn in the telecommunications sector, and the resulting models were comprehensively evaluated using explainable artificial intelligence techniques. Prior to the modeling phase, extensive data preprocessing was performed to improve data quality. In this context, missing data were analyzed and imputed using appropriate methods, outliers were identified, and their potential negative effects on the dataset were reduced, and categorical variables were transformed using suitable encoding techniques. To prevent class imbalance from negatively affecting model performance, a more balanced dataset was obtained using the SMOTE method. These preprocessing steps were found to improve the models' learning and contributed significantly to obtaining more accurate and consistent results.
Compared with other classification algorithms used in the study, the Random Forest model achieved the highest performance across accuracy, precision, and F1-score metrics. This finding indicates that tree-based methods achieved better results in this study.

To evaluate the model not only on performance but also on the interpretability of its decision-making process, explainable artificial intelligence methods were employed. Using the SHAP method, the variables influencing the model at the global level and their effects on the probability of customer churn were identified. The analysis results showed that the Satisfaction Score was the most influential variable in the model.

At the local level, SHAP and LIME analyses were conducted to examine in detail how model predictions are formed for individual customers. The results indicate that both methods generally provide consistent explanations. Since SHAP directly computes feature contributions, it provides a more reliable representation of the impact of variables on the model. LIME, on the other hand, supports interpretability by offering a locally approximated representation of the model's behavior.

Overall, the study concludes that machine learning algorithms can predict customer churn with high accuracy, and that explainable artificial intelligence methods significantly enhance the interpretability

and reliability of these predictions. Therefore, the findings are expected to contribute to the development of decision support systems in the telecommunications sector.

This study contributes to the literature by addressing customer churn prediction not only from a model performance perspective but also from an explainability perspective. The application of SHAP at both global and local levels, along with comparisons between SHAP and LIME, enables a more comprehensive understanding of model behavior.

6. Discussion

The findings of this study should be interpreted by considering several limitations. First, the use of a single telecommunications dataset limits the generalizability of the findings to other datasets.

Furthermore, the results indicated that the Satisfaction Score variable had a dominant effect on customer churn prediction. To evaluate its impact on the model, the Satisfaction Score variable was removed from the dataset, and the models were retrained and their performance metrics were re-evaluated. However, removing this variable resulted in a decrease in model performance metrics; therefore, the Satisfaction Score variable was retained in the final model. Nevertheless, the dominant influence of this variable on the model's decisions suggests that the prediction results may be strongly associated with customer satisfaction, which should be considered as a limitation of this study.

As shown in Figure 6, the LIME analysis yielded an R^2 value of 0.489, indicating that the local surrogate model represents the predictions of the original model only to a limited extent. The R^2 value ranges from 0 to 1, with values closer to 1 indicating that the LIME explanation more accurately represents the original model. The obtained R^2 value of 0.489 indicates that the LIME explanation reflects the model's decision only to a limited extent and that the local fidelity of the explanation is limited. Therefore, LIME explanations should be interpreted with caution, as the local explanation model may not adequately represent the overall decision-making behavior of the original model. This should be considered as a limitation of the explainability analysis conducted in this study.

Finally, the comparison between SHAP and LIME was performed using only a single customer instance. Therefore, the explainability results reflect the model's behavior only for the analyzed customer and may not be generalizable to different customer profiles. Future studies may provide a more comprehensive evaluation of model decisions by performing explainability analyses on different customer groups and across multiple datasets.

References

[1] M. A. Aydın, "Müşteri kaybı tahmininde sınıf dengesizliği problemi," *Politeknik Dergisi*, vol. 25, no. 1, pp. 351–360, 2022. doi: 10.2339/politeknik.734916

[2] O. Kaynar, M. F. Tuna, Y. Görmez, and M. A. Deveci, "Makine öğrenmesi yöntemleriyle müşteri kaybı analizi," *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, vol. 18, no. 1, pp. 1–14, 2017.

[3] İ. Kök, "Açıklanabilir Yapay Zekaya Dayalı Müşteri Kaybı Analizi ve Elde Tutma Önerisi," *Mühendislik Bilimleri ve Araştırmaları Dergisi*, vol. 6, no. 1, pp. 13–23, 2024. doi: 10.46387/bjesr.1344414

[4] I. Boukrouh and A. Azmani, "Explainable machine learning models applied to predicting customer churn for e-commerce," *Int. J. Artif. Intell*., vol. 14, no. 1, pp. 286–297, Feb. 2025. doi: 10.11591/ijai.v14.i1.pp286-297

[5] K. Arslanbay, "Explainability in Machine Learning Models for Churn Prediction," M.S. thesis, Faculty of Economics and Management, Czech University of Life Sciences Prague, Prague, Czech Republic, 2025.

[6] 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, 2020. doi: 10.21917/ijsc.2020.0291

[7] B. Uzak, "Telekomünikasyon sektöründe çalışan kaybı tahmini için makine öğrenmesi modeli seçimi," M.S. thesis, Inst. of Science, Bursa Uludağ University, Bursa, Turkey, 2022.

[8] U. Ahmed, A. Khan, S. H. Khan, A. Basit, I. U. Haq, and Y. S. Lee, "Transfer learning and meta classification based deep churn prediction system for telecom industry," arXiv preprint arXiv:1901.06091, 2019. [Online]. Available: https://arxiv.org/abs/1901.06091. [Accessed: May 4, 2026].

[9] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," SN Computer Science, 2021.

[10] T. W. L. Stuart, "The prediction of customer churn in a telecommunication company: Comparing different machine learning techniques," M.S. thesis, Tilburg University, Tilburg, Netherlands, 2021.

[11] F. Hassan Mohamed, "Churn prediction in telecommunication sector," M.S. thesis, Graduate School of Natural and Applied Sciences, Istanbul Commerce University, Istanbul, Turkey, 2021.

[12] S. K. Pasbunat, "Dynamic Churn Prediction Using Machine Learning in Telecom Industries," Ph.D. dissertation, California State University, Northridge, CA, USA, 2023.

[13] S. Zhao, "Customer Churn Prediction Based on the Decision Tree and Random Forest Model," *BCP Business & Management*, vol. 44, pp. 339–344, Apr. 2023. doi: 10.54691/bcpbm.v44i.4840

[14] F. Uyanık and M. C. Kasapbaşı, "Telekomünikasyon sektörü için veri madenciliği ve makine öğrenmesi teknikleri ile ayrılan müşteri analizi," *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, vol. 9, no. 3, pp. 172–191, May 2021. doi: 10.29130/dubited.807922

[15] M. Büyükkeçeci and M. C. Okur, "Müşteri Kayıplarının Tahmini Üzerine Bir Veri Madenciliği Uygulaması," *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi,* vol. 24, no. 72, pp. 887–900, Sep. 2022. doi: 10.21205/deufmd.2022247218

[16] G. Esteves and J. Mendes-Moreira, "Churn prediction in the telecom business," *in Proceedings of the 2016 11th International Conference on Digital Information Management (ICDIM), September 2016, Porto, Portugal*, pp. 254–259.

[17] N. Taskin, "Customer Churn Prediction Model in Telecommunication Sector Using Machine Learning Technique," M.S. thesis, Dept. of Statistics, Uppsala University, Uppsala, Sweden, 2023.

[18] P. Gandhi, "Explainable artificial intelligence," KDnuggets, Jan. 2019. [Online]. Available: https://www.kdnuggets.com/2019/01/explainable-ai.html. [Accessed: Feb. 26, 2026].

[19] C. Özkurt, "Transparency in Decision-Making: The Role of Explainable AI (XAI) in Customer Churn Analysis," Information Technology in Economics and Business, vol. 2, no. 1, pp. 1–11, Jan. 2025. doi: 10.69882/adba.iteb.2025011

[20] G. Visani, E. Bagli, F. Chesani, A. Poluzzi, and D. Capuzzo, "Statistical stability indices for LIME: Obtaining reliable explanations for machine learning models," Journal of the Operational Research Society, vol. 73, no. 1, pp. 91–101, 2022. doi: 10.1080/01605682.2020.1865846