Original Article

Check for updates

Available online at
www.heca-analitika.com/ijds

Infolitika Journal of Data Science

Vol. 2, No. 1, 2024

A Model-Agnostic Interpretability Approach to Predicting Customer Churn in the Telecommunications Industry

Teuku Rizky Noviandy [1,*], Ghalieb Mutig Idroes [1], Irsan Hardi [1], Mohd Afjal [2] and Samrat Ray [3]

[1] Interdisciplinary Innovation Research Unit, Graha Primera Saintifika, Aceh Besar, 23771, Indonesia; trizkynoviandy@gmail.com (T.R.N.); ghaliebidroes@outlook.com (G.M.I.); irsan.hardi@gmail.com (I.H.)
[2] VIT Business School, Vellore Institute of Technology, Vellore 632014, India; afzalmfc@gmail.com (M.A.)
[3] Business Analytics, International Institute of Management Studies, Pune 411002, India; s.ray@iimspune.edu.in (S.R.)

* Correspondence: trizkynoviandy@gmail.com

Article History

Received 13 March 2024
Revised 11 May 2024
Accepted 19 May 2024
Available Online 27 May 2024

Keywords:
Machine learning
Model interpretability
Customer retention
SHAP analysis
Predictive analytics

Abstract

Customer churn is critical for businesses across various industries, especially in the telecommunications sector, where high churn rates can significantly impact revenue and growth. Understanding the factors leading to customer churn is essential for developing effective retention strategies. Despite the predictive power of machine learning models, there is a growing demand for model interpretability to ensure trust and transparency in decision-making processes. This study addresses this gap by applying advanced machine learning models, specifically Naïve Bayes, Random Forest, AdaBoost, XGBoost, and LightGBM, to predict customer churn in a telecommunications dataset. We enhanced model interpretability using SHapley Additive exPlanations (SHAP), which provides insights into feature contributions to predictions. Here, we show that LightGBM achieved the highest performance among the models, with an accuracy of 80.70%, precision of 84.35%, recall of 90.54%, and an F1-score of 87.34%. SHAP analysis revealed that features such as tenure, contract type, and monthly charges are significant predictors of customer churn. These results indicate that combining predictive analytics with interpretability methods can provide telecom companies with actionable insights to tailor retention strategies effectively. The study highlights the importance of understanding customer behavior through transparent and accurate models, paving the way for improved customer satisfaction and loyalty. Future research should focus on validating these findings with real-world data, exploring more sophisticated models, and incorporating temporal dynamics to enhance churn prediction models' predictive power and applicability.

1. Introduction

Customer churn is a significant concern for businesses across various industries, representing the phenomenon where customers cease their relationship with a company [1]. High churn rates can adversely affect a company's revenue and growth, prompting the need for effective strategies to predict and mitigate churn [2, 3]. By understanding the factors that lead to customer churn, businesses can develop targeted interventions to retain valuable customers and enhance their overall service experience [4–6].

In the telecommunications industry, churn poses a particularly challenging problem due to the highly competitive market landscape [7]. Telecom companies face the constant threat of customers switching to rival providers, driven by factors such as pricing, service quality, and technological advancements [8]. Identifying the reasons behind customer churn and predicting which customers are likely to leave can provide telecom companies with crucial insights to tailor their retention strategies and improve customer satisfaction.

In recent years, machine learning has emerged as a powerful tool for predictive analytics, offering advanced techniques to analyze large datasets and uncover hidden patterns [9–11]. With their ability to process complex data and generate accurate predictions, machine learning models have been widely adopted across various domains, including customer churn prediction [12]. The application of machine learning in churn prediction allows companies to leverage data-driven approaches to enhance decision-making processes and optimize business outcomes.

Many studies use machine learning techniques to improve accuracy and provide useful insights in customer churn prediction. Researchers have tried various models to identify churn patterns and factors. For example, Bhuse et al. utilized support vector machines, XGBoost, random forests, and deep neural networks, with grid search hyperparameter tuning, finding that the random forest model achieved the highest accuracy at 90.96% [13]. Similarly, Zhang et al. applied Fisher discriminant equations and logistic regression analysis to develop a telecom customer churn prediction model, which achieved an accuracy of 93.94% [14]. Furthermore, Liu et al. implemented k-means clustering to categorize different consumer groups and used ensemble learning techniques to predict customer churn effectively [15].

Despite the predictive power of machine learning models, there is a growing demand for model interpretability and explainability [16–18]. Businesses and stakeholders seek to understand the reasoning behind model predictions to ensure trust and transparency in decision-making. Techniques such as SHAP (SHapley Additive exPlanations) have been developed to provide model-agnostic interpretability, offering insights into the contribution of each feature in the prediction process [19, 20]. This level of explainability is essential for gaining actionable insights and building confidence in deploying machine learning models in real-world scenarios.

While several methods are available for interpreting machine learning models, SHAP was chosen for its unique advantages. Unlike other techniques, such as LIME (Local Interpretable Model-agnostic Explanations) or feature importance scores, SHAP values provide a unified measure of feature importance that is consistent and interpretable across different models [21, 22]. SHAP uses principles from game theory to fairly distribute the prediction's output among the contributing features, ensuring that each feature's impact is considered about others. This method offers global interpretability (understanding overall model behavior) and local interpretability (explaining individual predictions), making it highly versatile and powerful.

This study aims to predict customer churn in the telecommunications industry using machine learning models and enhance model interpretability by applying SHAP. By analyzing a comprehensive dataset from a fictional telecommunications company, we strive to identify key factors influencing churn and develop robust predictive models. The ultimate goal is to provide telecom companies with valuable tools to understand customer behavior better and implement effective retention strategies, thereby reducing churn rates and improving customer loyalty.

2. Materials and Methods

The methodology for predicting customer churn involves several steps: dataset acquisition and data preprocessing. Subsequently, we implement and evaluate various machine learning models, including Naïve Bayes, Random Forest, AdaBoost, XGBoost, and LightGBM, to determine their effectiveness in predicting churn. Performance evaluation and model interpretability are crucial to ensure reliable and understandable results. The overall process is depicted in Figure 1.

2.1. Dataset

This study utilizes the telco customer churn dataset, which contains information about a fictional telecommunications company providing home phone and internet services to 7,043 customers in California during the third quarter [23]. The dataset is rich with details, including churn information, identifying customers who left within the last month. Of the total customers, 5,174 did not churn, while 1,869 did. This information is captured in the 'Churn' column. It also includes specifics about the services each customer has signed up for, such as phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies.

Additionally, the dataset provides comprehensive customer account information, detailing how long customers have been with the company, their contract type, payment methods, paperless billing status, monthly charges, and total charges. Furthermore, demographic information is included, offering insights into customers' gender, age range, and whether they have partners and

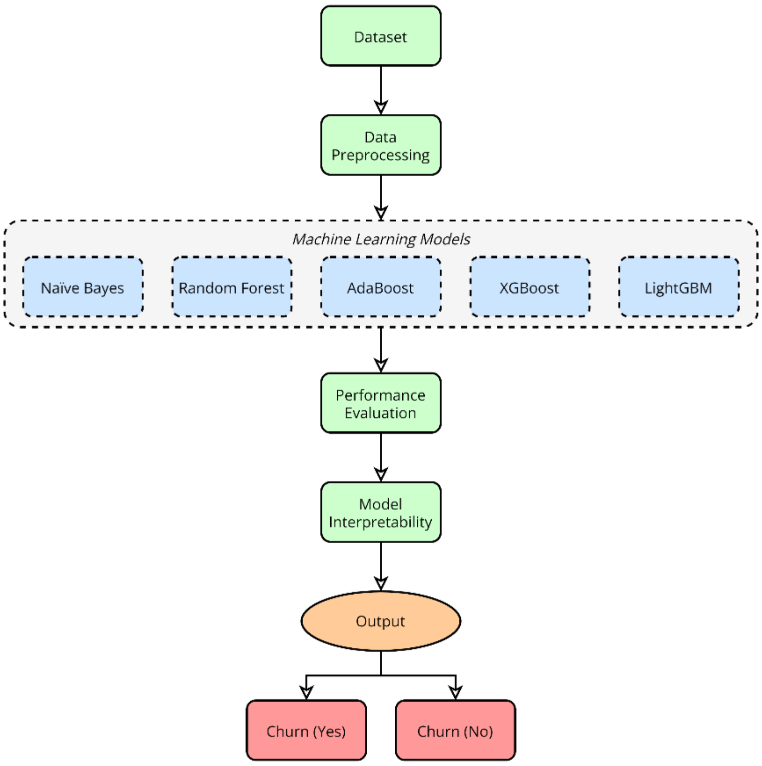

Figure 1. Workflow of this study.

Table 1. Detailed description of customer attributes and services in the telco customer churn dataset.

Variable Name	Type	Description
customerID	Categorical	Unique identifier for each customer
gender	Categorical	The gender of the customer
SeniorCitizen	Numerical	Indicates if the customer is a senior citizen (1: Yes, 0: No)
Partner	Categorical	Indicates if the customer has a partner (Yes/No)
Dependents	Categorical	Indicates if the customer has dependents (Yes/No)
tenure	Numerical	Number of months the customer has been with the company
PhoneService	Categorical	Indicates if the customer has phone service (Yes/No)
MultipleLines	Categorical	Indicates if the customer has multiple lines (Yes/No/No phone service)
InternetService	Categorical	Type of internet service provided (DSL/Fiber optic/No)
OnlineSecurity	Categorical	Indicates if the customer has online security service (Yes/No/No internet service)
OnlineBackup	Categorical	Indicates if the customer has online backup service (Yes/No/No internet service)
DeviceProtection	Categorical	Indicates if the customer has device protection service (Yes/No/No internet service)
TechSupport	Categorical	Indicates if the customer has tech support service (Yes/No/No internet service)
StreamingTV	Categorical	Indicates if the customer has streaming TV service (Yes/No/No internet service)
StreamingMovies	Categorical	Indicates if the customer has streaming movies service (Yes/No/No internet service)
Contract	Categorical	Type of contract (Month-to-month/One year/Two year)
PaperlessBilling	Categorical	Indicates if the customer has paperless billing (Yes/No)
PaymentMethod	Categorical	Payment method (Electronic check/Mailed check/Bank transfer (automatic)/Credit card (automatic))
MonthlyCharges	Numerical	Monthly charges incurred by the customer
TotalCharges	Numerical	Total charges incurred by the customer
Churn	Categorical	Indicates if the customer churned (Yes/No)

dependents. This extensive data enables a thorough analysis of customer behavior and factors influencing churn. The detailed descriptions of the customer attributes and services in the dataset are summarized in Table 1.

2.2. Data Preprocessing

The dataset contains categorical features, which are non-numeric data types representing categories or groups. To prepare these features for machine learning algorithms, we used label encoding. Label encoding transforms each category into a unique numerical value. Then, the dataset was split into two subsets: the training set (80%) and the testing set (20%) [24]. The training set builds and trains the model, allowing it to learn patterns and relationships within the data. The testing set evaluates the model's performance, providing an unbiased assessment of how well the model generalizes to new, unseen data.

2.3. Machine Learning Models

To build a model for customer churn prediction in this study, we employed five machine learning models: Naïve Bayes, Random Forest, AdaBoost, XGBoost, and LightGBM. Each of these models brings unique strengths to the table. Naïve Bayes, a probabilistic classifier based on Bayes' theorem, is known for its simplicity and effectiveness, particularly with large datasets [25]. Random Forest, an ensemble method, constructs multiple decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees, which helps improve accuracy and control over-fitting [26].

AdaBoost, short for Adaptive Boosting, combines the predictions of several weak learners to form a strong learner, enhancing the performance of models by focusing on the errors of previous iterations [27]. XGBoost, an advanced gradient boosting technique, is renowned for its efficiency and speed and is often used in machine learning competitions for its superior performance [28, 29]. LightGBM, or Light Gradient Boosting Machine, is designed to be highly efficient and scalable, easily handling large-scale data and offering fast training speeds. We used these different models to make solid and accurate predictions about customer churn [30, 31].

For this study, we implemented our machine learning models using scikit-learn version 1.2.0. We used the default hyperparameters for each model to ensure consistency and reproducibility of our results. The decision to use default hyperparameters was made to compare the intrinsic strengths and weaknesses of each algorithm without the influence of extensive hyperparameter tuning [32]. This approach allows for a straightforward comparison of model performance and simplifies the process for potential replication of the study by other researchers.

2.4. Performance Evaluation

We employed several metrics to evaluate the performance of the machine learning models used for customer churn prediction: accuracy, precision, recall, and F1-score [33, 34]. Accuracy measures the overall correctness of the model by calculating the proportion of correctly classified instances out of the total instances. Precision, the ratio of true positive predictions to the total positive predictions made by the model, indicates the model's ability to correctly identify churners among the predicted positive cases. Recall, also known as sensitivity or true positive rate, measures the proportion of actual churners correctly identified by the model. The F1-score, the harmonic mean of precision and recall, provides a balanced evaluation, especially in scenarios with imbalanced class distributions, ensuring that false positives and negatives are accounted for in the model's performance assessment. The accuracy, precision, recall, and F1-score equations can be seen in Equations 1-4, respectively.

$$Accuracy = \frac{TP + FN}{FP + FN + TP + TN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{FN + TP} \tag{3}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

Where TP (True Positives) are the correctly predicted positive instances, TN (True Negatives) are the correctly predicted negative instances, FP (False Positives) are the incorrectly predicted positive instances, and FN (False Negatives) are the incorrectly predicted negative instances.

Additionally, we calculated the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) to further evaluate the models. The AUC-ROC provides a comprehensive measure of the model's ability to distinguish between churners and non-churners across various threshold settings [35]. An AUC score of 0.5 indicates a model with no discriminative power, equivalent to random guessing, while a score of 1.0 represents a perfect model [36].

2.5. Model Interpretability

In addition to evaluating the predictive performance of our machine learning models, we aimed to enhance their interpretability using SHAP. This method provides a unified approach to explain the output of any machine learning model by assigning each feature an importance value for a particular prediction. This method is grounded in cooperative game theory, where the contribution of each feature is fairly distributed based on its impact on the prediction, enabling a deeper understanding of the underlying factors driving customer churn.

By applying SHAP, we can interpret individual predictions and aggregate the results to identify the most influential features across all instances. This level of interpretability allows us to pinpoint specific variables that significantly contribute to churn, such as contract type, monthly charges, and tenure. By visualizing these insights through

Table 2. Performance metrics for five machine learning models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Naïve Bayes	75.80	90.17	75.29	82.06
Random Forest	79.99	83.13	91.31	87.03
AdaBoost	80.55	84.83	89.58	87.14
XGBoost	79.35	83.41	89.77	86.47
LightGBM	80.70	84.35	90.54	87.34

Table 3. Confusion matrices for five machine learning models.

Model	Prediction	Actual	
		No	Yes
Naïve Bayes	No	780	256
	Yes	85	288
Random Forest	No	946	90
	Yes	192	181
AdaBoost	No	928	108
	Yes	166	207
XGBoost	No	930	106
	Yes	185	188
LightGBM	No	938	98
	Yes	174	199

SHAP summary and dependence plots, we provide telecom companies with actionable information to refine their customer retention strategies. The interpretability offered by SHAP not only helps gain stakeholders' trust but also ensures transparency in the decision-making process, facilitating the deployment of machine learning models in real-world scenarios with greater confidence.

3. Results and Discussion

In this section, we present the performance results of the machine learning models used for customer churn prediction and discuss the insights derived from these findings. The evaluation metrics for each model, including accuracy, precision, recall, and F1-score, are summarized in Table 2.

The results show that all five models perform well in predicting customer churn, with accuracy scores ranging from 75.80% to 80.70%. The LightGBM model achieved the highest accuracy at 80.70%, followed closely by AdaBoost with 80.55% and Random Forest with 79.99%. In terms of precision, the Naïve Bayes model stands out with a high precision score of 90.17%, indicating its strong ability to identify churners among the predicted positive cases correctly. However, it has a lower recall (75.29%) than the other models, which suggests it may miss some actual churners.

The recall scores highlight the models' ability to identify actual churners, with Random Forest and LightGBM achieving the highest recall values of 91.31% and 90.54%, respectively. The F1-score, which balances precision and recall, is highest for LightGBM at 87.34%, followed by AdaBoost at 87.14% and Random Forest at 87.03%. These results indicate that LightGBM and AdaBoost provide the most balanced performance across all metrics, making them particularly effective for customer churn prediction.

The superior performance of LightGBM and AdaBoost can be attributed to their advanced boosting techniques, which enhance the models' ability to capture complex patterns in the data. The high precision of Naïve Bayes, despite its lower overall accuracy and recall, suggests it could be useful in scenarios where the cost of false positives is particularly high. The consistent performance of Random Forest across different metrics underscores its robustness and reliability in handling large and complex datasets. These findings demonstrate the effectiveness of using machine learning models for predicting customer churn in the telecommunications industry, providing valuable insights for developing targeted retention strategies.

Next, we present the confusion matrices for each model (Table 3). These matrices provide a detailed breakdown of the models' performance by showing the number of true positives, true negatives, false positives, and false negatives. They offer additional insights into the models' predictive capabilities and potential areas for improvement.

The confusion matrices for the different models provide a deeper understanding of their performance by detailing the number of true positives, true negatives, false positives, and false negatives. The Naïve Bayes model, while exhibiting high precision with 90.17%, shows a significant number of false negatives (256), indicating its limitation in identifying all actual churners.

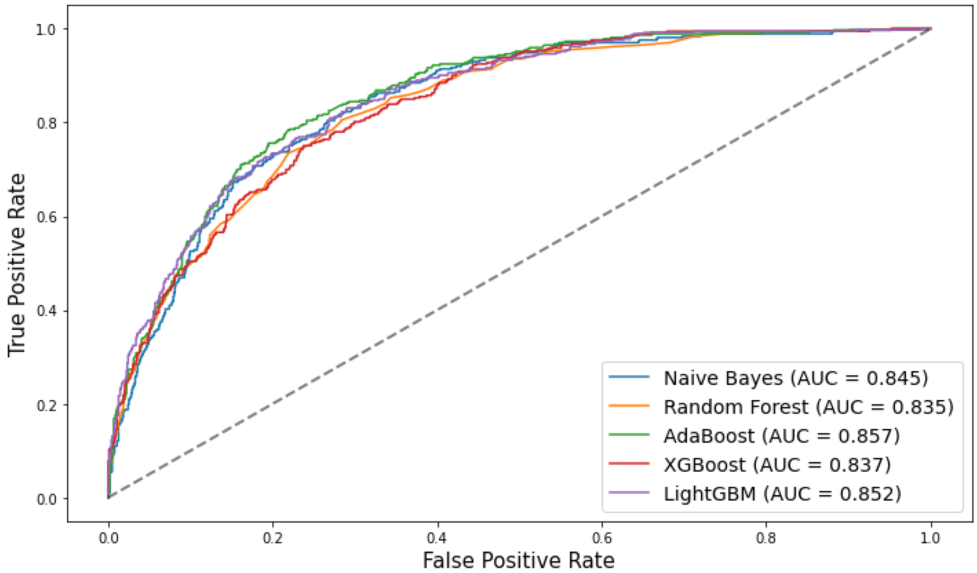

Figure 2. ROC curve for five machine learning models.

This is contrasted by its relatively lower false positives (85), suggesting it is more conservative in predicting churn. The Random Forest model balances its predictions more effectively, with 90 false negatives and 192 false positives, demonstrating a robust ability to capture most churners but at the cost of misclassifying some non-churners.

AdaBoost and XGBoost models display similar performance patterns, with AdaBoost having 108 false negatives and 166 false positives, and XGBoost showing 106 false negatives and 185 false positives. These models effectively identify churners but with varying degrees of false positive rates. LightGBM, the best performer in overall evaluation metrics, achieves the fewest false negatives (98) and maintains a reasonable number of false positives (174). This balance indicates LightGBM's superior predictive power and ability to minimize misclassifications. These insights from the confusion matrices highlight the practical implications of using different machine learning models, enabling telecom companies to make informed decisions in implementing effective customer retention strategies.

The ROC AUC curve depicted in Figure 2 visually represents the performance of each machine learning model in distinguishing between churners and non-churners. The AUC values for the models are as follows: Naïve Bayes (0.845), Random Forest (0.835), AdaBoost (0.857), XGBoost (0.837), and LightGBM (0.852). AdaBoost achieves the highest AUC score of 0.857, indicating its superior performance in identifying churners and non-churners across various threshold settings. LightGBM follows closely with an AUC of 0.852, further highlighting its strong predictive capabilities.

Naïve Bayes, despite its high precision, has an AUC of 0.845, reflecting its conservative prediction approach. XGBoost and Random Forest exhibit AUC scores of 0.837 and 0.835, respectively, demonstrating their robustness in churn prediction but with slightly lower overall discriminative power than AdaBoost and LightGBM. These AUC values corroborate the earlier evaluation metrics, confirming the effectiveness of AdaBoost and LightGBM in customer churn prediction. The ROC AUC curve thus reinforces the suitability of these models for deployment in real-world telecom scenarios, providing telecom companies with reliable tools to predict and mitigate customer churn effectively.

Given that the LightGBM model achieved the highest accuracy among the tested models, we selected it for further interpretability analysis using SHAP. LightGBM's robust performance, demonstrated by its balanced precision, recall, and high F1-score, makes it an ideal candidate for a deeper examination of the factors driving customer churn. By applying SHAP, we aim to uncover the contributions of individual features to the model's predictions, thereby providing actionable insights into customer behavior.

The SHAP bar plot for the LightGBM model, showcasing the mean SHAP values for the key features influencing customer churn predictions, is depicted in Figure 3. The feature 'tenure' emerges as the most significant predictor of customer churn, with a mean SHAP value of +0.13. This indicates that the duration of a customer's relationship with the telecom company strongly influences their likelihood of churning, with longer tenure generally reducing churn risk. The 'Contract' type is the second most important feature, with a mean SHAP value of

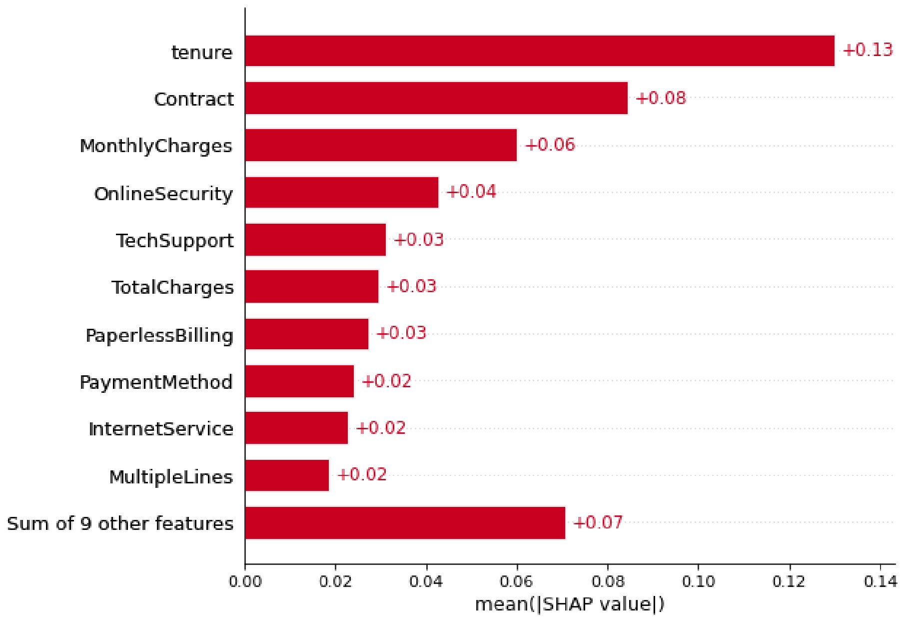

Figure 3. SHAP bar plot for LightGBM model's prediction.

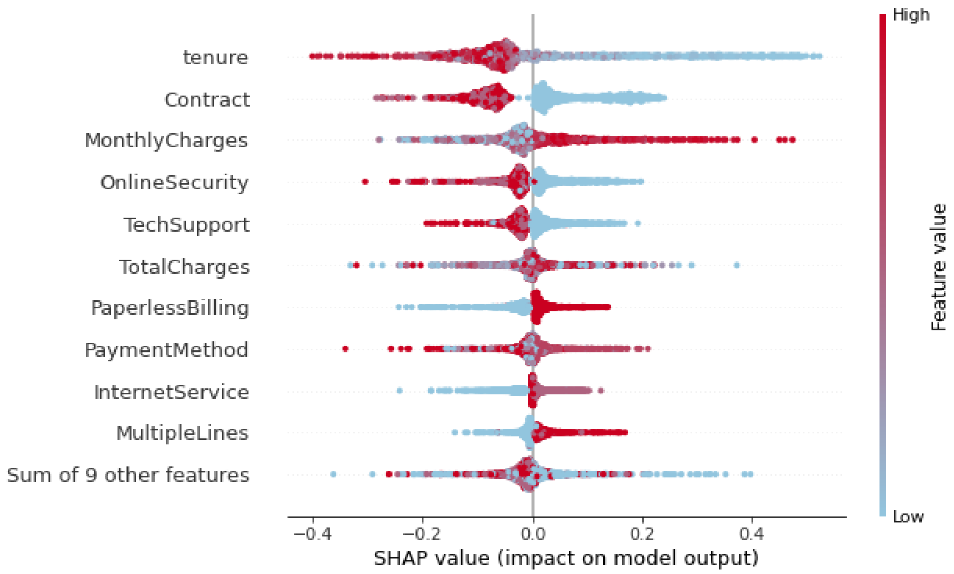

Figure 4. SHAP bee swarm plot illustrating the distribution of SHAP values for each feature in the LightGBM model.

+0.08, suggesting that the type of contract (e.g., month-to-month or long-term) significantly affects customer retention.

'MonthlyCharges' also plays a crucial role, with a mean SHAP value of +0.06, indicating that higher monthly charges may increase churn risk. Other notable features include 'OnlineSecurity' (+0.04), 'TechSupport' (+0.03), 'TotalCharges' (+0.03), 'PaperlessBilling' (+0.03), 'PaymentMethod' (+0.02), 'InternetService' (+0.02), and 'MultipleLines' (+0.02). The sum of nine other features contributes an additional +0.07 to the model's predictive power. These insights provide a comprehensive understanding of the key factors driving customer churn, enabling telecom companies to develop targeted strategies to enhance customer retention.

Figure 4 shows the SHAP bee swarm plot for the LightGBM model, which visualizes the impact of individual feature values on the model's predictions. While the bar plot provides a summary of the average importance of each feature, the bee swarm plot visualizes the distribution and specific influence of individual feature values on the model's predictions, with the color indicating the feature value (red for high values and blue for low values).

The feature 'tenure' shows a clear trend where high values (longer tenure) have negative SHAP values,

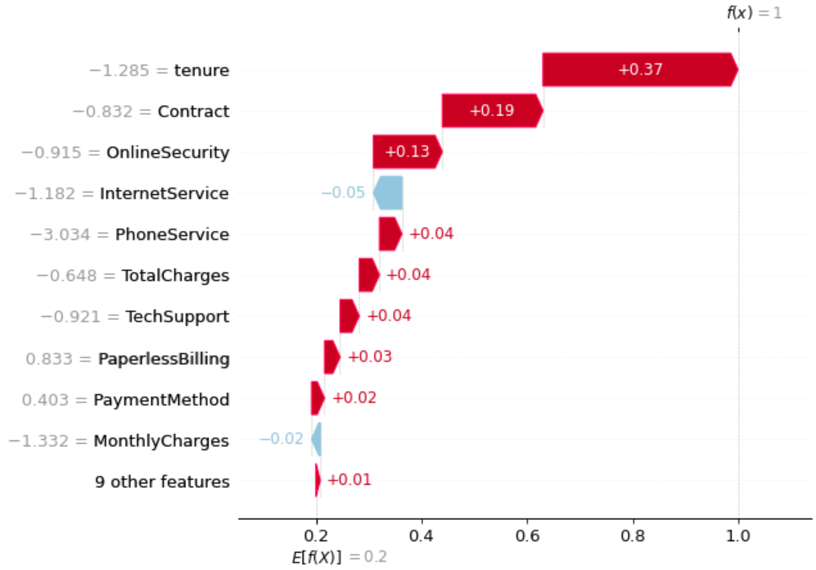

Figure 5. SHAP waterfall plot for an individual customer predicted to churn.

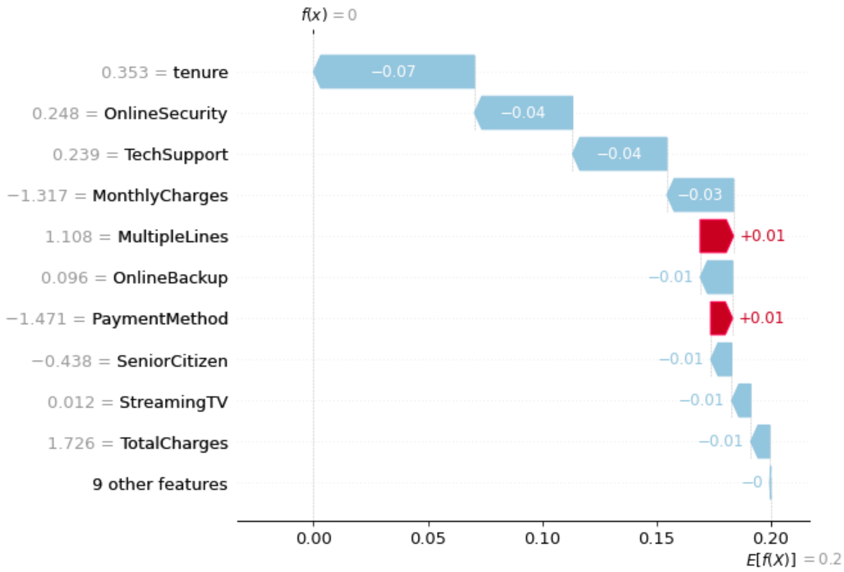

Figure 6. SHAP waterfall plot for an individual customer predicted not to churn.

reducing the likelihood of churn. This confirms that longer-term customers are less likely to leave the telecom company. The 'Contract' type also has a significant impact, with longer-term contracts associated with lower churn risk, as indicated by the concentration of blue points on the left side (negative SHAP values).

For 'MonthlyCharges', higher values (red points) generally lead to higher SHAP values, suggesting that increased charges are associated with a higher likelihood of churn. The features 'OnlineSecurity' and 'TechSupport' exhibit similar patterns. The presence of these services (red points) correlates with lower churn risk, indicated by negative SHAP values.

'TotalCharges' displays a more complex pattern, reflecting its cumulative nature over time, while features like 'PaperlessBilling', 'PaymentMethod', 'Internet Service', and 'MultipleLines' show varied impacts on churn based on their specific values. The plot highlights that the combination of these features contributes significantly to the model's predictive power, offering telecom companies valuable insights into the factors influencing customer churn and enabling them to refine their retention strategies effectively.

An individual example of SHAP values for a predicted churn outcome of "Yes" is illustrated in Figure 5. This figure provides a detailed breakdown of how specific feature values contribute to the final prediction. The plot shows that 'tenure' has the most substantial negative impact (-1.285), indicating that a shorter tenure significantly increases the likelihood of churn.

Conversely, features such as 'Contract' (-0.832), 'OnlineSecurity' (-0.915), and 'InternetService' (-1.182) also contribute negatively, reducing the likelihood of churn when present. On the positive side, features like 'PhoneService' (+0.04), 'TotalCharges' (+0.04), 'TechSupport' (+0.04), and 'PaperlessBilling' (+0.03) have smaller but additive effects that increase the probability of churn.

The individual contributions add up to shift the base value ($E[f(x)]$ = 0.2) towards the final model output ($f(x)$ = 1), indicating a strong prediction of churn. This visualization helps understand the specific factors influencing the prediction for this customer, highlighting the combined effects of various features and providing actionable insights for targeted interventions.

A specific example of SHAP values for a predicted churn outcome of "No" is presented in Figure 6. This figure demonstrates how various feature values contribute to the final customer retention prediction. The plot highlights that 'tenure' has a significant negative impact (-0.07), suggesting that a longer tenure reduces the likelihood of churn. Other features contributing to a lower churn probability include 'OnlineSecurity' (-0.04), 'TechSupport' (-0.04), 'MonthlyCharges' (-0.03), and 'TotalCharges' (-0.01). These features collectively decrease the churn risk, indicating that customers with these characteristics are likelier to stay with the company.

On the positive side, 'MultipleLines' (+0.01) and 'PaymentMethod' (+0.01) have minor positive contributions, slightly increasing the churn probability but not enough to outweigh the negative impacts of the other features. The combined effects of these features shift the base value ($E[f(x)]$ = 0.2) towards the final model output ($f(x)$ = 0), indicating a strong prediction of customer retention. This detailed visualization helps understand the specific factors that drive the prediction for this customer, providing valuable insights for enhancing retention strategies.

The findings of this study hold significant implications for the telecommunications industry. Telecom companies can use advanced machine learning models to better predict customer churn. Applying SHAP for model interpretability provides a transparent view of the factors influencing churn, allowing businesses to gain deeper insights into customer behavior. This interpretability is crucial for developing targeted retention strategies, as it helps identify the key drivers of churn, such as tenure, contract type, and monthly charges. These insights enable telecom companies to tailor their services and interventions to address their customers' specific needs and preferences, thereby improving customer satisfaction and loyalty. Additionally, using interpretable models fosters trust among stakeholders, making integrating these predictive tools into decision-making processes easier.

Despite the promising results, this study has several limitations. First, the dataset used is based on a fictional telecommunications company, which may not capture real-world telecom environments' full complexity and diversity. This could limit the generalizability of the findings to other contexts or industries. Second, while SHAP provides valuable insights into feature importance, it does not account for potential interactions between features that might influence churn. Future studies could benefit from exploring more sophisticated interpretability techniques that can capture these interactions. Furthermore, the study focuses on a limited number of machine learning models, and there may be other emerging models or techniques that could offer improved performance or interpretability. Finally, the study does not consider the temporal dynamics of customer behavior, which could be critical for understanding how churn predictors evolve.

Future research should further address the limitations this study identified to enhance the understanding and prediction of customer churn. Real-world datasets from multiple telecommunications companies could be used to validate and extend the findings, ensuring that the models are robust and applicable in various contexts. Investigating the interactions between features and their impact on churn predictions would provide a more comprehensive understanding of the underlying mechanisms driving customer behavior. Additionally, incorporating time-series analysis and recurrent neural networks could help capture the temporal aspects of churn, offering more dynamic and timely predictions. Lastly, extending the research to other industries with similar churn challenges, such as banking or retail, could help generalize the findings and develop industry-specific retention strategies.

4. Conclusions

This study demonstrates the effectiveness of using advanced machine learning models, particularly LightGBM, for predicting customer churn in the telecommunications industry, achieving an accuracy of 80.70%, precision of 84.35%, recall of 90.54%, and an F1-score of 87.34%. By using SHAP, we enhanced model interpretability, providing actionable insights into key factors influencing churn, such as tenure, contract type, and monthly charges. Despite limitations like using a fictional dataset and potential unaccounted feature interactions, the study highlights the potential of combining predictive analytics with interpretability methods to develop targeted retention strategies. Future research should validate these findings with real-world data, explore more sophisticated models, and incorporate temporal dynamics to refine churn predictions further.

Author Contributions: Conceptualization, T.R.N., G.M.I. and I.H.; methodology, T.R.N. and M.A.; software, T.R.N. and S.R.; validation, I.H. and M.H.; formal analysis, T.R.N. and S.R.; investigation, T.R.N.; resources, G.M.I.; data curation, I.H. and S.R.; writing—original draft preparation, T.R.N., I.H. and M.A.; writing—review and editing, T.R.N., G.M.I. and S.R.; visualization, T.R.N. and G.M.I.; supervision, T.R.N.; project administration, T.R.N.; funding acquisition, T.R.N. All authors have read and agreed to the published version of the manuscript.

Funding: This study does not receive external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study was retrieved from https://doi.org/10.34740/KAGGLE/DSV/8360350.

Conflicts of Interest: All the authors declare no conflicts of interest.

References

1. Zdravevski, E., Lameski, P., Apanowicz, C., and Ślęzak, D. (2020). From Big Data to Business Analytics: The Case Study of Churn Prediction, *Applied Soft Computing*, Vol. 90, 106164. doi:10.1016/j.asoc.2020.106164.
2. Tianyuan, Z., and Moro, S. (2021). Research Trends in Customer Churn Prediction: A Data Mining Approach, 227–237. doi:10.1007/978-3-030-72657-7_22.
3. Lemmens, A., and Gupta, S. (2020). Managing Churn to Maximize profits, *Marketing Science*, Vol. 39, No. 5, 956–973.
4. De, S., and Prabu, P. (2022). Predicting Customer Churn: A Systematic Literature Review, *Journal of Discrete Mathematical Sciences and Cryptography*, Vol. 25, No. 7, 1965–1985. doi:10.1080/09720529.2022.2133238.
5. Idroes, G. M., Hardi, I., Hilal, I. S., Utami, R. T., Noviandy, T. R., and Idroes, R. (2024). Economic Growth and Environmental Impact: Assessing the Role of Geothermal Energy in Developing and Developed Countries, *Innovation and Green Development*, Vol. 3, No. 3, 100144. doi:10.1016/j.igd.2024.100144.
6. Idroes, G. M., Hardi, I., Rahman, M. H., Afjal, M., Noviandy, T. R., and Idroes, R. (2024). The Dynamic Impact of Non-renewable and Renewable Energy on Carbon Dioxide Emissions and Ecological Footprint in Indonesia, *Carbon Research*, Vol. 3, No. 1, 35. doi:10.1007/s44246-024-00117-0.
7. Óskarsdóttir, M., Bravo, C., Verbeke, W., Sarraute, C., Baesens, B., and Vanthienen, J. (2017). Social Network Analytics for Churn Prediction in Telco: Model Building, Evaluation and Network Architecture, *Expert Systems with Applications*, Vol. 85, 204–220. doi:10.1016/j.eswa.2017.05.028.
8. Santouridis, I., and Trivellas, P. (2010). Investigating the Impact of Service Quality and Customer Satisfaction on Customer Loyalty in Mobile Telephony in Greece, *The TQM Journal*, Vol. 22, No. 3, 330–343. doi:10.1108/17542731011035550.
9. Noviandy, T. R., Maulana, A., Idroes, G. M., Maulydia, N. B., Patwekar, M., Suhendra, R., and Idroes, R. (2023). Integrating Genetic Algorithm and LightGBM for QSAR Modeling of Acetylcholinesterase Inhibitors in Alzheimer's Disease Drug Discovery, *Malacca Pharmaceutics*, Vol. 1, No. 2, 48–54. doi:10.60084/mp.v1i2.60.
10. Sasmita, N. R., Ramadeska, S., Kesuma, Z. M., Noviandy, T. R., Maulana, A., Khairul, M., and Suhendra, R. (2024). Decision Tree versus k-NN: A Performance Comparison for Air Quality Classification in Indonesia, *Infolitika Journal of Data Science*, Vol. 2, No. 1, 9–16. doi:10.60084/ijds.v2i1.179.
11. Noviandy, T. R., Nisa, K., Idroes, G. M., Hardi, I., and Sasmita, N. R. (2024). Classifying Beta-Secretase 1 Inhibitor Activity for Alzheimer's Drug Discovery with LightGBM, *Journal of Computing Theories and Applications*, Vol. 2, No. 2, 138–147. doi:10.62411/jcta.10129.
12. Matuszelański, K., and Kopczewska, K. (2022). Customer Churn in Retail E-Commerce Business: Spatial and Machine Learning Approach, *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 17, No. 1, 165–198. doi:10.3390/jtaer17010009.
13. Bhuse, P., Gandhi, A., Meswani, P., Muni, R., and Katre, N. (2020). Machine Learning Based Telecom-Customer Churn Prediction, *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, IEEE, 1297–1301. doi:10.1109/ICISS49785.2020.9315951.
14. Zhang, T., Moro, S., and Ramos, R. F. (2022). A Data-Driven Approach to Improve Customer Churn Prediction Based on Telecom Customer Segmentation, *Future Internet*, Vol. 14, No. 3, 94. doi:10.3390/fi14030094.
15. Liu, Y., Fan, J., Zhang, J., Yin, X., and Song, Z. (2023). Research on Telecom Customer Churn Prediction Based on Ensemble Learning, *Journal of Intelligent Information Systems*, Vol. 60, No. 3, 759–775. doi:10.1007/s10844-022-00739-z.
16. Noviandy, T. R., Idroes, G. M., and Hardi, I. (2024). Enhancing Loan Approval Decision-Making: An Interpretable Machine Learning Approach Using LightGBM for Digital Economy Development, *Malaysian Journal of Computing (MJOC)*, Vol. 9, No. 1, 1734–1745. doi:10.24191/mjoc.v9i1.25691.
17. Belle, V., and Papantonis, I. (2021). Principles and Practice of Explainable Machine Learning, *Frontiers in Big Data*, Vol. 4. doi:10.3389/fdata.2021.688969.
18. Noviandy, T. R., Maulana, A., Zulfikar, T., Rusyana, A., Enitan, S. S., and Idroes, R. (2024). Explainable Artificial Intelligence in Medical Imaging: A Case Study on Enhancing Lung Cancer Detection through CT Images, *Indonesian Journal of Case Reports*, Vol. 2, No. 1, 6–14. doi:10.60084/ijcr.v2i1.150.
19. Lundberg, S. M., and Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions, *Advances in Neural Information Processing Systems*, Vol. 30.
20. Le, T.-T.-H., Kim, H., Kang, H., and Kim, H. (2022). Classification and Explanation for Intrusion Detection System Based on Ensemble Trees and SHAP Method, *Sensors*, Vol. 22, No. 3, 1154. doi:10.3390/s22031154.
21. Barr Kumarakulasinghe, N., Blomberg, T., Liu, J., Saraiva Leao, A., and Papapetrou, P. (2020). Evaluating Local Interpretable Model-Agnostic Explanations on Clinical Machine Learning Classification Models, *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, 7–12. doi:10.1109/CBMS49503.2020.00009.
22. Moscato, V., Picariello, A., and Sperlí, G. (2021). A Benchmark of Machine Learning Approaches for Credit Score Prediction, *Expert Systems with Applications*, Vol. 165, 113986. doi:10.1016/j.eswa.2020.113986.
23. IBM Team. (2024). Telco Customer Churn (11.1.3+), Kaggle. doi:10.34740/KAGGLE/DSV/8360350.
24. Idroes, G. M., Noviandy, T. R., Maulana, A., Zahriah, Z., Suhendrayatna, S., Suhartono, E., Khairan, K., Kusumo, F., Helwani, Z., and Abd Rahman, S. (2023). Urban Air Quality Classification Using Machine Learning Approach to Enhance

Environmental Monitoring, *Leuser Journal of Environmental Studies*, Vol. 1, No. 2, 62–68. doi:10.60084/ljes.v1i2.99.

25. Wickramasinghe, I., and Kalutarage, H. (2021). Naive Bayes: Applications, Variations and Vulnerabilities: A Review of Literature with Code Snippets for Implementation, *Soft Computing*, Vol. 25, No. 3, 2277–2293. doi:10.1007/s00500-020-05297-6.

26. Boulesteix, A., Janitza, S., Kruppa, J., and König, I. R. (2012). Overview of Random Forest Methodology and Practical Guidance with Emphasis on Computational Biology and Bioinformatics, *WIREs Data Mining and Knowledge Discovery*, Vol. 2, No. 6, 493–507. doi:10.1002/widm.1072.

27. Chuttur, M. Y., and Bissonath, R. (2022). A Comparison of AdaBoost and SVC for Fake Hotel Reviews Detection, *2022 3rd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, IEEE, 1–6. doi:10.1109/ICCAKM54721.2022.9990075.

28. Wang, K., Li, M., Cheng, J., Zhou, X., and Li, G. (2022). Research on Personal Credit Risk Evaluation Based on XGBoost, *Procedia Computer Science*, Vol. 199, 1128–1135. doi:10.1016/j.procs.2022.01.143.

29. Noviandy, T. R., Maulana, A., Emran, T. B., Idroes, G. M., and Idroes, R. (2023). QSAR Classification of Beta-Secretase 1 Inhibitor Activity in Alzheimer's Disease Using Ensemble Machine Learning Algorithms, *Heca Journal of Applied Sciences*, Vol. 1, No. 1, 1–7. doi:10.60084/hjas.v1i1.12.

30. Noviandy, T. R., Nainggolan, S. I., Raihan, R., Firmansyah, I., and Idroes, R. (2023). Maternal Health Risk Detection Using Light Gradient Boosting Machine Approach, *Infolitika Journal of Data Science*, Vol. 1, No. 2, 48–55. doi:10.60084/ijds.v1i2.123.

31. Sevgen, E., and Abdikan, S. (2023). Classification of Large-Scale Mobile Laser Scanning Data in Urban Area with LightGBM, *Remote Sensing*, Vol. 15, No. 15, 3787. doi:10.3390/rs15153787.

32. Noviandy, T. R., Zahriah, Z., Yandri, E., Jalil, Z., Yusuf, M., Yusof, N. I. S. M., Lala, A., and Idroes, R. (2024). Machine Learning for Early Detection of Dropout Risks and Academic Excellence: A Stacked Classifier Approach, *Journal of Educational Management and Learning*, Vol. 2, No. 1, 28–34. doi:10.60084/jeml.v2i1.191.

33. Suhendra, R., Suryadi, S., Husdayanti, N., Maulana, A., Noviandy, T. R., Sasmita, N. R., Subianto, M., Earlia, N., Niode, N. J., and Idroes, R. (2023). Evaluation of Gradient Boosted Classifier in Atopic Dermatitis Severity Score Classification, *Heca Journal of Applied Sciences*, Vol. 1, No. 2, 54–61. doi:10.60084/hjas.v1i2.85.

34. Noviandy, T. R., Maulana, A., Idroes, G. M., Irvanizam, I., Subianto, M., and Idroes, R. (2023). QSAR-Based Stacked Ensemble Classifier for Hepatitis C NS5B Inhibitor Prediction, *2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE)*, IEEE, 220–225. doi:10.1109/COSITE60233.2023.10250039.

35. Klingspohn, W., Mathea, M., ter Laak, A., Heinrich, N., and Baumann, K. (2017). Efficiency of different measures for defining the applicability domain of classification models, *Journal of Cheminformatics*, Vol. 9, No. 1, 44. doi:10.1186/s13321-017-0230-2.

36. Berrar, D., and Flach, P. (2012). Caveats and Pitfalls of ROC Analysis in Clinical Microarray Research (and How to Avoid Them), *Briefings in Bioinformatics*, Vol. 13, No. 1, 83–97. doi:10.1093/bib/bbr008.