Machine Learning Explainability for Churn Prediction and Customer Retention Strategies: A Comparative Analysis

Anil Kumar Singh
Professor
Jagran Institute of Management, Kanpur, UP, India

ABSTRACT
Telecommunications companies face significant challenges when it comes to customer churn as loss of their clients negatively impacts revenue and business growth. This research introduces an explainable machine learning model for predicting customer churn and determining the most relevant variables related to customer churn. For the analysis of the Telco Customer Churn data set containing 7,043 observations, five models such as Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, and Extreme Gradient Boosting (XGBoost) were built. Accuracy, Precision, Recall, F1-score, ROC-AUC, PR-AUC, and five-fold cross-validation were used to assess model performance. Logistic Regression proved to be the best in terms of Accuracy (80.70%), Recall (56.15%), and F1-score (60.69%), whereas XGBoost showed the highest Precision (67.12%) and ROC-AUC (0.8425). Using SHapley Additive exPlanations (SHAP) as an interpretability technique, the variables affecting churn prediction, such as tenure, contract, Internet service, monthly charges, and total charges, were determined.

Keywords
Customer Churn; Machine Learning; Explainable AI; SHAP Analysis; Customer Retention.

1. INTRODUCTION
Customer retention has become an important business problem in the telecommunication industry since customers can change their service providers if they find better prices, services, contractual terms, or customer service somewhere else. By predicting the probability of a customer leaving the company before it happens, one can identify the customers at high risk and take targeted actions to retain them. Conventional approaches to analyzing churn involve descriptive statistics and manual segmentation of customers. While these approaches can detect some patterns, they do not necessarily reveal complicated dependencies between various demographic and psychographic factors, service utilization, contract, payment and billing details. One of the alternative approaches involves machine learning which learns patterns from data and makes individual predictions. However, good prediction accuracy is not always enough for decision making. The model can estimate that a specific customer is going to leave, but managers should also know the reasons behind the risk classification of a particular customer. This can be achieved with the help of Explainable Artificial Intelligence (XAI) that reveals the contribution of individual features into model's predictions. For example, SHAP (SHapley Additive exPlanations) approach can be used to analyze feature importance and individual predictions. In this research, the prediction of customer churn is examined through an approach to comparative machine learning with explainability. Demographic, service-related, contractual, and billing attributes will be used in this case to compare various classification models. Performance of the models is evaluated based on multiple classification metrics, whereas SHAP is applied to extract the major factors behind churn prediction.

2. OBJECTIVES OF THE RESEARCH
Prediction and explanation of customer churn together with development of a strategy for customer retention on their basis are the primary objectives of the research. In other words, this research unites the three processes that can be distinguished in churn analysis: prediction, explanation, and action.

3. LITERATURE REVIEW
Customer churn prediction has attracted significant attention as maintaining customers contributes to stability of revenue, customer lifetime value, and relationship management. Modern studies have shifted from classic classification to ensemble learning, class imbalance treatment, explainable artificial intelligence (XAI), and actionable strategies. Saleh and Saha [1] discussed churn prediction and customer retention in the field of telecommunications and stated the necessity for practical solutions to be provided by predictive models. Maan and Maan [2] used explainable machine learning techniques and showed that XGBoost is one of the best ways for churn prediction with SHAP for interpretation of feature importance. Chong et al. [3] evaluated different machine learning models for telecommunications customer churn prediction and showed that XGBoost has an accuracy of 79.67%, proving the importance of customer features for the development of personalized retention strategy. Zhou et al. [4] illustrated the potential of ensemble learning technique with an example of RF-AdaBoost model, which was able to demonstrate 93% recall, 99% precision, and 96% F1-score. Modern studies also involve different modeling techniques. Khattak et al. [5] explored a composite deep-learning model for customer churn prediction. Nevertheless, greater model complexity entails challenges in terms of explainability and implementation. To this end, Peng et al. [6] proposed to combine a more advanced XGBoost model with SHAP explainability and showed how it could help to enhance interpretation of predictions about churn customers. Additionally, Sikri et al. [7] emphasized the need to consider class imbalance in evaluation of churn prediction models and used other metrics besides accuracy. In 2024, the theme of explainability was gaining ground in telecom churn literature. Comparative works on tabular machine learning models introduced SHAP global and local explainability to understand the reasons behind churn. Sinaga and Widianto [9] used LightGBM along with SHAP and SMOTE to prove that explainability on the level of features could enhance prediction results. Other comparative works [10] included Random Forest, SVM, Gradient Boosting, XGBoost, and LightGBM models, stressing the significance of comparing different algorithms instead of using just one. Works published in 2025 further considered both prediction and explainability of models and their business application. A TriBoost architecture based on XAI was applied using the soft voting technique, with XGBoost,

CatBoost, and LightGBM, which showed an accuracy of 96.44%, precision of 92.82%, recall of 87.82%, and an F1-score of about 90.25%. The interpretations of the predictions using the SHAP and LIME techniques were also done. Abdelhady and Mohamed [12] developed an AI-based churn framework that was integrated with customer relationship management and stressed the importance of using the output of the model for engagement with customers. The more recent studies of 2026 have enhanced the emphasis on evaluation and explanation in decision-making processes. Zerine et al. [13] used nested cross-validation to compare five different models and evaluated the accuracy, precision, recall, F1-score, ROC-AUC, and PR-AUC, and then analyzed SHAP, partial dependence, and individual conditional expectation plots. El Attar and El-Hajj [14] created an explainable multi-model framework using the Telco Customer Churn dataset, which had 7,043 customers. The framework included feature engineering, SMOTE, machine-learning models, SHAP, and customer segmentation. The most common machine learning methods employed for structured churn datasets include XGBoost, Random Forest, LightGBM, and Gradient Boosting. Meanwhile, SHAP became the go-to technique for analyzing the model's decision process. However, more research is still needed that will address a combination of model analysis, feature-level explanations, and practical customer-retention techniques within a single approach. The current study fulfills this requirement through the creation and evaluation of machine learning models, the interpretation of churn prediction using SHAP, and the formulation of customer-retention strategies based on the findings.

4. MATERIAL AND METHODS

4.1 Dataset Description

In total, the dataset has 7,043 observations and 21 features. These features represent customer demographics, services, account information, billing, and the status of their churn from the service. The variable Churn is the target, representing whether a particular customer left the service. Categorical and numerical features are represented in the dataset. Although TotalCharges is a numerical amount of billing, it is provided in a text format and was converted to the numeric format in preprocessing. The CustomerID was dropped since it is merely an identifier. All categorical variables were one-hot encoded, whereas numerical variables (tenure, MonthlyCharges, TotalCharges) were used as features.

4.2 Target Variable - Churn Distribution

In this dataset, there are 5,174 non-churned customers, which accounts for 73.46% and 1,869 churned customers which account for 26.54% of customers. It can be considered a slightly unbalanced class with non-churned customers making up almost three-quarters of customers in the dataset.

4.3 Descriptive Statistics

According to Table 1 the average customer has a tenure of 32.37 months, paying an average monthly charge of 64.76 and total charges of 2,279.73 per month. There is a great variation in total charges due to the different amounts of spending by customers. The SeniorCitizen variable has a mean of 0.16, meaning that about 16% of customers are senior citizens

Table 1: Descriptive Statistics of Numerical Variables

Variables	Senior Citizen	Tenure	Monthly Charges	Total Charges
Number of Observations	7,043	7,043	7,043	7,043
Mean	0.16	32.37	64.76	2,279.73
Standard Deviation	0.37	24.56	30.09	2,266.79
Minimum	0.00	0.00	18.25	0.00
25th Percentile	0.00	9.00	35.50	398.55
Median	0.00	29.00	70.35	1,394.55
75th Percentile	0.00	55.00	89.85	3,786.60
Maximum	1.00	72.00	118.75	8,684.80

4.4 Correlation Matrix

Figure 1 presents the correlation analysis, it is clear that the factor with the highest correlation is tenure with r = −0.35, meaning that the higher the churn rate is for customers with lower tenure. Monthly Charges show weak positive correlation (r = 0.19) whereas Total Charges have a weak negative correlation (r = -0.20). Tenure and Total Charges have high correlation (r = 0.83).

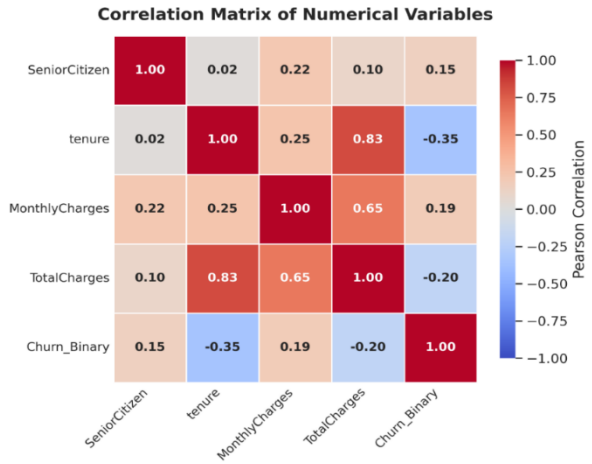

Figure 1: Correlation Matrix

4.5 Preprocessing and Features Overview

Post-preprocessing, the dataset consisted of 7,043 records, containing 19 input features, and 7,043 target values. Among those features, four features SeniorCitizen, tenure, MonthlyCharges, and TotalCharges were considered numerical features. The other 15 features represented categorical data, capturing customer demographics, offered services, contract and payment information. Thus, a good combination of numerical and categorical features served to represent customer features.

5. TRAINING AND TESTING

The data split was performed in a way that the dataset was divided into 5,634 training and 1,409 testing records, both having 19 predictive features. The class ratio is almost identical for both samples: 73.46% non-churn and 26.54% churn. Thus, a representative split while keeping the class imbalance was achieved. Model performance was assessed using accuracy, precision, recall, F1-score, ROC-AUC, and PR-AUC.

6. MODEL PERFORMANCE ANALYSIS

Table 2: Model Performance Comparison

Models	Accuracy	Precision	Recall	F1-Score	ROC-AUC
XGBoost	0.8055	0.6712	0.5241	0.5886	0.8425
Logistic Regression	0.8070	0.6604	0.5615	0.6069	0.8422
Gradient Boosting	0.7977	0.6540	0.5053	0.5701	0.8415
Random Forest	0.7949	0.6412	0.5160	0.5719	0.8269
Decision Tree	0.8006	0.6667	0.4973	0.5697	0.8219

As seen in Table 2, all five models performed almost equally. Logistic Regression performed best in terms of accuracy (80.70%), recall (56.15%), and F1-score (60.69%), while XGBoost performed best in terms of precision (67.12%) and ROC-AUC (0.8425). The performance of Gradient Boosting, Random Forest, and Decision Tree was marginally lower than other two models on most performance metrics.

6.1 Confusion matrix

Figure 2 shows the XGBoost confusion matrix, 939 customers who did not churn were classified as such by the model and 196 churned customers were accurately classified. However, 96 customers who were non-churning were misclassified as churn and 178 customers who were churning were misclassified as non-churning.

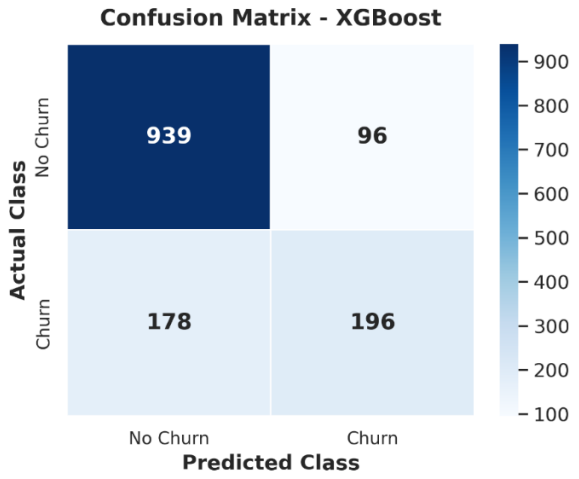

Figure 2: Confusion Matrix

6.2 ROC Curves for Customer Churn Prediction Models

It can be seen from Figure 3 that the discriminatory ability of all five models is quite high, with AUC ranging from 0.822 to 0.843. The best AUC was recorded by XGBoost, which had AUC of 0.843. Next to XGBoost were Logistic Regression and Gradient Boosting (AUC = 0.842).

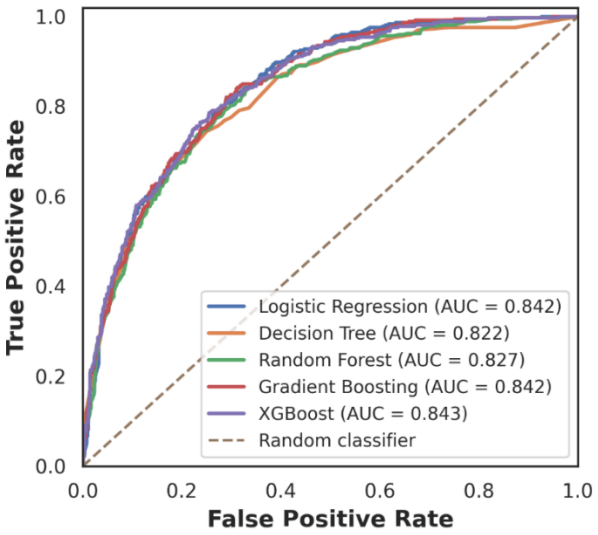

Figure 3: ROC Curves for Models of Customer Churn Prediction

6.3 Precision-Recall Curves for Customer Churn Prediction Models

Figure 4 presents the Precision-Recall values of the models are fair to moderate. The XGBoost model has the best average precision (AP) value of 0.658, followed by Gradient Boosting and Logistic Regression with 0.652 and 0.635, respectively. The Decision Tree model has the worst AP of 0.601.

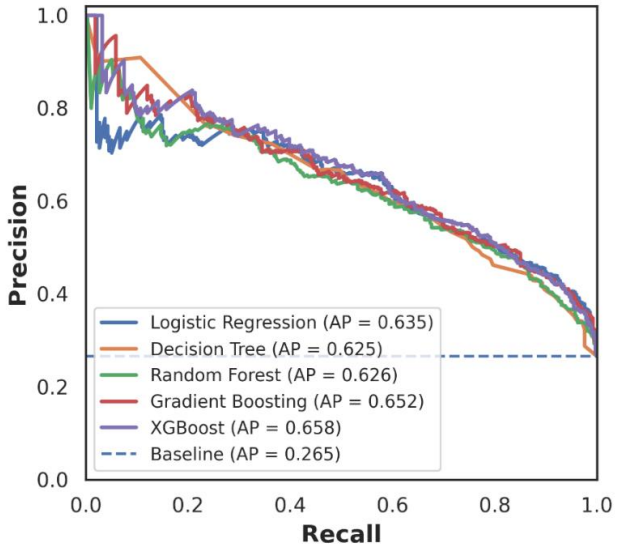

Figure 4: Precision-Recall Graphs of Churn Prediction

7. 5 Folds Cross Validation

According to Table 4, the Gradient Boosting and Logistic Regression models were consistent in their performance. The Gradient Boosting model had the highest average ROC-AUC and precision scores, whereas the Logistic Regression had the higher accuracy, recall and F1-Score. The small standard deviation values imply consistency across all five folds.

Table 4. Results of Five-Fold Cross Validation of Machine Learning Models

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC
Gradient Boosting	0.8051 ± 0.0050	0.6680 ± 0.0134	0.5281 ± 0.0174	0.5897 ± 0.0126	0.8471 ± 0.0111
Logistic Regression	0.8056 ± 0.0109	0.6592 ± 0.0216	0.5527 ± 0.0313	0.6011 ± 0.0272	0.8451 ± 0.0134
XGBoost	0.7994 ± 0.0088	0.6507 ± 0.0189	0.5265 ± 0.0327	0.5816 ± 0.0249	0.8424 ± 0.0097
Decision Tree	0.7879 ± 0.0076	0.6258 ± 0.0177	0.5050 ± 0.0810	0.5548 ± 0.0452	0.8226 ± 0.0092
Random Forest	0.7873 ± 0.0074	0.6271 ± 0.0130	0.4879 ± 0.0302	0.5486 ± 0.0240	0.8248 ± 0.0122

Table 5 indicates that the five-fold validation test, it is evident that Gradient Boosting had the highest average ROC-AUC. All the three best models, Gradient Boosting, Logistic Regression, and XGBoost, exhibited similar stable discrimination results.

Table 5. Results of Five Fold Cross Validation for ROC-AUC

Model	F 1	F 2	F 3	F 4	F 5	Avg R-A	Std.
GB	0.8609	0.8439	0.8588	0.8321	0.8396	0.8471	0.0124
LR	0.8546	0.8454	0.8636	0.8253	0.8364	0.8451	0.0150
XGBoost	0.8548	0.8395	0.8516	0.8279	0.8382	0.8424	0.0109
RF	0.8373	0.8226	0.8388	0.8056	0.8197	0.8248	0.0137
DT	0.8360	0.8207	0.8288	0.8091	0.8183	0.8226	0.0103

F=Fold, R-A=ROC-AUC, GB=Gradient Boosting, LR=Logistic Regression, RF=Random Forest, DT= Decision Tree

8. SHAP Analysis

Figure 5 indicates that tenure is the most significant variable, followed by the contract type, the Internet service type, monthly charges, and total charges. Long tenure is a factor that leads to low churn probability, while short tenure is associated with high churn probability. Moreover, two-year contracts lead to low churn probability, but fiber optic service and high monthly charges are associated with increased churn probability.

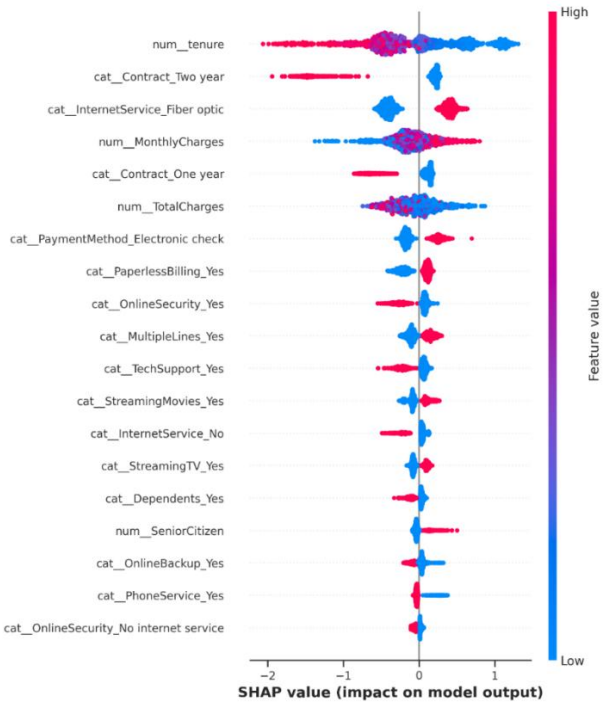

Figure 5: SHAP Analysis

8.1 SHAP Waterfall Plot

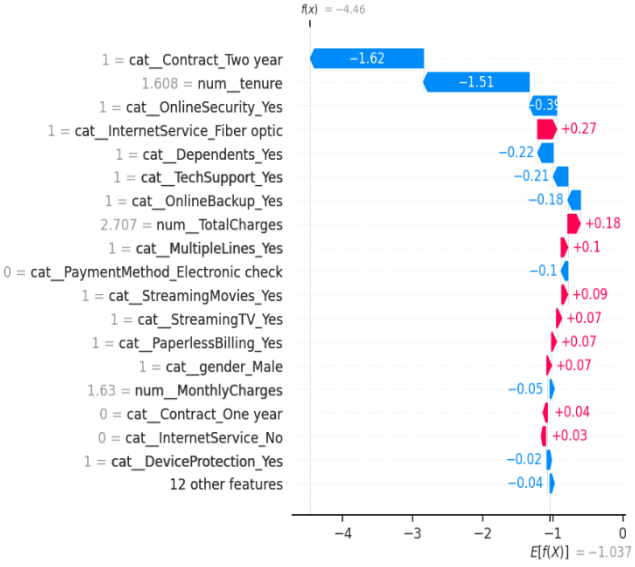

Figure 6: SHAP Waterfall Graph

Figure 6 reveals that contract type and tenure have the biggest impact on the prediction for the current observation. Contract period of two years (-1.62) and tenure (-1.51) decrease the probability of churn, whereas fiber internet service (0.27) and total charges (0.18) increase the probability. Other features have less impact on the result.

9. RETENTION STRATEGIES BASED ON CHURN

Table 6 show that retention campaigns should be designed for customers having high churn rates. Below is a table showing the results and retention strategies.

Table 6. Retention Strategies Based on Churn

Customer Segment	Customer Churn Rate	Recommendation for Retaining
Monthly contracts	42.71% churn	Promote long-term contracts through discounts, bundled deals, or extra services
Fiber optic customer	41.89% churn	Ensure reliability and offer proactive customer support.
Electronic-check payers	45.29% churn	Promote easy payment options like direct bank transactions or credit card payments.
Without technical support	41.64% churn	Offer proactive assistance and customer onboarding.
Seniors-citizens	41.68% churn	Create easy plans and proactively support seniors-citizens.
Short-term customers	Higher vulnerability	Focus on customer onboarding, engagement, and customer satisfaction

Generally, the results have proven that an intervention strategy that is both specific and based on data would be better than a generic one. The churn prediction model will provide the information needed in finding customers who are likely to leave so that specific measures such as contractual rewards, technical assistance, advice on payments, and special offers can be applied.

10. RESULTS AND DISCUSSION

The results of the analysis demonstrated evident dissimilarities in terms of churn behavior between the customer segments. Churn was higher for month-to-month contract type (42.71%) as opposed to one-year (11.27%) and two-year (2.83%) contracts. Customers of the fiber-optics plan exhibited higher churn rates (41.89%), similarly for electronic check payment (45.29%), those lacking technical support (41.64%), and senior citizens (41.68%). Relationship ($r = -0.35$) appeared to be the variable with the strongest association with the dependent variable. Five different machine-learning models were compared. Logistic regression exhibited the best accuracy (80.70%), recall (56.15%), and F1-score (60.69%). On the other hand, XGBoost demonstrated the best precision (67.12%) and ROC-AUC (0.8425). In terms of ROC, all three models demonstrated similar discrimination, while XGBoost demonstrated the best average precision (0.658) in the precision-recall analysis. The results of 5-fold cross-validation demonstrated stable model performance. Gradient boosting demonstrated the best ROC-AUC (0.8471), while logistic regression demonstrated slightly better recall (0.5527) and F1-score (0.6011). The relative low variation in folds proves the generation of similar results on different datasets by the models. It is easier to get insight into the model's predictions with the help of SHAP analysis. Tenure is the most influential feature followed by contract type, internet service, MonthlyCharges, and TotalCharges. The higher tenure and two-year contracts decreased the predicted churn probability, while the opposite features increased it. In addition, the individual SHAP explanation shows that contract type and tenure were the most influential in prediction. Thus, all results prove that the probability of churn correlates highly with the tenure and contract of a customer, as well as the service provided and payment features.

11. CONCLUSION

In this study, an explainable machine learning approach was proposed for customer churn prediction and the discovery of factors leading to customer attrition. Classification methods were tested on several evaluation criteria, then subject to 5-fold cross-validation and SHAP-based interpretation. As a result, it has been found that the Logistic Regression model produced the best accuracy, recall, and F1-score for the test dataset while XGBoost reached the highest precision and ROC-AUC value. Cross-validation has confirmed the same high performance for the leading models. Using the SHAP analysis, it has been discovered that the significant predictors of churn are tenure, Contract, InternetService, MonthlyCharges, and TotalCharges. Based on these findings, it can be concluded that customer retention will be more effective when potential churners are recognized in advance and the reasons for the predicted churn are known. Namely, customers with short tenure and month-to-month contract should be engaged more, while service-related issues can be fixed by providing technical assistance. Further studies on this topic can build upon the research presented herein through the use of longitudinal customer data, real-time behavioral data, cost-sensitive learning, and model monitoring.

12. REFERENCES

[1] S. Saleh and S. Saha, "Customer retention and churn prediction in the telecommunication industry: a case study on a Danish university," *Discover Applied Sciences*, vol. 5, article 173, 2023. doi: 10.1007/s42452-023-05389-6

[2] J. Maan and H. Maan, "Customer Churn Prediction Model using Explainable Machine Learning," 2023. arXiv:2303.00960.

[3] A. Y. W. Chong, K. W. Khaw, W. C. Yeong, and W. X. Chuah, "Customer Churn Prediction of Telecom Company Using Machine Learning Algorithms," *Journal of Soft Computing and Data Mining*, vol. 4, no. 2, pp. 1–22, 2023.

[4] Y. Zhou, W. Chen, X. Sun, and D. Yang, "Early warning of telecom enterprise customer churn based on ensemble learning," *PLOS ONE*, vol. 18, no. 10, e0292466, 2023. doi: 10.1371/journal.pone.0292466.

[5] A. Khattak, Z. Mehak, H. Ahmad, M. U. Asghar, M. Z. Asghar, et al., "Customer churn prediction using composite deep learning technique," *Scientific Reports*, vol. 13, article 17294, 2023.

[6] K. Peng, Y. Peng, and W. Li, "Research on customer churn prediction and model interpretability analysis," *PLOS ONE*, vol. 18, no. 12, e0289724, 2023. doi: 10.1371/journal.pone.0289724.

[7] A. Sikri, R. Jameel, et al., "Enhancing customer retention in telecom industry with machine learning driven churn prediction," *Scientific Reports*, vol. 14, article 13097, 2024. doi: 10.1038/s41598-024-63750-0.

[8] "Explaining customer churn prediction in telecom industry using tabular machine learning models," *Machine Learning with Applications*, vol. 17, article 100567, 2024. doi: 10.1016/j.mlwa.2024.100567.

[9] R. Sinaga and S. Widianto, "Understanding Telecommunication Customer Churn: Insights from LightGBM Predictive Modelling and SHAP Feature Interpretation," 2024. doi: 10.2139/ssrn.4758716.

[10] Raji Krishna, D. Jayanthi, D. S. Shylu Sam, K. Kavitha, Naveen Kumar Maurya, and T. Benil, "Application of machine learning techniques for churn prediction in the telecom business," *Results in Engineering*, vol. 24, article 103165, 2024. doi: 10.1016/j.rineng.2024.103165.

[11] [2]. Daniyal Asif, Muhammad Shoaib Arif, and Aiman Mukheimer, "A data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry," *Results in Engineering*, vol. 26, article 104629, 2025. doi: 10.1016/j.rineng.2025.104629.

[12] M. G. Abdelhady and K. A. Mohamed, "Leveraging artificial intelligence for predictive customer churn modeling in telecommunications: a framework for enhanced customer relationship management," *Scientific Reports*, vol. 15, article 43826, 2025.

[13] I. Zerine, M. M. Islam, et al. "Explainable churn prediction in telecom with tabular ML five model benchmark and SHAP analysis," *Discover Artificial Intelligence*, vol. 6, article 263, 2026. doi: 10.1007/s44163-026-00983-0.

[14] A. El Attar and M. El-Hajj, "Explainable AI-driven customer churn prediction: a multi-model ensemble approach with SHAP-based feature analysis," *Frontiers in Artificial Intelligence*, vol. 9, 2026. DOI: 10.3389/frai.2026.1748799.