Customer Churn Risk Analysis Using a TabNet and SHAP Model in Banking Sector

Ryansyah Fathi
Department of Data Science
Faculty of Informatics
Telkom University
Bandung, Indonesia
akufathin@student.telkomuniversity.ac.id

Dede Rohidin
Department of Infromatic
Faculty of Informatics
Telkom University
Bandung, Indonesia
dederohidin@telkomuniversity.ac.id

Jondri
Department of Data Scienc
Faculty of Informatics
Telkom University
Bandung, Indonesia
jondri@telkomuniversity.ac.id

Customer churn presents a significant challenge for the banking industry, impacting revenue and business sustainability. However, identifying at risk customers often relies on traditional methods that may lack accuracy or interpretability. This study develops a robust and interpretable churn prediction model using TabNet (Tabular Network) and SHAP (SHapley Additive exPlanations). The model was trained on a large-scale dataset of 115,640 banking customers with a significant class imbalance (12.2% churn) without employing synthetic sampling. The experimental results demonstrate that the proposed TabNet model achieved superior performance, yielding an Accuracy of 99,87% and an F1-score of 99,48% for the churn class, significantly outperforming baseline models, including MLP, DNN, and Simple RNN. Furthermore, SHAP analysis successfully identified 'Balance', 'NumComplaints', 'NumOfProducts', and 'Credit Score' as the most critical features influencing churn risk. This combined approach provides financial institutions with a highly accurate predictive tool and actionable insight to formulate targeted retention strategies.

Keywords: Customer Churn; TabNet; SHAP; Banking; Model Interpretability

1. INTRODUCTION

In the fast-paced and competitive digital era, customer loyalty has become one of the most valuable assets for companies, including financial institutions such as banks[1]. The ease of access to various digital banking services, coupled with the increasing options for customers to switch providers, presents a tangible challenge in customer retention. The phenomenon of customer churn, where customers stop using a service, poses a serious challenge [1][2]. Studies show that acquiring new customers can cost 5 to 6 times as much as retaining an existing one [3]. Therefore, accurately predicting and preventing churn has become a critical business objective [1].

Previous studies have demonstrated that the ability to predict churn is a significant competitive advantage accurately. Various machine learning approaches have been applied across different sectors. For instance, hybrid models combining K-Means and Support Vector Machines (SVM) have been utilized in e-commerce to segment customers before prediction [3]. In the telecommunications sector, which also faces high churn rates, complex deep learning models such as the FCLCNN-LSTM [4] and Bi-LSTM [5] have been developed to assess risk. Other studies in this sector have used GA-XGBoost [6], hybrid neural networks (e.g., CNN-BiLSTM) [7], or ensemble methods [1]. In the banking sector specifically, early research applied various machine learning models to telemarketing campaign data [8][9], while more recent research has focused on comparing traditional algorithms like Random Forest, XGBoost, and Logistic Regression [10].

However, while these models demonstrate strong predictive performance, their focus is often limited to accuracy. The aspect of interpretability and the ability to understand the reasons behind a prediction is frequently overlooked [6]. In a highly regulated sector like banking, transparency and the ability to explain predictive outcomes to stakeholders are critical. This presents a clear research gap and the need for an approach that simultaneously delivers robust churn prediction while also providing clear, human understandable explanations for its decisions, specifically on the tabular data dominant in the banking sector.

To address this challenge, this study proposes a combined model utilizing TabNet (Tabular Network) [11] and SHAP (Shapley Additive exPlanations) [12]. TabNet is a deep learning architecture designed explicitly for tabular data that employs sequential attention to select features automatically and has shown competitive performance [11]. Unlike traditional methods that often require extensive feature engineering or external sampling, TabNet can learn a framework that reduces reliance on labeled data [14]. Conversely, SHAP is a game theoretic unified approach to explain the output of any machine learning model [12]. It is increasingly used to interpret 'black-box' models by quantifying the contribution of each feature to a specific prediction [6] [15] [16].

Therefore, the primary objective of this study is to develop an accurate customer churn classification model tailored to banking tabular data using the TabNet algorithm. Furthermore, to ensure the results are transparent and understandable to stakeholders, the study implements the SHAP method to provide robust interpretations of the model's predictions. Through this approach, the research seeks to identify the key features that serve as the most significant drivers of customers churn [7]. By combining the predictive power of TabNet with the explanatory power of SHAP, this research contributes a reliable and data-driven solution to support wiser retention strategies in the banking industry.

2. METHODS

This research was conducted through a systematic series of stages, as illustrated in the research system design (Figure 1). The process begins with data collection and preprocessing, followed by model development, and concludes with model evaluation and interpretation.

2.1 TabNet (Tabular Network)

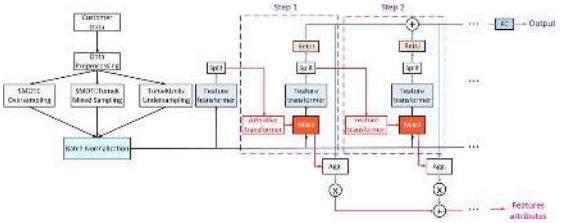


FIGURE 1
TabNet Architecture

2.2 SHAP (SHapley Additive exPlanations)

To interpret the predictions of the TabNet "black-box" model, this study employed SHAP (SHapley Additive exPlanations). SHAP is a game theoretic approach that provides a unified method for interpretable machine learning. It calculates the SHAP value for each feature, representing its contribution to pushing the model's output from a base value to the final prediction [6] [12].

This method allows for both global and local interpretability [16]. Global interpretability identifies which features are most important on average across the entire dataset, while local interpretability explains why a specific prediction was made [12] [15].

The dataset used in this research is the Bank Customer Churn dataset sourced from the Kaggle platform [18]. This dataset provides a comprehensive, synthetic collection of customer information from a fictional bank in Botswana, designed explicitly for churn analysis. It contains 115,640 unique customer records organized in a tabular format. The features encompass a wide range of attributes, including demographic details, financial information (e.g., Balance, Income), and product usage (e.g., NumOfProduct, NumComplaints).

2.3.1. Exploratory Data Analysis (EDA)

Initial exploratory data analysis (EDA) was conducted to understand the data's characteristics and distribution [10]. The primary challenge identified was a severe class imbalance. As illustrated in Figure 2, the dataset is highly skewed, with 87,8% of customers labeled as 'No Churn (0)' and only 12,2% labeled as 'Churn(1)'.

Confusion Matrix (Table 1) provides a detailed breakdown of correct and incorrect predictions by categorizing them into four quadrants, like first is True Positive (TP), representing churn customers correctly identified, second is False Negative (FN), representing churn customers missed by the model, third is False Positive (FP), representing non-churn customers incorrectly flagged as churn, and last is True Negative (TN), representing non-churn customer correctly identified [17].

2.3. Dataset and Preprocessing

The primary model employed in this study is TabNet (Tabular Network), a deep learning architecture explicitly designed for tabular data [11]. TabNet utilizes an attention mechanism to perform instance-wise feature selection, enabling it to focus on the most relevant features for each prediction [11] [13].

As detailed in the methodology, the TabNet architecture (shown in Figure 1) processes features through a sequential multi step framework. Each step consists of a Feature Transformer and an Attentive Transformer [11]. The Feature Transformer processes the feature representations, while the attentive Transformer calculates a mask using Sparsemax and processes in subsequent steps to produce the final prediction [11] [14].

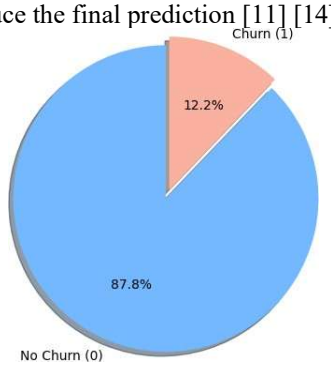


FIGURE 2
Customer Churn Distribution

Further analysis involved examining the relationship between features and the target variable. A correlation heatmap (Figure 3) was generated to assess multicollinearity between numerical features. Additionally, the correlation of each feature with the 'Churn Flag' was plotted (Figure 4).

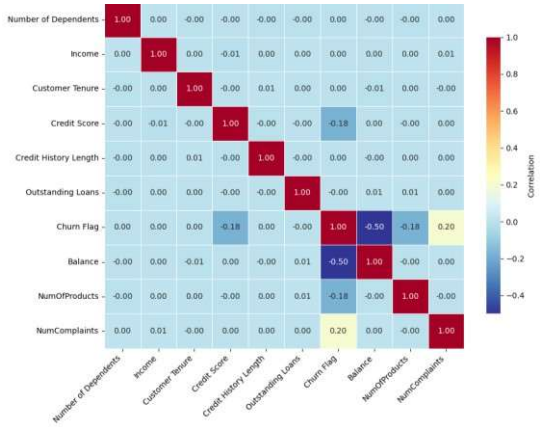


FIGURE 3
Correlation Matrix of Numerical Features

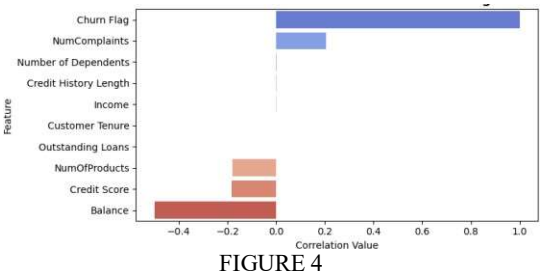


FIGURE 4
Correlation of Each Feature with Churn Flag

As shown in Figure 4, this initial analysis confirmed that specific attributes are correlated with the likelihood of churn. For example, 'NumComplaints' shows a clear positive correlation, while 'Balance' and 'NumOfProducts' show negative correlations. This validates the dataset's suitability for building a predictive model, as these features provide predictive signals [7] [10].

2.4 Model Evaluation

To evaluate the performance of the trained classification model, this study uses several standard metrics derived from the Confusion Matrix [17]. The

TABLE 1
Confusion Matrix

	Class 0	Class 1
Actual: Churn (1)	True Positive (TP)	False Negative (FN)
Actual: No Churn (0)	False Positive (FP)	True Negative (TN)

From this matrix, the primary evaluation metrics are calculated. Accuracy measures the ratio of all correct predictions (both TP and TN) against the total number of samples. However, accuracy can be misleading in imbalanced datasets, such as those common in churn prediction [17].

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

Precision measures the model's exactness in optimistic predictions, indicating how many customers predicted to churn actually churned [17]. This metric is crucial for minimizing false positives.

$$Precision = (TP) / (TP + FP)$$

Recall measures the model's completeness, indicating its ability to detect all intellectual churn cases [17]. This is critical for minimizing false negatives, as failing to identify a customer who is truly at risk is a significant business loss.

$$Recall = (TP) / (TP + FN)$$

Finally, the F1-Score is used as the primary evaluation metric. Provides the harmonic mean of precision and Recall, offering a single, balanced score that is particularly robust for imbalanced data. A high F1- score is only achieved when both precision and Recall are high [17]

$$F1\text{-}Score = (2 \times TP) / (2 \times TP + FP + FN)$$

3. RESULTS AND DISCUSSIONS

This section presents the experimental findings of the proposed model, followed by a comparative analysis against other deep learning baselines. Subsequently, the discussion section provides an in depth interpretation of these results and the key factors influencing churn as identified by the SHAP analysis.

3.1 Results

TABLE 2
. Performance of the Proposed TabNet Model

Metric	Class 0	Class 1	Weighted Average
Support	20309	2819	23128

The experiments were conducted on the preprocessed test set, which (based on the model outputs) consisted evaluated, and the results are presented below.
The proposed TabNet model achieved a final test accuracy of 99.87%. The detailed classification report is presented in Table 2, and the corresponding confusion matrix is shown in Figure 5.
To validate its effectiveness, the TabNet model was benchmarked against four other deep learning architectures: a Multi-Layer Perceptron, a standard DNN Baseline, a Simple RNN, and a 1D CNN. The comparative results, focusing on the critical churn class (Class 1), are summarized in Table 3.

TABLE 3
Comparative Performance of Model

Model	Accuray	Precision	Recall	F1-Score
TabNet	0.9987			
No Churn (0)		0.9988	0.9998	0.9993
Churn (1)		0.9986	0.9911	0.9948
MLP Classifier	0.9979			
No Churn (0)		0.9993	0.9884	0.9988
Churn (1)		0.9884	0.9947	0.9915
Simple RNN	0.9951			
No Churn (0)		0.9951	0.9974	0.9963
Churn (1)		0.9809	0.9649	0.9728
CNN 1D	0.9627			
No Churn (0)		0.9709	0.9871	0.9789
Churn (1)		0.8944	0.7868	0.8371
DNN	0.9966			
No Churn (0)		0.9966	0.9977	0.9971
Churn (1)		0.9835	0.9752	0.9793

3.2 Discussions

The results indicated that the proposed TabNet model achieved the highest performance among all tested architectures. As shown in Table 2, the model obtained an F1 score of 0.9948 for the minority churn class. This high score demonstrates an exceptional balance between Precision (0.9986) and Recall (0.9911). The confusion matrix reinforces this, showing only 25 False Negatives. This is a critical business outcome, as it means the model successfully identified 99,11% of all customers who were truly at risk of churning, without being artificially biased by oversampling. Furthermore, only 4 non-churning customers were incorrectly flagged (False Positive), demonstrating high reliability.

The comparative analysis in Table 3 solidifies this finding. The MLPClassifier also performed exceptionally well (F1-Score 0.9915), proving to be a strong baseline. However, the proposed TabNet model still yielded superior results. This suggests that TabNet's specialized architecture, which utilizes sequential attention and feature masking, provides a tangible advantage in modeling complex customer
behaviors and identifying predictive signals directly from imbalance tabular data.

To interpret why the model made these decisions, SHAP (SHapley Additive exPlanations) analysis was conducted. This analysis identifies the key drivers influencing the model's predictions. Figure 6 ranks the features by their mean absolute impact, while Figure 7 provides a detailed summary plot illustrating the direction and magnitude of each feature's effect.



FIGURE 6
SHAP Global feature Importance

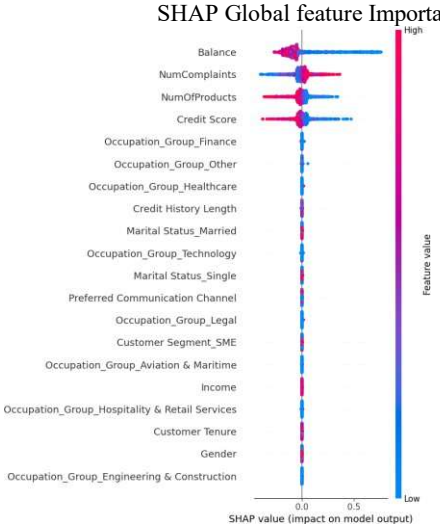


FIGURE 7
SHAP Summary Plot of Feature Impacts

The SHAP analysis reveals that the top four most significant factors in predicting customer churn are 'Balance', 'NumComplaint', 'NumOfProduct', and 'Credit Score'.

The interpretation of these features provides clear, actionable insight:

1. Balance : This is the most dominant feature. As shown in Figure 7, a low feature value (blue dots) corresponds to high positive SHAP value, strongly pushing the model's prediction toward churn(1). Conversely, customers with a high balance (red dots) have negative SHAP values, indicating they are significantly less likely to churn.
2. NumComplaints: A high number of complaints (red dots) clearly correlates with positive SHAP values, increasing the churn risk. Customers with few or no complaints (blue dots) have negative SHAP values, indicating higher loyalty.
3. NumOfProduct: This feature demonstrates that customer engagement is key. Having a low number of products (blue dots) increases churn probability (positive SHAP value), while customers with many products (red dots) are associated with negative SHAP values, reducing churn risk.
4. Credit Score: A lower score (blue dots) is shown to increase the likelihood of churn, while a higher credit score (red dots) helps to retain the customer.

These insights provide data-driven evidence for the bank. For instance, retention strategies can be proactively targeted at customers with consistently low balances, those who have registered complaints, or those who drop from multiple products to single product. The model's ability to not only predict churn with high accuracy but also to explain why provides a powerful tool for developing targeted inventions.

4. CONCLUSIONS

This study has successfully developed and validated an accurate and interpretable model for predicting customer churn in the banking sector. By implementing the TabNet architecture on a large scale (115,640 records) and a highly imbalanced (12,2 % minority class) dataset. The proposed model achieved superior performance, yielding an overall Accuracy of 99,87% and a critical F1-score of 0,9948 for the churn class, significantly outperforming baseline models including MLP, DNN, and SimpleRNN. Furthermore, integrating SHAP provides crucial model transparency, moving beyond "black-box" predictions. The SHAP analysis identified Balance as the most dominant predictor, followed by NumComplaints, NumOfProducts, and Credit Score. This analysis revealed actionable insight, such as low balances and a high number of complaints being strong indicators of churn. The primary contribution of this work is the demonstration of the TabNet+SHAP framework as a robust, end-to-end solution that provides both high precision predictions and the necessary interpretability for data driven retention strategies in the banking industry.

REFERENCES

[1] A. Li, T. Yang, X. Zhan, Y. Shi, and H. Li, "Utilizing Data Science and AI for Customer Churn Prediction in Marketing," Journal of Theory and Practice of Engineering Science, vol. 4, no. 5, pp. 72-79, 2024, doi: 10.53469/jtpes.2024.04(05).10.

[2] X. Xiahou and Y. Harada, "B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM," Journal of Theoretical and Applied Electronic Commerce Research, vol. 17, no. 2, pp. 458-475, 2022, doi: 10.3390/jtaer17020024.

[3] P. Kotler and K. L. Keller, Marketing Management, 15th global ed. London: Pearson Education Limited, 2016.

[4] C. Wang, C. Rao, F. Hu, X. Xiao, and M. Goh,

"Risk assessment of customer churn in telco using FCLCNN-LSTM model," Expert Systems with Applications, vol. 248, p. 123352, 2024, doi: 10.1016/j.eswa.2024.123352.

[5] L. Zhang and Q. Wei, "Personalized and contextualized data analysis for e-commerce customer retention improvement with Bi-LSTM churn prediction," IEEE Transactions on Consumer Electronics, vol. 70, no. 1, pp. 1-12, 2024, doi: 10.1109/TCE.2024.3376672.

[6] K. Peng and Y. Peng, "Research on Telecom Customer Churn Prediction Based on GA- XGBoost and SHAP," Journal of Computer and Communications, vol. 10, no. 11, pp. 107-120, 2022, doi: 10.4236/jcc.2022.1011008.

[7] X. Liu, G. Xia, X. Zhang, W. Ma, and C. Yu, "Customer churn prediction model based on hybrid neural networks," Scientific Reports, vol. 14, p. 30707, 2024, doi: 10.1038/s41598-024-79603-9.

[8] S. Moro, P. Cortez, and P. Rita, "A data-driven approach to predict the success of bank telemarketing," Decision Support Systems, vol. 62, pp. 22-31, 2014, doi: 10.1016/j.dss.2014.03.001.

[9] S. Moro, P. Cortez, and P. Rita, "Extreme Gradient Boosting Trees with Efficient Bayesian Optimization for Bank Telemarketing Campaigns," Journal of Open Innovation: Technology, Market, and Complexity, 2025 (In Press).

[10] P. P. Singh, F. I. Anik, R. Senapati, A. Sinha, N. Sakib, and E. Hossain, "Investigating customer churn in banking: a machine learning approach and visualization app for data science and management," Data Science and Management, vol. 7, no. 1, pp. 7-16, 2024, doi: 10.1016/j.dsm.2023.09.002.

[11] S. Ö. Arik and T. Pfister, "TabNet: Attentive interpretable tabular learning," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 8, pp. 6679-6687, 2021, doi: 10.1609/aaai.v35i8.16826.

[12] S. M. Lundberg and S.-I. Lee, "A unified approach to interpretable machine learning," in Advances in Neural Information Processing Systems 30 (NIPS 2017), 2017, pp. 4765–4774. [Online]. Available: https://arxiv.org/abs/1705.07874.

[13] Q. Shi and Y. Shen, "Explainable customer churn prediction model based on deep learning," Expert Systems with Applications, vol. 248, p. 123416, 2024, doi: 10.1145/3654823.3654875.

[14] J. Kali, "Self-supervised learning for customer churn prediction," M.S. thesis, Tilburg University, Netherlands, 2023. [Online]. Available: https://www.kaggle.com/datasets/blastchar/telco- customer-churn.

[15] S. Pyo and S. O. Choi, "Regional innovation and economic growth: Empirical insights from FGLS, FE-DKSE, and XGBoost-SHAP approach," Journal of Open Innovation: Technology, Market, and Complexity, vol. 11, p. 100524, 2025, doi: 10.1016/j.joitmc.2025.100524.

[16] Y. Sun and H. Ma, "Interpretable analysis of transformer winding vibration characteristics: SHAP and multi-classification feature optimization," International Journal of Electrical Power and Energy Systems, vol. 166, p. 110585, 2025, doi: 10.1016/j.ijepes.2025.110585.

[17] H. Zhou et al., "Application of conversational intelligent reporting system based on artificial intelligence and large language models," Journal of Theory and Practice of Engineering Science, vol. 4, no. 3, pp. 176-182, 2024, doi: 10.53469/jtpes.2024.04(03).16.

[18] S. D. Mfazi, "Bank Customer Churn Dataset," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/sandiledesmond mfazi/bank-customer-churn.