# MACHINE LEARNING-BASED SYSTEM ARCHITECTURE FOR PROACTIVE PREDICTION OF CREDIT CARD CUSTOMER CHURN IN THE BANKING SECTOR

**Ricky Pieter Palembangan[1), ]Ahmad Nurul Fajar[2)]**
[1, 2)]Magister Sistem Informasi, Universitas Bina Nusantara
Jakarta, Indonesia
e-mail: ricky.palembangan@binus.ac.id[1)], afajar@binus.edu[2)]

**ABSTRACT**

*The rapid digital transformation in the banking industry has increased competition and strengthened the urgency of pro-active customer retention, especially in the credit card business which has a direct impact on profitability and long-term value of customers. This research aims to develop and evaluate comparatively machine learning models to predict credit card customer churn at PT XYZ and design a system architecture that supports proactive retention actions. Using the CRISP-DM framework, this study analyzed 11,314 customer data covering demographic, transactional, and account char-acteristics. Three classification algorithms, namely Logistic Regression, Decision Tree, and Random Forest, were evaluat-ed using accuracy, precision, recall, F1-score, and AUC-ROC metrics. The results showed that Random Forest provided the best performance with an AUC value of 0.9860 and an accuracy of 97.19%. A feature importance analysis shows that the number of transactions, the number of active cards, and the length of time since the last transaction are the main predic-tors of churn. Based on these findings, this study proposes a machine learning-based system architecture that integrates prediction outputs with CRM and monitoring dashboards to support more timely and targeted retention strategies*

***Keywords**: Credit card customer churn, Machine learning, Proactive retention, Banking information systems, Predictive analytics, System architecture.*

## I. INTRODUCTION

The contemporary financial services landscape is undergoing rapid transformation due to digitalization, data-driven innovation, and changing consumer expectations for faster, personalized, and always on services. In this context, traditional banks are facing increasingly strong competitive pressures, not only from fellow financial institutions, but also from fintech companies that offer more agile, user-oriented payments, financing, and service experience solutions. These changes have also shifted the basis of competition in the banking industry from just operational efficiency to the ability to build customer experience, loyalty, and long-term relationships based on technology. Therefore, customer retention is a strategic issue because it is directly related to profitability, firm value, and the sustainability of customer-bank relationships. In this increasingly digital and competitive environment, retention approaches that are only reactive are inadequate, so banks need analytics capabilities that are able to identify churn risks early and support more proactive interventions [1] [2].

Overall, the financial services landscape has changed fundamentally due to the acceleration of digital transformation, fintech expansion, and the increasing utilization of artificial intelligence in banking operations. In the credit card sector, these competitive pressures make customer loyalty even more fragile and increase the need for an information system that is able to identify churn risks early so that banks do not only react when customers have decided to leave [3] [4] [5].

Customer churn represents more than just an operational metric, it is a strategic threat capable of eroding market share and profitability. Customer churn impacts revenue, service costs, retention campaign effectiveness, and customer lifetime value. Recent literature shows that organizations that still rely on reactive interventions tend to lose momentum to retain customers, while a proactive approach based on machine learning allows for earlier risk identification and more efficient allocation of retention resources [1] [6] [7].

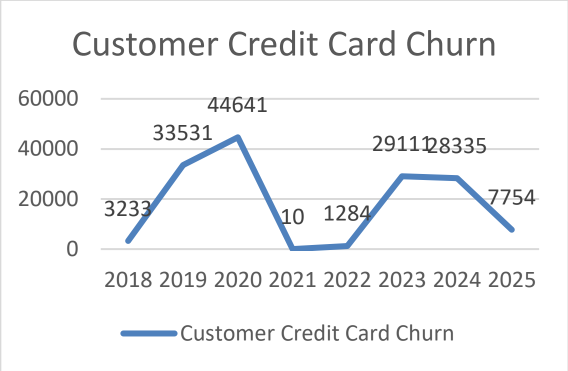


Figure 1. Historical Credit Card Churn Trends at PT XYZ (2018–2025)

Based on the internal historical data archive of PT XYZ's credit card unit for the 2018–2025 period, the descriptive statistics of the churn rate demonstrate significant historical fluctuation. The absolute number of closed or non-renewed cards exhibited a massive surge, increasing by approximately 935% from 3,238 accounts in 2018 to 33,531 accounts in 2019, and peaking at 44,641 accounts in 2020. Although it experienced an extreme contraction to just 10 accounts in 2021, the volume rebounded significantly in the subsequent periods, stabilizing at an average of 18,488 churned customers annually throughout the observed timeframe. This recurring quantitative trend indicates that churn is not a random anomaly, but rather a persistent systemic issue that has the potential to suppress the sustainability of credit card-based revenue. Therefore, Figure 1 is positioned as preliminary empirical evidence that PT XYZ needs an early prediction mechanism integrated with the retention business process, not just historical reporting.

The research problem departs from the fact that PT XYZ already has rich transactional and demographic data assets, but churn handling is still predominantly reactive. In other words, organizations already have enough data to build an early-warning system, but they don't yet have a validated predictive model and system architecture design that bridges analytics output with operational actions in CRM, call centers, and managerial dashboards.

In the past six years, studies of churn prediction in the banking, financial services, and subscription-based services sectors have shown that machine learning algorithms generally outperform traditional approaches. Random Forest, gradient boosting, and ensemble models are often reported to have high performance for capturing nonlinear relationships and inter-feature interactions, while cross-sector studies also confirm the practical value of machine learning pipelines for churn prediction in different service contexts [8] [9] [10] [11] [12]. However, other studies show that composite deep learning models, profit driven approaches, or explainability enriched pipelines can offer different advantages, for example in optimizing business objectives, interpretability, or resilience to data complexity [13] [14].

A review of the literature also shows that most research stops at the evaluation stage of the offline model. Some studies have begun to emphasize explainability, probability calibration, local-global interpretation, or visualization applications, but implementation within the framework of information systems that connect data pipelines, serving models, risk score storage, CRM, and executive dashboards is still relatively limited [15] [16] [17] [18]. On the other hand, recent surveys confirm that the gap between model accuracy and business operationalization remains an important research agenda in contemporary churn prediction studies [19] [6].

Thus, the research gap of this study lies in the limited studies that simultaneously: (1) compare several churn prediction models in the context of banking credit cards, (2) identify relevant churn drivers for business interventions, and (3) translate the results of the model into a system architecture design that can be used in an information systems environment. In addition, there are findings that are not always consistent in the literature: some studies place the Random Forest and ensemble tree as the most stable models [20] [21], while other studies show the advantages of hybrid models, deep learning, or explainable AI based approaches in different contexts and metrics [13] [22] [7].

The novelty of this research lies in the combination of two contributions that are rarely presented simultaneously in the study of credit card customer churn, namely the validation of machine learning-based predictive models and the design of a system architecture that supports proactive implementation at the operational level. Different from studies that mainly focus on algorithm benchmarking, this study places model output as a core component in the design of ETL flows, serving models, risk tables, CRM integration, and monitoring dashboards. This position is in line with recent literature trends that demand the integration of predictability, explain ability, and actionable business decision-making [23] [24] [25].

Based on this background, this study aims to: (1) develop and compare the performance of the

Logistic Regression, Decision Tree, and Random Forest models in predicting credit card customer churn at PT XYZ; (2) identify the variables that have the most influence on the probability of customer churn; and (3) designing a system architecture that supports the proactive implementation of customer retention strategies. Academically, this research contributes to the development of studies on the operationalization of machine learning in the domain of information systems in the banking sector, especially related to the integration of predictive models into business processes, decision-making, and data-oriented system design. In practical terms, this research provides an empirical foundation for banking institutions to develop retention strategies and campaigns that are more responsive, measurable, and aligned with business priorities, thereby supporting increased effectiveness of interventions against customers at risk of churn.

## II. LITERATURE REVIEWS

In the banking context, customer churn is defined as a situation in which a customer stops using a bank's products or services, either voluntarily or due to dissatisfaction with the services provided. Keramati et al. [26], in their study on electronic banking services, emphasize that customer retention is a fundamental need for banking organizations, and as cited from Nie et al. [27] in that study, a 5% increase in customer retention rates has the potential to boost bank profits by up to 85%. Specifically in the credit card segment, Li and Yan [14] state that predicting credit card customer churn plays a crucial role in the banking industry's risk management, helping banks identify potential risks and implement appropriate mitigation measures.

Conventional rule-based methods and customer surveys are no longer deemed sufficient to address the ever-changing dynamics of customer behavior. As emphasized by Al-Quraishi et al. [28], machine learning (ML) models and artificial intelligence enable banks to identify hidden churn trends and generate automated decision-making processes to support more proactive retention strategies. This has spurred the development of various machine learning approaches in the domain of banking churn prediction. AL-Najjar et al. [8] developed a credit card customer churn prediction model using a feature selection method combined with five machine learning algorithms, including Bayesian Network, C5 Tree, CHAID Tree, Classification and Regression Tree, and Neural Network, with independent variables covering customer demographics, transaction behavior, credit limits, and the duration of the customer's relationship with the bank.

Regarding algorithm performance comparisons, a comparative study conducted by Boozary et al. [29] shows that ensemble models specifically XGBoost and Random Forest consistently outperform conventional classifiers such as K-Nearest Neighbors and Logistic Regression in terms of accuracy, precision, recall, and F1-score, making them the most reliable choice for accurate and trustworthy churn prediction needs. It should be noted that the study used a dataset from the telecommunications industry, so its validation in a banking context remains relevant but requires feature adjustments. More specifically in the credit card domain, Li and Yan [30] demonstrated that an XGBoost model combined with four data balancing techniques Random Oversampling, SMOTE, Borderline-SMOTE, and ADASYN successfully achieved high performance with accuracy, precision, recall, F1-score, and AUC values of up to 97% on a dataset of 10,127 credit card customer samples. The study also integrated SHAP values for model interpretability and R-learner causal inference to analyze the causal impact of variables on customer churn decisions. More complex ensemble stacking approaches also yielded competitive results, with Vu [31] designing a two-level model that integrates KNN, XGBoost, Random Forest, and SVM at Level 0, and Logistic Regression, Recurrent Neural Network, and Deep Learning Neural Network at Level 1, achieving a precision of 98.74%, a recall of 91.27%, an accuracy of 95.13%, and ROC-AUC and AUC-PR of 99.17% and 99.27%, respectively.

Regarding the identification of key churn determinants, Li and Yan [14] found that the total number of transactions, total transaction value, the number of bank products held by customers, and changes in the number and value of transactions from the fourth quarter to the first quarter were proven to have a significant influence on credit card customer churn, thereby providing a strong theoretical foundation for data-driven customer management. Meanwhile, at the level of account activity-based features, Budiarto et al. [32] identified that the frequency of financial transactions and the volume of customer loans are the primary predictive features, with Random Forest consistently outperforming other models across various evaluation metrics. However, these findings were derived from a synthetic dataset, so their generalizability to real banking data requires further verification.

A technical challenge that must be addressed in churn prediction modeling is class imbalance, given that the number of churning customers is far smaller than that of active customers. Various resampling techniques, such as SMOTE and its variants, have proven effective in balancing class distributions and improving model performance, as applied in several studies discussed earlier. Furthermore, model interpretability is now a growing concern; as emphasized by Li and Yan [14], traditional machine learning methods are no longer sufficient for effective customer management without the ability to transparently explain prediction results.

Recent developments are moving toward the design of prediction systems that are not only statistically accurate but also integrated into real-time customer retention business processes. Singh et al. [33] developed an ML-based data visualization application equipped with real-time monitoring capabilities, enabling banks to track attrition rates, customer behavior, and other relevant indicators directly, thereby supporting rapid and targeted decision-making. The study also found that different predictive algorithms are more appropriately applied to different customer segments, resulting in predictions that are more practically impactful. Thus, the development of a machine learning system architecture that is proactively integrated with business retention processes has become an urgent need and forms the foundation of this research.

## III. RESEARCH METHOD

This study uses an explanatory quantitative approach based on predictive analytics with a case study design on PT XYZ. This approach was chosen because the main purpose of the research is not only to describe churn, but to model the probability of churn empirically, compare the performance of several algorithms, and translate the results into an operable information system design.

The methodological framework used in this study is the Cross-Industry Standard Process for Data Mining (CRISP-DM), which is a standard process model that systematically prepares data mining projects through six main stages, namely business understanding, data understanding, data preparation, modeling, evaluation, and deployment. In the business understanding phase, researchers define credit card customer churn as a strategic retention issue. In the data understanding phase, the researcher examines the data structure, variable distribution, missing values, outliers, and initial patterns of churn through descriptive statistics and exploratory visualization. In the data preparation phase, imputation, logarithmic transformation, categorical variable encoding, and feature reduction that have the potential to cause data leakage are carried out. In the modeling phase, the model is trained and tuned using stratified split and GridSearchCV. In the evaluation phase, the model was compared using accuracy, precision, recall, F1-score, and AUC-ROC with an emphasis on recall and F1-score. In the deployment phase, the research compiled a system architecture blueprint for the integration of the model with CRM processes and operational dashboards. Recent literature shows that CRISP-DM is still a widely used and relevant framework for data science projects that demand linkages between model performance and business implementation [34] [35].

Most churn prediction studies use the CRISP-DM method, to ensure alignment between business objectives and analytical processes. The selection of CRISP-DM is based on thre considerations. First, it provides a systematic flow from business understanding to deployment so that it is suitable for churn research that demands a close relationship between business objectives and analytical steps. Second, CRISP-DM is widely used in customer analytics and banking analytics studies because it is flexible for data preparation, modeling, and evaluation stages. Third, from an information systems perspective, CRISP-DM facilitates the translation of model results into process design and implementation architecture [20] [5] [18]. Within this framework, the Random Forest algorithm introduced by Breiman has become the gold standard for class-ification tasks involving unbalanced datasets, due to the nature of the ensemble and the mechanism of innate feature importance [36] [37].

The research population comprises all PT XYZ credit card customers recorded in the internal historical database for the 2018–2025 period. Initially, the research sample consisted of 11,314 customer records encompassing demographic attributes, account status, transaction behavior, and payment patterns. However, during the data pre-processing phase, the dataset was reduced to 11,017 records due to the targeted removal of 297 observations. This exclusion was explicitly performed because these specific records exhibited systematic missingness in fundamental transactional variables, namely LastTrxDate and MonthsSinceLastTrx, indicating that these customers likely never initiated a transaction following their card acquisition. Deleting these 297 records, rather than applying data imputation, was a necessary methodological decision to prevent the introduction of data noise and bias that could distort the recency calculation and mislead the model's pattern recognition. Thus, this study utilizes refined organizational administrative data that represents a census of active analysis units, not survey samples.

The data required includes account identity attributes, customer characteristics, transaction frequency, billing amount, usage limit, card active period, and churn target variables. The data collection technique is carried out through the extraction of operational data archives and internal documentation which is then anonymized for analysis purposes. Given the moderate class imbalance in the data set (churn: 27.9%), the Synthetic Minority Over-sampling Technique (SMOTE) was applied during the training phase to balance the class distribution. The selection of SMOTE is based on its ability to produce synthetic samples that are often used to reduce bias against the majority class, although its effectiveness still needs to be evaluated empirically on each dataset [15] [38].

The analysis methods used include descriptive analysis, exploratory data analysis, data preprocessing, classification modeling, performance evaluation, feature importance analysis, and system architecture design. Three algorithms were chosen for different but complementary reasons: Logistic Regression was maintained as an interpretive baseline; Decision Tree was chosen because it is easy to explain to business stakeholders; and Random Forest was chosen because it is known for its strong handling of non-linearity, feature interactions, and data variations in churn problems and banking analytics [8] [9] [12].

To ensure computational efficiency and reproducibility (random_state=42), the entire analytics process was executed using Python as the primary programming language. Specific machine learning libraries were utilized according to their functions: pandas and NumPy for data manipulation and preprocessing, Matplotlib for exploratory data visualization, imbalanced-learn for handling class imbalance via the Synthetic Minority Over-sampling Technique (SMOTE), and scikit-learn for building and evaluating the predictive models. Furthermore, to achieve optimal predictive performance, the models were fine-tuned using GridSearchCV. The final optimized configurations deployed in this study were: Random Forest (n_estimators=200, max_depth=10, min_samples_split=2, min_samples_leaf=1, class_weight='balanced'), Decision Tree (max_depth=10, min_samples_split=2, min_samples_leaf=1, class_weight='balanced'), and Logistic Regression (C=0.01, penalty='l2', class_weight='balanced'). In more detail each phase in the context of the research will be explained in the following subsection.

*A. Data Understanding*

The data set used in this study consists of 11,314 historical records of PT XYZ credit card customers for the period 2018–2025. The binary target variable (Churn) is set based on the card's last expiration date (LastExpDate) and the absence of an active card (TotalActiveCard = 0), which reflects the business logic that customers who don't renew a card that has matured are treated as churners. Exploratory analysis was conducted through class distribution tabulation, histograms, boxplots, and initial correlation checks to understand the data structure prior to modeling.

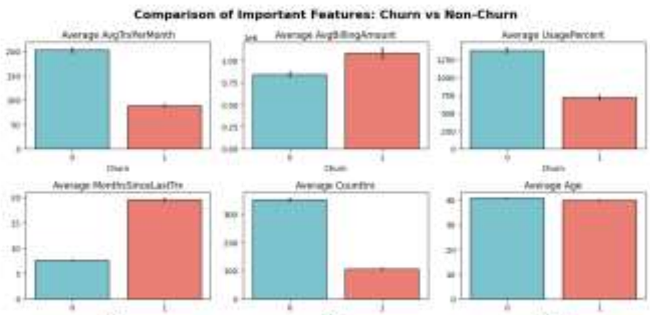


Figure 2. Comparison of Important Features

The dataset shows a moderate class imbalance with 27.9% of customers in churn status. The descriptive findings show a consistent pattern: churn customers tend to have lower transaction frequency, longer periods of inactivity, and narrower product engagement than relenting customers. Features such as credit limits, billing amounts, and usage percentages also show long tail distributions, so transformation and control of the influence of extreme values are needed to make the model more stable.

*B. Data Preparation*

In line with the CRISP-DM framework, the data preparation stage is carried out through four main steps, namely: (1) standardization of scale and numerical transformation, (2) handling missing values, (3) coding and elimination of features that are redundant or at risk of causing data leakage, and (4) conservative handling of outliers. This step structure is used to make pre-processing decisions more transparent and searchable by both reviewers and users of the model.

To improve interpretability and reduce computational instability due to scale differences, the high-magnitude financial variables, namely CustLimit, AvgBillingAmount, and TotalBillingAmount, were converted into thousands of rupiahs. In addition, the highly right-skewed numerical features are transformed logarithmically to stabilize the variance and reduce the dominance of extreme observations.

Missing values are addressed using context-sensitive strategies for the data. Categorical features such as CustSex and CustMaritalStatus that have a low missing rate are imputed using the mode, while observations that miss key behavioral indicators especially MonthsSinceLastTrx are excluded from the final dataset because they have the potential to degrade the validity of churn signals and create bias in the model.

Categorical variables are encoded with a one-hot encoding approach and reference categories are omitted to prevent multicollinearity in linear models. At the same time, non-predictive identifiers such as CIF and CustomerName, as well as temporal features that have been aggregated in other variables, are removed to make the feature space more concise and freer from the risk of information leakage.

Outliers on features like CustLimitInThousand, UsagePercent, and CountTrx aren't automatically removed because some of them represent premium customer behavior or usage intensity that's indeed legitimate for business. Therefore, this study prefers logarithmic transformations to aggressive trimming, in line with cutting-edge practices in churn modeling and credit analytics that seek to keep edge case information available to the model [21] [22].

A systematic approach to data preparation results in a feature matrix that is cleaner, more statistically stable, and more suitable for cross algorithm comparison in the modeling phase.

### C. *Modeling Approach*

Following the CRISP-DM framework, the preprocessed results (data_processed) dataset consists of 11,017 customer records with 14 predictor features and one binary target variable (Churn). This composition is considered adequate to evaluate the performance of the baseline model and the ensemble model in binary classification problems in the context of credit card customer churn.

The main transformations at this stage include: (1) log transformations for highly skewed features; (2) feature selection to eliminate irrelevant or leakage-prone attributes; (3) binary coding for simple categorical variables; and (4) standardization of specific features for Logistic Regression with StandardScaler. The dataset is then separated into training and testing data using a stratified split of 80:20 to maintain class proportions.

The datasets were then partitioned into training and test sets using 80:20 graded separation to maintain the original class distribution (churn: 27.9%, non-churn: 72.1%). This resulted in 8,813 training samples and 2,204 test samples. To accommodate the sensitive nature of Logistic Regression, the training and testing sets for this model are standardized using StandardScaler. In contrast, the Decision Tree and Random Forest models, which are invariant to feature scaling, are trained on native (non-scaled) data.

All models are initialized with random_state = 42. Hyperparameter tuning is done using GridSearchCV with stratified 3-fold cross-validation. This strategy was chosen to keep comparisons between models fair and to reduce the likelihood that the best models only excel because of a particular default configuration, not because of their match to the data structure.

Explain the research chronology, including the research design, research procedures (in the form of algorithms, pseudocode, or other formats), and how to test and collect data [1], [3]. The description of the research flow must be supported by references so that the explanation can be accepted scientifically.

The following tables and figures are centered, as shown in the tables and figures below, and are referenced in the article manuscript.

## IV. RESULTS AND DISCUSSIONS

### A. *Model performance*

Model performance was evaluated on the hold-out test set using accuracy, precision, recall, F1-score, and AUC-ROC. Although accuracy is still reported for completeness, the study places the recall and F1-score as key metrics because the problem of churn is unbalanced and the cost of misclassification of churns is higher in business. The latest literature also confirms that F1-score is more representative for evaluating the precision-recall balance in the context of unbalanced data-based business predictive modeling [25].

The results of the evaluation are summarized in Table 1 and the subsequent analysis.

Table 1. Comparison of Machine Learning Model Performance on Test Sets

| Models | Accuracy | Precision | Recall | F1 Score | AUC-ROC |
|---|---|---|---|---|---|
| Random Forest | 0.9719 | 0.9711 | 0.9271 | 0.9486 | 0.9860 |
| Decision Tree | 0.9601 | 0.9287 | 0.9287 | 0.9287 | 0.9505 |
| Logistic Regression | 0.9646 | 0.9484 | 0.9238 | 0.9360 | 0.9763 |

Interpretation of the results shows that Random Forest (RF) is the best model in Table 1. This model achieved an accuracy of 97.19%, precision of 97.11%, recall of 92.71%, F1-score of 94.86%, and an AUC of 0.9860. The dominance of RF is in line with banking churn and financial institution analytics studies which show that ensemble trees tend to be stable in capturing nonlinear interactions and are more resistant to noise than single models [20] [9].

Logistic Regression (LR) continues to show strong performance and is highly competitive, especially in terms of recalls, so this model remains relevant as a comparator that is easier to interpret. LR's superiority in interpretability makes it important for managerial decision-making contexts, although overall its discriminatory power still falls under Random Forest.

Decision Tree (DT) performs well, but lower AUC-ROC indicates an overall discriminating power that is not as strong as RF and LR. These findings are consistent with the literature that assesses single trees to be susceptible to overfitting and split structure instability compared to ensemble models or more calibrated approaches [12] [7].

Table 2. Performance Comparison Model with & without SMOTE

| Models | Scenario | Recall | F1 Score | AUC-ROC |
|---|---|---|---|---|
| Random Forest | With SMOTE | 0.9271 | 0.9486 | 0.9860 |
| | Without SMOTE | 0.9465 | 0.9419 | 0.9887 |
| Decision Tree | With SMOTE | 0.9287 | 0.9287 | 0.9505 |
| | Without SMOTE | 0.9417 | 0.9230 | 0.9515 |
| Logistic Regression | With SMOTE | 0.9238 | 0.9360 | 0.9763 |
| | Without SMOTE | 0.9238 | 0.9253 | 0.9756 |

Table 2 shows that the effects of SMOTE in this study were not entirely linear. In contrast to a number of studies that reported an increase in recall after oversampling, in this dataset, Random Forest without SMOTE actually produced slightly higher recall (0.9465) and higher AUC (0.9887) compared to the scenario with SMOTE (recall 0.9271; AUC 0.9860), although scenarios with SMOTE result in a higher F1-score (0.9486 versus 0.9419). This pattern indicates that at moderate class imbalances, synthetic oversampling can improve the balance of precision recall but also have the potential to increase overlap between classes or over-refine decision boundaries. These findings do not deny the usefulness of SMOTE, but confirm that its benefits should be empirically tested on each dataset, as also emphasized by recent methodological studies on variation in the performance of resampling techniques [38].

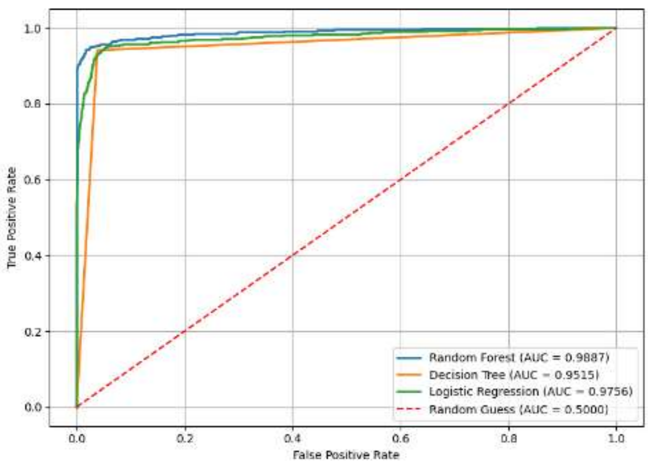


Figure 3. Random Forest ROC Curve, Decision Tree, Logistic Regression

Based on a comprehensive evaluation, Random Forest was established as the main model for credit card customer churn prediction at PT XYZ. The ROC curve in Figure 2 shows that the RF curve is closest to the upper left corner, which means that the model has the best classification ability in distinguishing churners and non-churners across different thresholds. From a business perspective, this performance is important because banks need a model that is not only accurate, but also capable of capturing the majority of at-risk customers without generating too many false alarms. These findings support the results of previous studies that position RF as a strong option in the context of banking churn, although other literature has also shown that deep learning models or profit-driven classifiers can excel when analytical goals are focused on profit optimization, probability calibration, or advanced interpretability [13] [14].

Table 3. Optimal Hyperparameter Configuration

| Models | Best Parameters |
|---|---|
| Random Forest | n_estimators = 100, max_depth = 10, min_samples_split = 10, min_samples_leaf = 4 |
| Decision Tree | max_depth = 20, min_samples_split = 2, min_samples_leaf = 1 |
| Logistic Regression | C = 10, penalty = L2 |

To ensure a fair comparison, hyperparameter tuning is performed using GridSearchCV with stratified 3-fold cross validation. The optimal configuration in Table 3 shows that the best performance is not achieved through a very deep tree or an overly complex configuration, but through a balance between model capacity and overfitting control.

Table 4. Top Features Importance Ranking

| Ratings | Feature Name | Importance Score | Business Interpretation |
|---|---|---|---|
| 1 | Calculate Transactions | 41.59% | Transaction volume indicator |
| 2 | TotalActive Card | 19.64% | Diversity of product portfolio |
| 3 | MonthSinceLastTrx | 15.94% | Novelty of customer transactions |
| 4 | TrxPerMonth Average | 9.25% | Historical transaction frequency |
| 5 | Usage Percentage | 4.42% | Credit utilization pattern |
| 6 | TotalBilling-InThousand | 3.79% | Total amount of billing |
| 7 | CustVisaCard | 1.77% | Visa card ownership |
| 8 | AverageBilling-InThousand | 1.24% | Average billing behavior |
| 9 | CustLimitInThousand | 0.82% | Number of credit limits |
| 10 | CustMastercard-Card | 0.75% | Mastercard Ownership |

Business interpretation of feature importance outcomes provides actionable insights for the development of retention strategies. Transaction volume analysis shows that customers with a declining number of transactions show an exponentially increased churn probability. Critical thresholds arise when monthly transaction volume declines by more than 40 percent compared to historical averages, providing clear guidance for automated alert systems and proactive outreach times.

Some active card relationships emerged as a strong retention factor, with customers retaining two or more active cards showing a 65 percent lower probability of churn compared to single card holders. These findings point to opportunities for cross selling strategies that focus on increasing the depth of the product portfolio rather than breadth, emphasizing complementary card products that increase customer engagement.

An analysis of customer transaction inactivity revealed that customers who were inactive for more than four months showed a churn probability of 78 percent, while customers who were inactive for six months exceeded a churn probability of 85 percent. Critical intervention thresholds appear at two months of inactivity, providing clear guidance for proactive outreach timing and campaign triggering mechanisms.

## B. *Proposed architecture and dashboard*

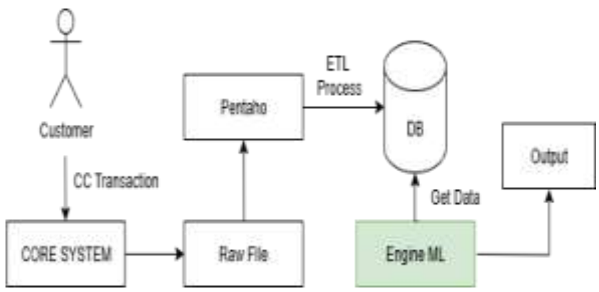


Figure 3. Proposed blueprint architecture machine learning

Although the proposed architecture is still a conceptual blueprint and has not been fully implemented in PT XYZ's production environment, this design shows how the Random Forest model can be operationalized in the banking information systems ecosystem. In practical terms, the model can be packaged as an API service, linked to a scheduled ETL pipeline, and store the risk score in a custom table that is then called by the CRM to trigger a retention intervention. This approach is in line with the principles of MLOps and operational analytics which emphasize the integration between data pipeline, service models, monitoring, and business actions [18].

Figure 4. Proposed mockup dashboard for PT XYZ credit card churn prediction system

The proposed dashboard serves as a decision-making interface for marketing and risk management teams. The integration of the model's output with business indicators allows for monitoring the distribution of churn risks, evaluating the effectiveness of retention campaigns, and reviewing intervention thresholds. From the perspective of information systems, this dashboard is not just a visualization, but a medium of coordination between analytical functions, operational functions, and managerial functions so that predictive insights are truly translated into timely and measurable actions.

### *C. Comparison with Related Jobs*

Previous studies of customer churn have generally focused on algorithm benchmarking, feature enrichment, or improving model interpretability. In the context of banking, studies have evaluated Logistic Regression, Decision Tree, Random Forest, gradient boosting, and deep learning with mixed results, depending on the characteristics of the data, optimization goals, and evaluation metrics used [8] [39] [12]. Recent systematic surveys and reviews also confirm that the churn prediction literature is moving towards ensemble learning, deep learning, explainable AI, and evaluation that is more sensitive to class imbalances [6] [19].

However, these results are not always consistent. Some studies report the advantages of Random Forest or a feature optimized tree based model for the context of banking and internet banking service marketing [9] [21], while other research shows that composite models, deep learning, or profit driven approaches can provide additional advantages in different contexts [13] [7]. This contradiction shows that no single model is universally superior. Performance remains influenced by data structures, churn definitions, and business decision goals.

In this context, the novelty of this research lies not only in the empirical proof that Random Forest is best suited for the PT XYZ dataset, but also in the preparation of a system architecture that places the model as an operational component in the CRM and dashboard workflows. Thus, this study bridges the gap between model-centric research and deployment-oriented information systems research.

The contribution of the research can be summarized in three ways. First, this study provides empirical evidence on the comparative performance of three algorithms in the context of credit card customer churn in Indonesia. Second, this study identifies churn drivers that can be directly transformed into business intervention rules. Third, this study offers an implementation blueprint that can be further developed towards a real-time analytics system, monitoring model, and explainable decision support for bank management.

## V. CONCLUSION

This research shows that machine learning is effectively used to predict credit card customer churn in the context of banking. Of the three models compared, Random Forest provided the best performance, so it was considered the most suitable to support the early identification of customers at risk of churn. The research findings also show that customer behavior variables, specifically transaction intensity, number of active cards, and duration of transaction inactivity, are the main predictors that affect churn probability. Thus, this research not only contributes to the prediction aspect, but also offers a conceptual basis for the development of a system architecture that proactively supports customer retention strategies.

The main contribution of this research operates on two complementary levels. At the analytical level, this study contributes to machine learning research in the banking sector by providing a transparent and methodologically rigorous comparative evaluation of multiple machine learning models specifically calibrated to the context of credit card customer churn, employing internal customer data from PT XYZ's credit card unit spanning the 2018–2025 period thereby producing evaluation results that are empirically grounded in real banking operations rather than relying on publicly available or synthetic datasets. At the information systems level, this study contributes to the operational implementation of churn prediction systems by proposing a system architecture that explicitly maps the pathway from model output to business action, demonstrating how machine learning outputs can be integrated into an Extract-Transform-Load (ETL) pipeline, a serving model layer, a Customer Relationship Management (CRM) module, and an executive decision-making dashboard bridging the critical gap between predictive modeling and operational deployment so that the practical value of machine learning does not remain confined to the experimental stage but is translated into actionable retention strategies within the banking business process.

This research has several limitations. First, the data comes from a single institution so cross-bank generalizations need to be done carefully. Second, the proposed system architecture is still conceptual and has not been tested in a full production environment. Third, this study limits model comparisons to three main algorithms so that it does not include gradient boosting models, hybrid deep learning, or probability calibration schemes that in some studies have proven promising.

Further research is suggested to test the model on multi-institutional datasets to assess generalizability across different banking environments and customer profiles. On the modeling side, future studies are encouraged to evaluate more cutting-edge algorithms such as XGBoost, LightGBM, and hybrid neural networks, as well as to incorporate explainable AI (XAI) methods such as SHAP and LIME to support transparency and accountability in retention decision-making. Regarding system implementation, which falls outside the scope of the current study, future research could examine the feasibility of deploying the proposed architecture in a full production environment, including the development of a streaming and MLOps-based infrastructure to enable real-time churn scoring, automated model monitoring, and continuous retraining pipelines. Such implementation studies would be essential to evaluate whether the system architecture proposed in this research can sustain performance under the dynamic and high-volume conditions characteristic of banking operational environments, and to identify practical challenges that may arise during the transition from experimental modeling to live deployment.

## REFERENCES

[1] J. B. G. Brito *et al.*, “A framework to improve churn prediction performance in retail banking,” *Financ. Innov.*, vol. 10, no. 1, p. 17, Jan. 2024, doi: 10.1186/s40854-023-00558-3.

[2] L. Kim, T. Jindabot, and S. F. Yeo, “Understanding customer loyalty in banking industry: A systematic review and meta analysis,” *Heliyon*, vol. 10, no. 17, p. e36619, Sep. 2024, doi: 10.1016/j.heliyon.2024.e36619.

[3] D. Broby, “Financial technology and the future of banking,” *Financ. Innov.*, vol. 7, no. 1, p. 47,

Dec. 2021, doi: 10.1186/s40854-021-00264-y.

[4] S. Gupta, A. Leszkiewicz, V. Kumar, T. Bijmolt, and D. Potapov, "Digital Analytics: Modeling for Insights and New Methods," *J. Interact. Mark.*, vol. 51, no. 1, pp. 26–43, Aug. 2020, doi: 10.1016/j.intmar.2020.04.003.

[5] 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 Sci. Manag.*, vol. 7, no. 1, pp. 7–16, Mar. 2024, doi: 10.1016/j.dsm.2023.09.002.

[6] M. Shahabikargar, A. Beheshti, X. Zhang, J. Foo, and A. Jolfaei, "A comprehensive survey on customer churn analysis studies," *J. Inf. Telecommun.*, vol. 10, no. 1, pp. 24–70, Jan. 2026, doi: 10.1080/24751839.2025.2528440.

[7] N. M. AbdelAziz, M. Bekheet, A. Salah, N. El-Saber, and ..., "A comprehensive evaluation of machine learning and deep learning models for churn prediction," *Information*, 2025.

[8] D. AL-Najjar, N. Al-Rousan, and H. AL-Najjar, "Machine Learning to Develop Credit Card Customer Churn Prediction," *J. Theor. Appl. Electron. Commer. Res.*, vol. 17, no. 4, pp. 1529–1542, Nov. 2022, doi: 10.3390/jtaer17040077.

[9] R. A. de Lima Lemos, T. C. Silva, and B. M. Tabak, "Propension to customer churn in a financial institution: a machine learning approach," *Neural Comput. Appl.*, vol. 34, no. 14, pp. 11751–11768, Jul. 2022, doi: 10.1007/s00521-022-07067-x.

[10] P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi, "Customer churn prediction system: a machine learning approach," *Computing*, vol. 104, no. 2, pp. 271–294, Feb. 2022, doi: 10.1007/s00607-021-00908-y.

[11] Y. Suh, "Machine learning based customer churn prediction in home appliance rental business," *J. Big Data*, vol. 10, no. 1, p. 41, Apr. 2023, doi: 10.1186/s40537-023-00721-8.

[12] A.-G. Văduva, S.-V. Oprea, A.-M. Niculae, A. Bâra, and A.-I. Andreescu, "Improving Churn Detection in the Banking Sector: A Machine Learning Approach with Probability Calibration Techniques," *Electronics*, vol. 13, no. 22, p. 4527, Nov. 2024, doi: 10.3390/electronics13224527.

[13] A. Khattak, Z. Mehak, H. Ahmad, M. U. Asghar, M. Z. Asghar, and A. Khan, "Customer churn prediction using composite deep learning technique," *Sci. Rep.*, vol. 13, no. 1, p. 17294, Oct. 2023, doi: 10.1038/s41598-023-44396-w.

[14] Y. Li and K. Yan, "Prediction of bank credit customers churn based on machine learning and interpretability analysis," *Data Sci. Financ. Econ.*, vol. 5, no. 1, pp. 19–34, 2025, doi: 10.3934/DSFE.2025002.

[15] S. C. K. Tékouabou, Ștefan C. Gherghina, H. Toulni, P. N. Mata, and J. M. Martins, "Towards Explainable Machine Learning for Bank Churn Prediction Using Data Balancing and Ensemble-Based Methods," *Mathematics*, vol. 10, no. 14, p. 2379, Jul. 2022, doi: 10.3390/math10142379.

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

[17] K. W. De Bock *et al.*, "Explainable AI for Operational Research: A defining framework, methods, applications, and a research agenda," *Eur. J. Oper. Res.*, vol. 317, no. 2, pp. 249–272, Sep. 2024, doi: 10.1016/j.ejor.2023.09.026.

[18] H. J. Watson and D. Larson, "MLOps: From a Cottage Industry to a Factory Approach," *Int. J. Bus. Intell. Res.*, vol. 15, no. 1, pp. 1–22, Nov. 2024, doi: 10.4018/IJBIR.358916.

[19] M. Imani, M. Joudaki, A. Beikmohammadi, and H. Arabnia, "Customer Churn Prediction: A Systematic Review of Recent Advances, Trends, and Challenges in Machine Learning and Deep Learning," *Mach. Learn. Knowl. Extr.*, vol. 7, no. 3, p. 105, Sep. 2025, doi: 10.3390/make7030105.

[20] J. B. G. Brito *et al.*, "A framework to improve churn prediction performance in retail banking," *Financ. Innov.*, vol. 10, no. 1, p. 17, Jan. 2024, doi: 10.1186/s40854-023-00558-3.

[21] F. Ehsani and M. Hosseini, "Customer churn analysis using feature optimization methods and tree-based classifiers," *J. Serv. Mark.*, vol. 39, no. 1, pp. 20–35, Jan. 2025, doi: 10.1108/JSM-04-2024-0156.

[22] F. M. Talaat, A. Aljadani, M. Badawy, and M. Elhosseini, "Toward interpretable credit scoring: integrating explainable artificial intelligence with deep learning for credit card default prediction," *Neural Comput. Appl.*, vol. 36, no. 9, pp. 4847–4865, Mar. 2024, doi: 10.1007/s00521-023-09232-2.

[23] S. S. Poudel, S. Pokharel, and M. Timilsina, "Explaining customer churn prediction in telecom industry using tabular machine learning models," *Mach. Learn. with Appl.*, vol. 17, p. 100567, Sep. 2024, doi: 10.1016/j.mlwa.2024.100567.

[24] G. Marín Díaz, R. Gómez Medina, and J. Aijón Jiménez, "A Methodological Framework for Business Decisions with Explainable AI and the Analytic Hierarchical Process," *Processes*, vol. 13, no. 1, p. 102, Jan. 2025, doi: 10.3390/pr13010102.

[25] K. M. Sujon, R. Hassan, K. Choi, and M. A.

Samad, "Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models," *J. Big Data*, vol. 12, no. 1, p. 268, Dec. 2025, doi: 10.1186/s40537-025-01313-4.

[26] A. Keramati, H. Ghaneei, and S. M. Mirmohammadi, "Developing a prediction model for customer churn from electronic banking services using data mining," *Financ. Innov.*, vol. 2, no. 1, p. 10, Dec. 2016, doi: 10.1186/s40854-016-0029-6.

[27] G. Nie, W. Rowe, L. Zhang, Y. Tian, and Y. Shi, "Credit card churn forecasting by logistic regression and decision tree," *Expert Syst. Appl.*, vol. 38, no. 12, pp. 15273–15285, Nov. 2011, doi: 10.1016/j.eswa.2011.06.028.

[28] T. Al-Quraishi, O. Albahri, A. Albahri, A. Alamoodi, and I. M. Sharaf, "Bridging Predictive Insights and Retention Strategies: The Role of Account Balance in Banking Churn Prediction," *AI*, vol. 6, no. 4, p. 73, Apr. 2025, doi: 10.3390/ai6040073.

[29] P. Boozary, S. Sheykhan, H. GhorbanTanhaei, and C. Magazzino, "Enhancing customer retention with machine learning: A comparative analysis of ensemble models for accurate churn prediction," *Int. J. Inf. Manag. Data Insights*, vol. 5, no. 1, p. 100331, Jun. 2025, doi: 10.1016/j.jjimei.2025.100331.

[30] Y. Li and K. Yan, "Prediction of bank credit customers churn based on machine learning and interpretability analysis," *Data Sci. Financ. Econ.*, vol. 5, no. 1, pp. 19–34, 2025, doi: 10.3934/DSFE.2025002.

[31] V.-H. Vu, "An Efficient Customer Churn Prediction Technique Using Combined Machine Learning in Commercial Banks," *Oper. Res. Forum*, vol. 5, no. 3, p. 66, Jul. 2024, doi: 10.1007/s43069-024-00345-5.

[32] R. A. Prayoga, H. A. Budiarto, M. F. Afif, A. S. Pradipta, and A. S. Lestari, "Peningkatan Minat Baca Siswa Melalui Pekan Literasi dan Pembiasaan 15 Menit Membaca: Studi Kasus MI Mulyadarama Girimukti," *War. LPM*, pp. 388–400, Oct. 2023, doi: 10.23917/warta.v26i4.1870.

[33] 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 Sci. Manag.*, vol. 7, no. 1, pp. 7–16, Mar. 2024, doi: 10.1016/j.dsm.2023.09.002.

[34] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.

[35] A. M. Shimaoka, R. C. Ferreira, and A. Goldman, "The evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks," *SBC Rev. Comput. Sci.*, vol. 4, no. 1, pp. 28–43, Oct. 2024, doi: 10.5753/reviews.2024.3757.

[36] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[37] P. Chapman, *CRISP-DM 1.0: Step-by-step data mining guide*, no. 1–73. SPSS Inc, 200AD.

[38] S. Matharaarachchi, M. Domaratzki, and S. Muthukumarana, "Enhancing SMOTE for imbalanced data with abnormal minority instances," *Mach. Learn. with Appl.*, vol. 18, p. 100597, Dec. 2024, doi: 10.1016/j.mlwa.2024.100597.

[39] A. Muneer, R. Faizan Ali, A. Alghamdi, S. Mohd Taib, A. Almaghthawi, and E. A. A. Ghaleb, "Predicting customers churning in banking industry: A machine learning approach," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 26, no. 1, p. 539, Apr. 2022, doi: 10.11591/ijeecs.v26.i1.pp539-549.