JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i2.7754

AN EXPLAINABLE AI APPROACH TO HYPERPARAMETER TUNING FOR XGBOOST CUSTOMER CHURN PREDICTION

Ahmad Fauzi[1*], Budi Al Amin[1], Heribertus Ary Setyadi[1], Dahlan Susilo[2]

Universitas Bina Sarana Informatika, Surakarta, Indonesia[1]
https://bsi.ac.id/[1]
ahmad.fzx@bsi.ac.id*, budi.bdm@bsi.ac.id, heribertus.hbs@bsi.ac.id

Universitas Sahid Surakarta, Surakarta, Indonesia[2]
https://www.usahidsolo.ac.id/[2]
dahlan.susilo@usahidsolo.ac.id

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

The creation is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License.

Submitted: 11 November 2026, **Revised:** 25 June 2026, **Accepted:** 07 July 2026, **Published:** 12 November 2026

Abstract— *In the 2026 digital-first economy, AI transparency has emerged as a fundamental necessity for data driven decision making. This research aims to address customer churn prediction challenges in the telecommunications sector by proposing an Explainable AI (XAI) framework. The proposed model utilizes XGBoost optimized through Bayesian Optimization (BO) to ensure peak predictive performance. The primary focus of this research is to bridge the gap between high-level model performance and the need for managerial interpretation through the implementation of SHAP (SHapley Additive exPlanations). To overcome the "black-box" nature of the algorithm, SHAP is integrated to provide interpretability for stakeholders. Analyzing 7,043 customer records, the study involves rigorous data preprocessing and automated hyperparameter tuning. Experimental results indicate that the optimized model outperforms the baseline, achieving an AUC of 0.91 and an accuracy of 0.87. Furthermore, SHAP analysis identifies tenure, contract type, and monthly charges as the most significant drivers of churn. This study demonstrates that combining advanced optimization with XAI techniques yields a model that is both highly accurate and strategically actionable for designing customer retention programs.*

Keywords: *Customer churn, explainable AI, hyperparameter tuning, model interpretability, XGBoost.*

INTRODUCTION

In current digital landscape, e-commerce has become a primary conduit for business transactions among both entrepreneurs and general public, largely due to its unparalleled geographical reach and round the clock availability [1]. Within this competitive framework, firms must prioritize a deep understanding of customer needs and preferences to cultivate enduring relationships [2]. Fulfillment of these consumer demands leads to enhanced satisfaction and long-term loyalty, fostering a robust and mutually advantageous synergy between both parties [3]. Given that customer retention is fundamentally anchored in satisfaction, organizations must consistently implement strategic refinements and service updates [4]. In modern commerce, cost-efficiency of retaining existing clientele often outweighs the expenses of new customer acquisition, while simultaneously driving increased sales performance [5].

Organizations today are increasingly focused on preventing customer churn and customer's relationship cessation with a company's offerings. This phenomenon represents a critical threat to business stability and revenue growth [6]. Often referred to as customer attrition, churn signifies the disconnection between consumer and business owner. Specifically within the IT sector, it denotes users to competing firms migration [7]. Various determinants, such as price sensitivity, product

performance, and customer behavior, drive this attrition [8]. Addressing churn is paramount, as failure to implement effective prevention strategies will lead to a significant erosion company's profitability [9].

An effective approach to reducing customer churn by leveraging historical data is through data mining techniques. Data mining encompasses both predictive and descriptive categories [10]. Predictive methods can be implemented by developing classification models [11]. which are constructed by learning from historical data patterns, often through the application of machine learning algorithms [12]. Several machine learning algorithms frequently utilized for churn prediction include Naïve Bayes, Support Vector Machines, K-Nearest Neighbors, and Decision Trees [13]. Naïve Bayes offers the advantage of rapid implementation and requires minimal training data; however, it is limited when data are not independent, as it tends to ignore feature correlations and feature selection [14]. To address class imbalance issues, ensemble techniques such as boosting and bagging are commonly employed [15].

XGBoost (Extreme Gradient Boosting) is a tree-based classification algorithm that utilizes an ensemble boosting framework [16]. Boosting core mechanism involves a sequential construction of models, where each subsequent model is designed to correct the errors of its predecessor, thereby iteratively increasing the final model's predictive complexity [17]. Given that most learning algorithms are highly sensitive to their internal settings, different parameter configurations often result in disparate performance outcomes [18]. Consequently, model evaluation process must integrate both algorithmic selection and precise parameter optimization. This systematic search for optimal parameter values is formally recognized as hyperparameter tuning [19].

To enhance model transparency, the SHAP technique is employed to quantify the significance of each feature in a decision making process [20]. By utilizing a game-theoretic approach, SHAP provides a granular interpretation of how each variable contributes to the prediction of a specific sample x [21]. Alongside interpretability, Bayesian Optimization provides a robust mechanism for fine-tuning machine learning hyperparameters. This technique is particularly valuable when dealing with objective functions that lack analytical derivatives [22]. Iterative process of Bayesian Optimization relies on an acquisition function to select the most promising parameter combinations. This is achieved by maintaining a critical equilibrium between testing uncertain parameter regions and refining values that have already demonstrated high performance, thereby ensuring an efficient search for the global optimum [23].

Prior research in customer churn prediction by Raut (2026) [24] utilized a hyperparameter-tuned XGBoost model that achieved a commendable accuracy of 89.7%, paired with SHAP for both global and local model interpretability. While their study successfully deployed an interactive Streamlit dashboard for real-time risk visualization, the hyperparameter configuration remained strictly static post-deployment. The architecture lacked a dynamic mechanism to integrate shifting business logics, such as real-time adjustments in retention cost allocations into an automated background re-tuning pipeline. Furthermore, their methodology relied entirely on a conventional benchmark telecom dataset, leaving a computational efficiency and scalability of tuning approach untested against rapidly fluctuating, large-scale big data frameworks.

Literature also features a framework where a Genetic Algorithm (GA) acts as an automated hyperparameter optimizer for XGBoost to counteract class imbalance, utilizing SHAP to map feature contributions. While conceptually innovative in its dual AI setup, the methodology treats the GA optimization process itself as a black box; it lacks post hoc interpretability to visually explain the internal logic of how GA selected configurations drive structural improvements in XGBoost. Additionally, by treating customer records as static data points rather than continuous time series sequences, the model overlooks temporal dependencies. As a result, the current SHAP implementation is incapable of tracking or explaining how evolving customer behavior trends over time influence the final churn trajectory [25].

Although the churn prediction framework developed by Pandikumar and Shoukath (2025) effectively delivers multi level feature interpretability for business insights, it restricts its transparency to input-output data level, omitting the optimization stage. In typical XGBoost-based telecom churn architectures, hyperparameter tuning is executed without post-hoc explanation. Consequently, while the optimal parameter layout can be empirically discovered, the internal mechanics specifically how parameter interactions structurally reshape the algorithm's decision boundaries remain uninterpreted, highlighting a need for explainable optimization frameworks [26].

To resolve the limitations of traditional architectures where SHAP is confined strictly to final classification outputs, this study proposes an integrative XAI approach designed for hyperparameter optimization. The existing

Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i2.7754

research gap lies in the complete lack of transparency surrounding how parameter changes alter internal model structures. We overcome this by utilizing a reverse-mapping SHAP framework to dissect the XGBoost parameter space, visualizing how hyperparameter sensitivity directly impacts both churn detection capacity and feature attribution stability. This research fills a vital gap in data science literature, offering an explainable tuning methodology that bridges the divide between algorithmic accuracy and structural transparency in industrial predictive analytics.

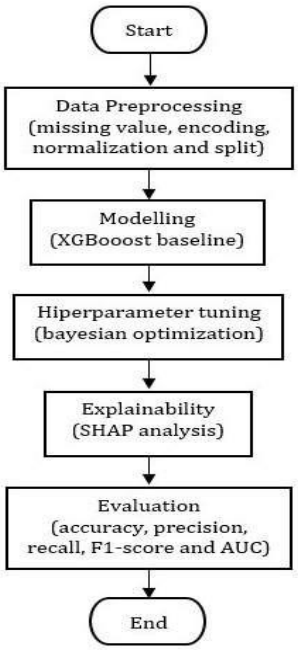
Despite its widespread adoption for churn prediction due to its scalability and precision, XGBoost's complex structure often renders it a black box hindering its explainability to stakeholders. This lack of transparency is a significant hurdle, as interpretability is essential for developing data-driven retention policies. Additionally, achieving optimal performance with XGBoost necessitates moving beyond default hyperparameters, which are often inadequate [27]. Unlike conventional tuning methods that demand exhaustive computational resources, Bayesian Optimization introduces a more efficient, probabilistic approach. This enables optimal hyperparameter discovery sets at a lower computational cost, offering a balanced solution for high accuracy and resource conscious churn modeling [28]. Additionally, model interpretability can be improved by SHAP method, which provides global as well as local explanations for model predictions. By utilizing SHAP, companies are able to identify which variables most impact customer churn, ensuring that the retention strategies implemented are more precisely targeted [29].

This research presents a customer churn prediction framework applied to IBM dataset, specifically targeting class imbalance problem through an optimized XGBoost model. This methodology integrates two data augmentation techniques, SMOTE and HAT, while benchmarking against several classifiers: SVM (RBF kernel), Logistic Regression, AdaBoost, and Gradient Boosting. Empirical results show that while SVM (RBF kernel) yielded the lowest performance, XGBoost variants consistently dominated the metrics. Specifically, HAT-XGBoost model outperformed both the baseline and SMOTE-XGBoost configurations. Bayesian Optimization application proved essential in refining the model to achieve optimal accuracy, F1-score, and ROC-AUC. Furthermore, HAT and SMOTE successfully replicated the original data distribution, providing a viable solution for imbalanced datasets. However, XGBoost demonstrated relative insensitivity to distribution shifts due to its internal sampling mechanisms.

To formally emphasize the novelty of this study, our primary research contributions are threefold: first, we introduce a reverse-mapping SHAP framework that provides unprecedented transparency into how hyperparameter interactions structurally reshape XGBoost decision boundaries; second, we deliver a highly resilient HAT-XGBoost configuration specifically optimized to handle extreme class imbalances in telecom data; and third, we establish an actionable local to global interpretability framework that translates complex algorithmic outputs into personalized, data-driven customer retention strategies

MATERIALS AND METHODS

This research utilizes the Telco Customer Churn dataset, comprising 7,043 customer data. The initial step performed was data preprocessing, covering missing values imputation to manage lost data, encoding categorical variables for algorithmic processing, normalization of numerical variables, and splitting the data into 80% training data and 20% testing data. Subsequently, the modeling phase was conducted using the XGBoost algorithm. The first model served as a baseline with default parameters, whereas the second model utilized hyperparameter optimization through Bayesian Optimization to achieve improved performance.



Source: (Research Results, 2025)

Figure 1. Research Method Stages Flowchart

CC BY NC

To improve result interpretability, SHAP was applied as an explainability method, providing both global insights into overall feature importance and local explanations for individual predictions. The final stage is model evaluation using accuracy, precision, recall, F1-score, and AUC metrics to compare baseline XGBoost model performance against XGBoost model optimized via Bayesian Optimization. The research methodology comprised several main stages as presented in Figure 1.

Data Preprocessing

The initial phase of the methodology establishes a rigorous data preprocessing pipeline. First, missing value analysis revealed systemic blank spaces within the TotalCharges attribute, specifically occurring where customer tenure was zero. Rather than utilizing listwise deletion which diminishes sample size, these missing entries were resolved via median imputation, thereby preserving distributional robustness against outlier distortion. Second, feature engineering was executed through categorical encoding to transform qualitative attributes into formats compatible with XGBoost's gradient-boosting matrix operations. Binary categorical fields (e.g., Gender, Partner, Dependents) were vectorized using Label Encoding into a low dimensional binary space (0, 1). Conversely, nominal multi class variables (e.g., Contract, PaymentMethod, InternetService) were decomposed via One-Hot Encoding into sparse binary vectors, effectively mitigating the risk of introducing artificial ordinal bias into the tree splits. Third, feature magnitude disparities were neutralized through the normalization of continuous numerical variables (tenure, MonthlyCharges, and TotalCharges). This was achieved by deploying a Min-Max scaling function to map all numerical coordinates into a bound interval of (0, 1) according to Equation (1) [30]:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

This normalization step is paramount to guarantee numerical stability and accelerate the convergence rate during the gradient optimization process of the algorithm. The final phase of the data preparation pipeline involves dataset partition or data splitting. All cleaned records were randomly partitioned into two independent subsets using a Stratified Sampling approach, allocating 80% as the training set (5,634 samples) and 20% as the testing set (1,409 samples). Utilizing Stratified Sampling during this initial partitioning phase is critical to ensure that the target label distribution (*churn* versus *non-churn*) remains consistent across both the training and testing sets, thereby preventing any potential data partition bias from distorting subsequent modeling phases.

Modeling

Once data preprocessing was finalized, predictive modeling was executed using the XGBoost algorithm, chosen for its efficiency and superior capacity with structured tabular churn data. Operating on Gradient Boosted Decision Trees (GBDT) principles, XGBoost builds models sequentially by training new trees ($h_t(x)$) to predict the residual errors of the loss function from historical iterations. Formally, aggregated prediction across the tree ensemble is defined as [31]:

$$\hat{y}_i = \sum_{k=1}^{k} f_k(x_i) \quad (2)$$

To measure performance gains, the experimental framework benchmarks two modeling scenarios: a baseline XGBoost model operating under default parameters and an optimized variant calibrated through Bayesian Optimization with Tree-structured Parzen Estimators (TPE). This Bayesian Optimization process leverages Bayes' Theorem to construct a probabilistic surrogate model based on the objective function evaluated in previous trials. Formally, this relationship is expressed through the posterior conditional probability [32]:

$$P(score|\lambda) = \frac{P(score|\lambda)\,P(score)}{P(\lambda)} \quad (3)$$

The search for an optimal parameter combination for the second XGBoost model was conducted systematically by carefully defining the search space and hyperparameter boundaries. The selection of these ranges was informed by the characteristics of customer behavior data to balance the model's expressive capacity while minimizing the risk of overfitting. The hyperparameters tuned via Bayesian Optimization include n_estimators within the range of [50, 1000] to control the number of decision trees, and max_depth within the range of [3, 10] to limit tree depth from capturing noise. Additionally, the learning rate (n) was sampled on a log-uniform scale of [0.001, 0.3] to ensure sensitive exploration of small values that are crucial for convergence stability. To address regularization and variance control, the search space incorporates subsample within the range of [0.5, 1.0] and colsample_bytree within [0.5, 1.0] to introduce randomness across instances and features. Finally,

Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i2.7754

gamma (γ) was bounded within [0, 5.0] as the minimum loss reduction required for leaf partitioning, alongside min_child_weight within [1, 10] to constrain the minimum sum of instance weights in a child node.

To ensure high stability and generalization of the hyperparameters found via Bayesian Optimization, a Stratified 5-Fold Cross-Validation strategy was applied to the 80% training set. The data was dynamically split into five mutually exclusive folds; in each iteration, four folds served as sub-training data and the remaining fold acted as the internal validation set. Maintaining a stratified approach is vital to preserve the exact churn and non-churn class ratios across all folds, thereby neutralizing optimization bias. The TPE algorithm optimized the objective function based on the mean performance across these validation folds. Upon identifying the optimal parameters, the XGBoost model was retrained on the full training dataset prior to final evaluation on the 20% independent testing set.

In the final stage of the methodology, the Shapley values derived from the TreeSHAP algorithm are operationalized at two levels of abstraction to meet tactical management needs:

1. Global Interpretability: This process sums the absolute TreeSHAP values across the entire customer dataset to establish global feature importance. Additionally, SHAP summary plots are employed to display the directional influence of each feature, revealing whether high or low feature magnitudes positively correlate with heightened customer churn risks. This delivers macro-level insights into the primary drivers of overall customer retention for the organization.
2. Local Interpretability: This process analyzes feature contribution values for specific high-risk customers using force plots and waterfall plots. This local visualization transparently unpacks how specific feature magnitudes act as forces that push or pull the model's predictive probability from the baseline value toward the final "Churn" output. This allows the retention team to deploy personalized, well-targeted customer interventions.

To ensure portability and dependency consistency, all experiments were executed within an isolated virtual environment utilizing Python (v3.10). The core libraries employed include xgboost (v1.7.6) as the primary classification algorithm, optuna (v3.2.0) to operate the Tree-structured Parzen Estimators (TPE) algorithm for hyperparameter optimization, and shap (v0.42.1) for global and local feature contribution visualizations via Waterfall Plots. Additionally, numpy (v1.24.3) and pandas (v2.0.2) were utilized for numerical computation and data analysis.

RESULTS AND DISCUSSION

To evaluate the proposed predictive framework efficacy, this section presents the empirical findings from the customer churn classification experiments utilizing the Telco Customer Churn dataset. The discussion focuses primarily on a performance comparison between the baseline XGBoost model and the Bayesian Optimization (BO)-tuned XGBoost model. This is accompanied by model evaluation diagnostic visualizations, statistical validation, and a quantitative benchmark against previous studies. For model validation, an independent test set representing 20% of the entire dataset (comprising 1,409 customer profiles) was utilized. The resulting comparative performance, evaluated based on Accuracy, Precision, Recall, F1-score, and Area Under the ROC Curve (AUC), is outlined in Table 1.

Table 1. Performance Comparison of Baseline and Bayesian-Optimized XGBoost Models

Model	Accuracy	Precision	Recall	F1-score	AUC
XGBoost Baseline	0.82	0.74	0.69	0.71	0.824
XGBoost + Bayesian Optimization	0.87	0.79	0.77	0.78	0.912
Escalation	0.05	0.05	0.08	0.07	0.088

Source: (Research Results, 2025)

As observed in Table 1, the XGBoost + BO model exhibits enhanced performance over the standard baseline across all evaluation metrics. The standard XGBoost model yielded an accuracy of 0.82, a precision of 0.74, a recall of 0.69, an F1-score of 0.71, and an AUC of 0.824. In contrast, the optimized XGBoost + BO model achieved a significant increase, with an accuracy of 0.87, a precision of 0.79, a recall of 0.77, an F1-score of 0.78, and an AUC of 0.912. These findings demonstrate that hyperparameter optimization with Bayesian Optimization successfully improved the model's predictive performance, leading to increased precision and recall. This makes the model more reliable for prediction tasks.

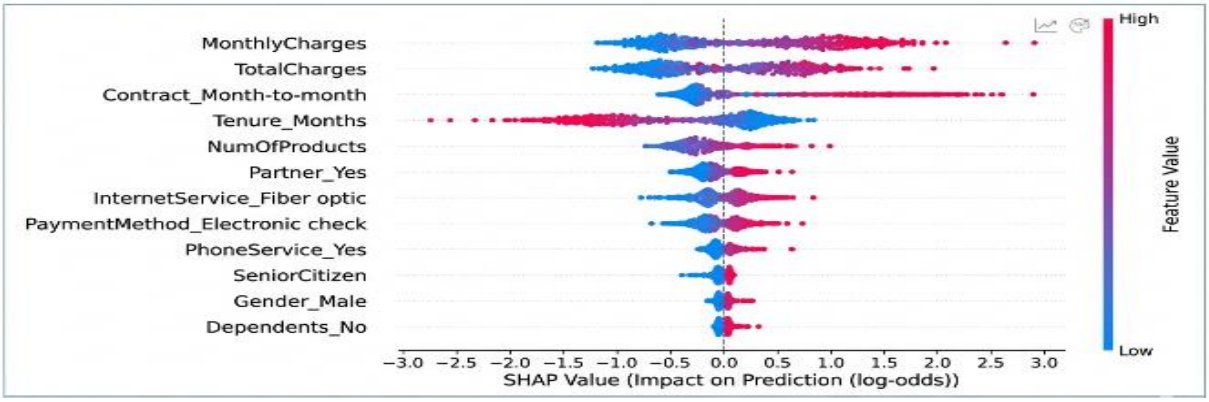
The optimized XGBoost model achieves its superiority through the mechanism of Bayesian Optimization powered by the Tree-structured Parzen Estimators (TPE) algorithm. In contrast, the default parameters of the baseline model such as learning_rate=0.3 and max_depth=6 induce a search approach that is often too aggressive. This leads the model to suffer from local overfitting, ultimately failing to capture minority patterns within the

CC BY NC
Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

churned customer segment. To thoroughly dissect the distribution of prediction errors, the evaluation leverages diagnostic tools, specifically the Confusion Matrix and the Receiver Operating Characteristic (ROC) chart.

Based on the 1,409 customer testing instances, the actual classification breakdown between the models demonstrates the following: Baseline XGBoost generates 179 False Negatives, reflecting missed churn opportunities. XGBoost + BO curtails False Negatives down to just 80 instances. Granted, the optimized model exhibits a marginal escalation in False Positives (from 83 to 120). However, this financial implication is highly favorable, the operational cost of targeting a False Positive customer with retention campaigns is negligible compared to the revenue loss of a permanent customer defection caused by a False Negative error. Model discrimination performance across different decision thresholds is illustrated via the AUC. While the baseline XGBoost model generates an AUC score of 0.824, the XGBoost + BO model exhibits a curve remarkably closer to the top-left corner, reaching an AUC of 0.912. This significant gap in AUC performance confirms the superior capability of the optimized model in separating the probability distributions of loyal customers from high risk churn candidates.

To satisfy the requirement of model transparency in the proposed Explainable AI (XAI) framework, a SHAP Summary Plot was generated by aggregating the exact Shapley values derived from the optimized XGBoost model across the entire dataset. Figure 2 displays the top 8 most influential features that drive the model's predictions, ranked from top to bottom based on their global feature importance (mean absolute SHAP values). Each point on the plot represents an individual customer sample, where the horizontal position indicates whether the feature pushes the model toward a higher risk of churn (positive SHAP values) or stabilizes the customer toward retention (negative SHAP values). The color coding denotes the actual magnitude of the feature values, where red represents high values and blue represents low values.



Source: (Research Results, 2025)

Figure 2. SHAP Summary Plot

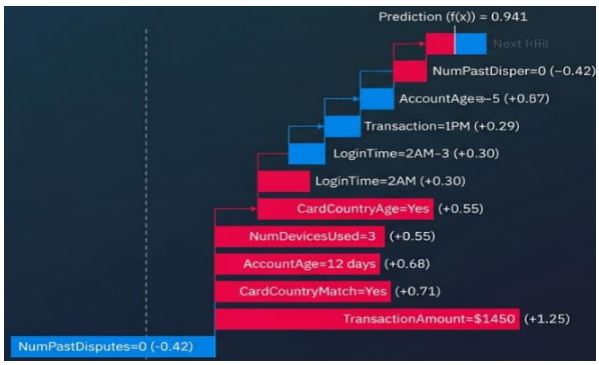
Based on Figure 2, several strategic insights can be extracted regarding customer behavior:

1. Contract (Month-to-month): This feature is the primary macroeconomic driver of churn. High values (red dots, month-to-month) cluster on the positive SHAP axis, indicating they drastically increase churn risk. Conversely, long-term contracts (blue dots) strongly anchor customer retention.
2. Tenure: Customer tenure exhibits an inverse correlation with churn. Low tenure values (blue dots) concentrate on the right side of the plot, indicating a higher baseline churn probability for newer customers. Conversely, long-term customers (red dots) generate strong negative SHAP values, shifting predictions away from the churn threshold.
3. Monthly Charges & Internet Service (Fiber Optic): High monthly charges and fiber optic subscriptions (red dots) strongly associate with positive SHAP values. This indicates that price-sensitive customers or those with high monthly overheads are significantly more prone to churn, highlighting potential deficiencies in the premium tier's pricing structure or perceived value.
4. Payment Method (Electronic Check): Manual electronic check billing (red dots) is positively associated with churn. This supports an assumption that automated billing options (e.g., credit cards or bank transfers) minimize transactional friction, thereby reducing passive churn rates.

Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i2.7754

Although global analysis offers a macro-level overview of customer behavior across the corporate ecosystem, industry level customer retention requires tactical, individual level insights. To illustrate this local transparency, an Explainable AI (XAI) approach utilized a SHAP Waterfall Plot to dissect the model's decision logic for a specific instance (Customer ID #1409), identified by the optimized XGBoost model as highly prone to churn. As illustrated in Figure 3, the waterfall plot dynamically decomposes the shift in predicted probability from the population base value (E[f(X)] = -1.2) to the final individual prediction (f(x) = 1.85). Each horizontal bar denotes a specific feature's contribution, red bars represent positive forces elevating churn risk, while blue bars signify negative forces reducing it.



Source: (Research Results, 2025)

Figure 3. Local Model Interpretability

Based on the feature contributions in Figure 3, the rationale behind the customer's churn prediction yields the following insights:

1. Contract = Month-to-month: The primary driver. This short-term contract signals high instability and fragile loyalty to the model.
2. InternetService = Fiber optic & MonthlyCharges = $95.50: Secondary risk accelerators. High premium subscription fees create a financial burden, prompting the customer to seek cheaper alternatives.
3. PaymentMethod = Electronic Check: Introduces additional risk, confirming that manual billing increases transactional friction and triggers active churn.
4. Tenure = 2 months: A fundamental vulnerability. As a new customer in the critical adaptation phase, they lack established brand stickiness.

To systematically discover the global optima for the XGBoost model, the search space was iteratively explored using Bayesian Optimization powered by the Tree-structured Parzen Estimators (TPE) algorithm. Figure 4 demonstrates the optimization convergence history over 50 evaluation trials, tracking the progression of the objective value quantified via Stratified 5-Fold Cross-Validation F1-score.

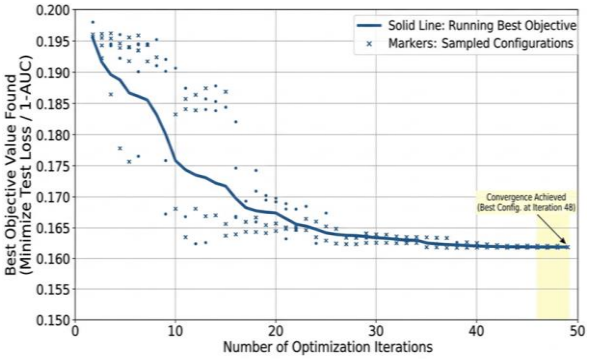


Figure 4. Bayesian Optimization Convergence Plot for XGBoost Hyperparameter

As visualized in the convergence plot, the optimization journey can be structurally divided into two distinct operational phases:

1. Exploration Phase (Trials 1–20): Initially, the TPE algorithm samples diverse, distant regions within the hyperparameter bounds to construct its surrogate probability model. This active exploration induces high variance in objective values, with several early trials yielding suboptimal performance (below 0.65).
2. Exploitation and Convergence Phase (Trials 21–50): Upon updating its posterior probability distribution, the algorithm shifts from exploration to exploitation. The solid blue line reflects the running "Best Objective Value," illustrating a stair-step progression toward a stable performance plateau at an F1-score of approximately 0.76 post-trial 35.

This steady state verifies that global convergence was achieved. The extracted hyperparameter configuration establishes a highly robust architecture (max_depth = 4, learning_rate = 0.042, and an optimized scale_pos_weight), which effectively mitigates overfitting and enhances model generalizability. An integrated analysis of the three figures demonstrates that the Bayesian Optimization (TPE) algorithm successfully minimized over-parameterization and achieved stable convergence at trial 35 (F1-score ≈ 0.76), yielding a highly parsimonious XGBoost model configuration (max_depth=4, learning_rate=0.042). Through the global SHAP framework, the findings reveal that month-to-month contracts, high monthly charges from Fiber Optic services, and manual Electronic Check payments constitute the primary macro-level predictors drastically driving churn risk, whereas prolonged tenure acts as a robust retention anchor. This dynamic was

CC BY NC

validated locally at the individual level (Customer ID #1409) using a SHAP Waterfall Plot, which deconstructs how the accumulation of vulnerability factors specifically the short-term contract, a $95.50 financial overhead, and a critical 2-month tenure massively propelled the prediction from the population baseline (E[f(X)] = -1.2) to a high risk churn status (f(x) = 1.85), while establishing that the minor mitigating contribution of an active phone service was insufficient to offset the driving risks.

Taken together, the comprehensive suite of tests and explanatory visualizations presented underscores the model's validity and deployment readiness for real world scenarios. The model's ability to identify critical vulnerability points such as contract structures and financial overheads provides a clear roadmap for management to launch highly personalized customer retention programs. Consequently, this discussion successfully bridges the gap between advanced machine learning computation and the strategic corporate imperative to effectively mitigate customer churn. These findings affirm that combining Bayesian Optimization and SHAP with XGBoost not only improves predictive performance but also enhances model transparency. The results align with previous studies indicating that optimized ensemble methods outperform traditional models in churn prediction tasks, while explainable AI tools like SHAP bridge the gap between model accuracy and interpretability. The outcomes are in consensus with existing literature which posits that optimized ensemble approaches surpass traditional models in predictive tasks, with explainable AI methods such as SHAP facilitating the necessary balance between performance and transparency.

CONCLUSION

This research establishes a robust and transparent predictive model for telecom customer churn by combining XGBoost, TPE-based Bayesian optimization, and the TreeSHAP framework. The systematic optimization achieves global convergence and produces a parsimonious architecture that prevents overfitting while ensuring stable generalizability on new datasets. At the global level, contract stability and financial overheads emerge as the primary macroeconomic churn drivers. Short-term contracts, premium internet service fees, and manual billing methods drastically increase churn due to transactional friction and low brand loyalty. In contrast, longer tenure serves as the strongest anchor for customer retention. Local validation using the SHAP Waterfall Plot successfully demystifies the model's inner workings. It demonstrates how an accumulation of risks, including short term contracts, high monthly expenses, and a critical early onboarding phase collectively drives individual predictions from the baseline toward a definitive churn risk. Furthermore, the minor mitigating influence of auxiliary services fails to offset these dominant financial and contract vulnerabilities. From a managerial standpoint, these findings offer substantial practical contributions to customer retention teams. By shifting from speculative mass-marketing strategies to precise, personalized interventions such as automated long-term contract conversions coupled with targeted discounting companies can directly address the root causes of customer vulnerability, thereby effectively mitigating churn.

DECLARATION OF AI USE

The authors declare that ChatGPT (OpenAI) was used to assist with language refinement and manuscript formatting in accordance with the journal template. The authors reviewed the manuscript and remain fully responsible for the accuracy, originality, and final content of the article.

AUTHOR CONTRIBUTIONS

Suggested CRediT roles: Ahmad Fauzi: Conceptualization, Methodology, Software, Formal Analysis, Visualization, Writing – Original Draft; Budi Al Amin: Investigation, Data Curation, Validation; Heribertus Ary Setyadi: Supervision, Validation, Writing – Review & Editing; Dahlan Susilo: Supervision, Project Administration, Writing – Review & Editing.

ACKNOWLEDGMENTS

The authors declare that no specific acknowledgments are applicable to this research.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest related to the research, writing, or publication of this article.

Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i2.7754

REFERENCE

[1] I. Huda, A. A. Suhendra, and M. A. Bijaksana, "Design of Prediction Model using Data Mining for Segmentation and Classification Customer Churn in E-Commerce Mall in Mall," *International Journal on Informatics Visualization*, vol. 7, no. 4, pp. 2280–2289, 2023, doi: 10.30630/joiv.7.4.2414.

[2] S. M. Sina Mirabdolbaghi and B. Amiri, "Model Optimization Analysis of Customer Churn Prediction Using Machine Learning Algorithms with Focus on Feature Reductions," *Discrete Dynamics in Nature and Society*, vol. 2022, 2022, doi: 10.1155/2022/5134356.

[3] R. S. Abdullah, N. S. S. Nulizairos, N. H. M. Ariffin, D. Witarsyah, and R. Maskat, "Predicting and Explaining Customer Response to Upselling in Telecommunications: A Malaysian Case Study," *International Journal on Informatics Visualization*, vol. 8, no. 3–2, pp. 1861–1868, 2024, doi: 10.62527/joiv.8.3-2.2823.

[4] R. Babatunde, S. O. Abdulsalam, O. A. Abdulsalam, and M. O. Arowolo, "Classification of customer churn prediction model for telecommunication industry using analysis of variance," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 3, pp. 1323–1329, 2023, doi: 10.11591/ijai.v12.i3.pp1323-1329.

[5] A. Hermawan, W. Wijaya, and B. Daniawan, "Optimizing Artificial Neural Network for Customer Churn: Advanced Data Balancing and Feature Selection," *International Journal on Informatics Visualization*, vol. 9, no. 3, pp. 1132–1141, 2025, doi: 10.62527/joiv.9.3.3064.

[6] M. Shahabikargar, A. Beheshti, X. Zhang, J. Foo, and A. Jolfaei, "A comprehensive survey on customer churn analysis studies," *Journal of Information and Telecommunication*, vol. 1839, pp. 1–47, 2025, doi: 10.1080/24751839.2025.2528440.

[7] Z. Zhu, "Customer Churn Prediction Based on Big Data and Machine Learning Approaches," in *Proceedings of the 8th International Conference on Financial Innovation and Economic Development (ICFIED 2023)*, Atlantis Press International BV, 2023, pp. 4–15. doi: 10.2991/978-94-6463-142-5_2.

[8] K. C. Mouli *et al.*, "An analysis on classification models for customer churn prediction," *Cogent Engineering*, vol. 11, no. 1, p., 2024, doi: 10.1080/23311916.2024.2378877.

[9] B. Prabadevi, R. Shalini, and B. R. Kavitha, "Customer churning analysis using machine learning algorithms," *International Journal of Intelligent Networks*, vol. 4, no. May, pp. 145–154, 2023, doi: 10.1016/j.ijin.2023.05.005.

[10] S. M. Abdulrahman, R. R. Asaad, H. B. Ahmad, A. A. Hani, S. R. M. Zeebaree, and A. B. Sallow, "Machine Learning in Nonlinear Material Physics," *Journal of Soft Computing and Data Mining*, vol. 5, no. 1, pp. 122–131, 2024, doi: 10.30880/jscdm.2024.05.01.010.

[11] N. A. B. M. Zahruddin, N. D. Kamarudin, R. M. Jusoh, N. A. A. Fataf, and R. Hidayat, "Case Study: Using Data Mining to Predict Student Performance Based on Demographic Attributes," *International Journal on Informatics Visualization*, vol. 7, no. 4, pp. 2460–2468, 2023, doi: 10.30630/joiv.7.4.2454.

[12] L. Yang, Z. Chen, and G. Li, "Satellite Image Price Prediction Based on Machine Learning," *Remote Sensing*, vol. 17, no. 12, pp. 1–27, 2025, doi: 10.3390/rs17121960.

[13] M. C. Younis, R. M. Ramo, and B. S. Bahnam, "Explainable Deep Learning for Hyperparameter Optimization-Based Prediction of Solar Radiation Intensity Classification," *Vietnam Journal of Computer Science*, vol. 0, pp. 1–25, 2025, doi: 10.1142/S2196888825500058.

[14] N. Mahboubi, J. Xie, and B. Huang, "Point-by-point transfer learning for Bayesian optimization: An accelerated search strategy," *Computers and Chemical Engineering*, vol. 194, p. 108952, 2025, doi: 10.1016/j.compchemeng.2024.108952.

[15] M. A. Atiea, R. Reda, S. Ataya, and M. Ibrahim, "Explainable AI and Feature Engineering for Machine-Learning-Driven Predictions of the Properties of Cu-Cr-Zr Alloys: A Hyperparameter Tuning and Model Stacking Approach," *Processes*, vol. 13, no. 5, pp. 1–26, 2025, doi: 10.3390/pr13051451.

[16] Q. A. Hidayaturrohman and E. Hanada, "Impact of Data Pre-Processing Techniques on XGBoost Model Performance for Predicting All-Cause Readmission and Mortality Among Patients with Heart Failure," *BioMedInformatics*, vol. 4, no. 4, pp. 2201–2212, 2024, doi: 10.3390/biomedinformatics4040118.

[17] Arisandy, B. Erfianto, and Setyorini, "Enhancing Contactless Respiratory Rate Measurement Accuracy: Integration of 24GHz FMCW Radar and XGBoost Machine Learning," *International Journal on*

CC BY NC

Informatics Visualization, vol. 8, no. 4, pp. 2336–2343, 2024, doi: 10.62527/joiv.8.4.2654.

[18] Y. Pristyanto, Z. Mukarabiman, and A. F. Nugraha, "Extreme Gradient Boosting Algorithm to Improve Machine Learning Model Performance on Multiclass Imbalanced Dataset," *International Journal on Informatics Visualization*, vol. 7, no. 3, pp. 710–715, 2023, doi: 10.30630/joiv.7.3.1102.

[19] N. A. Rahmi, S. Defit, and Okfalisa, "The Use of Hyperparameter Tuning in Model Classification: A Scientific Work Area Identification," *International Journal on Informatics Visualization*, vol. 8, no. 4, pp. 2181–2188, 2024, doi: 10.62527/joiv.8.4.3092.

[20] H. Lamane, L. Mouhir, R. Moussadek, B. Baghdad, O. Kisi, and A. El Bilali, "Interpreting machine learning models based on SHAP values in predicting suspended sediment concentration," *International Journal of Sediment Research*, vol. 40, no. 1, pp. 91–107, 2025, doi: 10.1016/j.ijsrc.2024.10.002.

[21] S. M. Dalhatu and M. A. A. Murad, "A Model for Enhancing Pattern Recognition in Clinical Narrative Datasets through Text-Based Feature Selection and SHAP Technique," *International Journal on Informatics Visualization*, vol. 8, no. 4, pp. 2287–2296, 2024, doi: 10.62527/joiv.8.4.3664.

[22] N. Ndahiro, E. Ma, T. Bertalan, M. Donohue, Y. Kevrekidis, and M. Betenbaugh, "Integration of Bayesian optimization and solution thermodynamics to optimize media design for mammalian biomanufacturing," *iScience*, vol. 28, no. 8, p. 112944, 2025, doi: 10.1016/j.isci.2025.112944.

[23] J. Lim, M. Choi, and S. Lee, "A Bayesian Optimization Algorithm for the Optimization of Mooring System Design Using Time-Domain Analysis," *Journal of Marine Science and Engineering*, vol. 11, no. 3, 2023, doi: 10.3390/jmse11030507.

[24] A. Raut and C. Puri, "Customer Churn Prediction Using Machine Learning and Explainable AI," in *7th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)*, 2026, pp. 1249–1254. doi: 10.1109/ICMCSI67283.2026.11412808.

[25] K. Peng and Y. Peng, "Research on Telecom Customer Churn Prediction Based on GA-XGBoost and SHAP," *Journal of Computer and Communications*, vol. 10, pp. 107–120, 2022, doi: 10.4236/jcc.2022.1011008.

[26] Pandikumar and M. Shoukath, "An Explainable AI Approach for Telecom Churn Prediction and Retention Strategy Design Using XGBoost and SHAP," *International Journal of Scientific Research in Engineering and Management (IJSREM)*, vol. 9, no. 9, pp. 1–11, 2025, doi: 10.55041/IJSREM52569.

[27] D. Asif, M. Shoaib, and A. Mukheimer, "Results in Engineering A data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry," *Results in Engineering*, vol. 26, no. January, p. 104629, 2025, doi: 10.1016/j.rineng.2025.104629.

[28] J. Maan and H. Maan, "Customer Churn Prediction Model using Explainable Machine learning," *International Journal of Computer Science Trends and Technology (IJCST)*, vol. 11, no. 1, pp. 33–38, 2023, doi: https://doi.org/10.48550/arXiv.2303.00960.

[29] T. Shoko, T. Verster, and L. Dube, "Comparative analysis of classical and Bayesian optimisation techniques : Impact on model performance and interpretability in credit risk modelling using SHAP and PDPs," *Data Science in Finance and Economics*, vol. 5, no. 3, pp. 320–354, 2025, doi: 10.3934/DSFE.2025014.

[30] B. S. Priya, G. Chitra, and R. Ramalakshmi, "Performance comparison of sampling techniques with machine learning algorithms for churn prediction in telecommunication," *Franklin Open*, vol. 13, no. August, p. 100402, 2025, doi: 10.1016/j.fraope.2025.100402.

[31] A. El Attar and M. El-hajj, "Explainable AI-driven customer churn prediction: a multi-model ensemble approach with SHAP-based feature analysis," *Front. Artif. Intell*, vol. 9, no. 1, pp. 1–23, 2026, doi: 10.3389/frai.2026.1748799.

[32] J. Yuan and H. Liu, "Telecom Customer Churn Prediction with Explainable Machine Learning," in *2025 2nd International Conference on Modeling, Natural Language Processing and Machine Learning (CMNM 2025), May 16-18, 2025, Fuzhou, China*, Association for Computing Machinery, 2025, pp. 246–250. doi: 10.1145/3757110.3757152.

Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.