JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9576
Hal 1541-1551
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Comparison of XGBoost and SVM for Predicting User Behaviour Segment Shifts of QRIS Merchants Based on Dynamic RFM with SHAP Interpretation

Zulham Sitorus*, Hanna Willa Dhany, Ade Iskandar

Postgraduate, Master of Information Technology, Universitas Pembangunan Panca Budi, Indonesia
Email: [1*]zulhamsitorus@dosen.pancabudi.ac.id, [2]hdhany@dosen.pancabudi.ac.id, [3,]suratiskandar@gmail.com
Email Penulis Korespondensi: zulhamsitorus@dosen.pancabudi.ac.id*
Submitted **25-08-2026**; Accepted **30-10-2026**; Published **31-10-2026**

Abstract

With the development of QRIS, transaction traces that record user behaviour can be maintained, but retail merchants rarely utilise them to anticipate changing customer behaviour. In predicting the shift in user behaviour segments of QRIS at PMart, this study compares Extreme Gradient Boosting (XGBoost) and Support Vector Machine (SVM). Segments were formed using K-Means from RFM (Recency, Frequency, and Monetary), which were recalculated every month with a three-month window. The status of segment movement for the following month is up, stable, or down, and is predicted using dynamic RFM features and their derivatives. In the simulation data, which contains 66,158 transactions from 1,168 users, both models were validated temporally, tuned with Optuna, tested with the McNemar test and bootstrap, and explained with SHAP. Three active segments were formed, and only 63.3% of observations remained in their segments from month to month. XGBoost had a lower error ($p < 0.001$), but the macro-F1 difference was not significant. SVM achieved a macro-F1 of 0.612 and an MCC of 0.488. SVM is more sensitive to high-value users who are moving down the segment (recall 76.3% versus 60.7%). The explanations of both models are consistent; the main driving factors for segment migration are frequency, shopping trends, and frequency (Spearman $\rho = 0.868$). This framework can be used as a basis for an early warning system for customer retention in QRIS stores.

Keywords: QRIS; Dynamic RFM; Segment Transfer; XGBoost; Support Vector Machine; SHAP

1. INTRODUCTION

The widespread adoption of the Quick Response Code Indonesian Standard (QRIS) in the past five years has marked a significant transformation in Indonesia's payment system. QRIS was established by Bank Indonesia as the national standard for QR payment codes that can be used together to receive payments from various bank applications and digital wallets [1]. This standardisation allows cashless payments to be used by retail stores, including small and medium-sized businesses. The development of QRIS is very rapid. As many as 96.68% of QRIS sellers are MSMEs. Throughout 2025, the volume of QRIS transactions reached 15.51 billion transactions, growing 148.54% year-on-year, with a value of Rp1,420.66 trillion [2]. QRIS merchants reached 49.4 million, and QRIS users reached 70.6 million [3].

This development shows that QRIS payments have become a part of people's daily lives. Social influence and supporting conditions, as well as habits, beliefs, and perceived risks, are some of the reasons why consumers choose QRIS [4]. However, this research does not focus on user transaction patterns after regular use of QRIS at a merchant; instead, it focuses on decisions made with QRIS. Each QRIS transaction conducted by the seller leaves an automatically recorded digital record, including the date and amount of payment, as well as the masked identity of the payer's account. Digital payments accelerate transactions, reduce recording errors, and provide data for data-driven loyalty programs, [5]. Transaction trails should be information about customer behaviour rather than sales records.

Data cannot store user behaviour. Because the values sought by customers change, segment membership becomes unstable, [6] considering customer segments as "moving targets." [7] empirically shows that customer purchasing patterns change over time, which means marketing strategies must adapt to segment shifts. Habits are the main factor determining the sustainability of use in QRIS [8], so changes in shopping habits will directly impact the value and frequency of transactions. One of the minimarket stores, PMart, located within the campus of Pembangunan Panca Budi University, serves QRIS payments. From 2025 to 2026, there was an increase in the number of QRIS transactions from the number of students who attend school at 6:00 PM. This data is an analytical asset that has not been properly utilised. The Recency, Frequency, and Monetary (RFM) model is used to describe customer transaction behaviour by using the recency of the last transaction, transaction frequency, and total spending value [9] [10], these three aspects represent adequate statistics to estimate the future value of customers. Because the three variables can be directly calculated from transaction logs without including demographic data, RFM is very suitable for QRIS data.

However, RFM is usually static. According to research conducted [11] on 105 publications, RFM has dominance in customer representation, and K-Means has dominance in segmentation, which is usually calculated within a single time window. The same pattern was also found in recent studies, as shown in [12] [13]. This method produces a very brief snapshot, making it impossible to observe customer movement between segments. [14] states that changes in the RFM time window and changes in customer value are common shortcomings in RFM research. Segment shifts can be seen as events such as moving up to a higher-value segment, remaining, or moving down to a riskier segment. [7] simulates inter-segment shifts of customers over time, [15] shows that the likelihood of inter-segment shifts contributes to customer value

Copyright © 2026 The **Zulham Sitorus**, Page 1541
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9576
Hal 1541-1551
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

calculation. With this formulation, the research question is changed from "to which segment will this user move in the next period?" to "which segment does this user belong to?" Because the second question is predictive in nature, a supervised classification model is needed.

The prediction of segment migration is formulated as classification. The status of segment migration is predicted using dynamic RFM features and their derivatives in the t+1 period. The selection of two algorithms is because each represents a different learning paradigm. To address overfitting, a group of decision trees based on gradient boosting called Extreme Gradient Boosting (XGBoost) is used. By using a kernel function, Support Vector Machine (SVM) can find a separating hyperplane with a maximum margin and capture nonlinear decision boundaries [16]. In general, tree-based models tend to excel on tabular data, and a review [17] of 112 studies concluded that ensemble methods consistently outperform conventional classifiers in churn prediction.

Decision-makers at PMart are not satisfied with the prediction accuracy. A user is predicted to churn because XGBoost and SVM with a nonlinear kernel are black box models. Predictions are difficult to translate into appropriate retention actions without explanation. [18], interpretability is one of the main issues in research on customer behaviour prediction. The Shapley value from cooperative game theory is used to calculate the contribution of each feature to individual predictions. SHAPley Additive exPlanations (SHAP) does this [19]. TreeSHAP calculates the values accurately and generates both global and local explanations for XGBoost [20]. Because KernelSHAP is model-agnostic, it is used for SVM. Therefore, both models can be compared in terms of prediction performance and explanation consistency.

Previous studies have shown that SHAP has the ability to detect actionable behavioural factors. [20] [21] found that the main drivers of churn are the number of transactions, transaction value, and their changes over periods; [22] combined RFM segmentation, XGBoost, and SHAP, and found that the drivers of churn actually differ among customer value groups. SHAP hopes to explain every type of segment shift among QRIS users at PMart, for example, whether the decline is caused by lower frequency, lower frequency, or lower shopping value. Theoretically, there are three contributions made by this research. First, this research develops QRIS from the domain of intention and attitude to the domain of actual transaction behaviour. Second, this research develops RFM from a descriptive segmentation tool to a basis for predicting temporal segment shifts. Third, this research adds to digital payment data with empirical evidence comparing XGBoost and SVM and the consistency of SHAP explanations for both.

2. RESEARCH METHODOLOGY

For this research, a longitudinal secondary data-based comparative-predictive design was used. The unit of observation is the user-period (month) pair, while the unit of analysis is the QRIS user account that transacts at PMart. The target variable is the status of segment transition for the following month, whether it is up, stable, or down, and the predictor variable is the dynamic RFM features along with their derivatives. Two algorithms, XGBoost and SVM, were compared fairly. Both have the same features, data splitting, and hyperparameter tuning budget, so it is not the data treatment that is responsible for the performance difference, but the algorithm that is responsible for the performance difference. SHAP is then used to explain the prediction results.

2.1 Research Framework

The research framework is structured following the Cross-Industry Standard Process for Data Mining (CRISP-DM), which is the most widely adopted data science process model and consists of six phases [23]. The mapping of the research stages to the six phases is presented in Figure 1. The CRISP-DM phases are mapped from the stages shown in Figure 1. The business understanding phase formulates the problem, such as PMart's inability to anticipate that QRIS users would reduce shopping activity. The data understanding phase includes collecting QRIS transaction data and studying its characteristics. Preprocessing, dynamic RFM creation and user segmentation, segment shift labelling, and training and test data splitting are all phases of data preparation. XGBoost and SVM are trained in parallel with the same data and schema during the modelling phase. Evaluation and model comparison tests, as well as SHAP interpretation, are included in the evaluation phase. The deployment phase produces the best model and retention suggestions for PMart. When the evaluation shows that the definition of segments, labels, or features needs improvement, the process returns to the data preparation stage, in accordance with the iterative nature of CRISP-DM.

The research methodology used in this study is shown in Figure 1. The procedure starts with gathering QRIS transaction data, which is subsequently preprocessed to manage missing values, eliminate duplicates, and standardise data formats. The users are then divided based on their transaction behaviour using a dynamic RFM (Recency, Frequency, Monetary) analysis over a series of time frames. The goal variable is segment transitions, which are defined as changes in each user's segment between periods. Two classification algorithms, XGBoost and Support Vector Machine (SVM), are trained in parallel after the labelled dataset is split into training and testing sets. To ascertain whether the differences between the two models are statistically significant, they are compared using standard performance indicators. The contribution of

Copyright © 2026 The **Zulham Sitorus**, Page 1542
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9576
Hal 1541-1551
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

each feature to the model's predictions is then interpreted using SHAP (SHapley Additive exPlanations). Ultimately, the model that performs the best is chosen, and useful suggestions are developed based on the results.

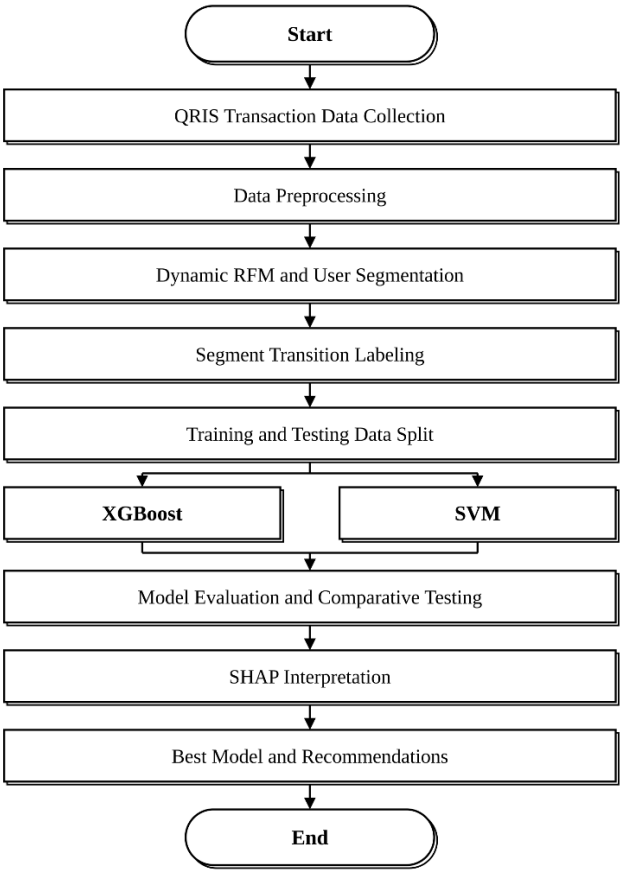


Figure 1. Research Framework

2.2 Data Science

Data science is a field that encompasses many disciplines using computation, statistics, and domain knowledge to obtain information that aids decision-making [24] [25], data science covers the entire data work cycle, from data collection and preparation, exploration, representation and transformation, computation, modelling, to visualisation and presentation of results. Thus, data science is not just about creating models; it also includes how data is transformed into relevant representations. Data science is used on QRIS transaction data. Transaction logs, which originally served only as sales records, are transformed into representations of user behaviour using dynamic RFM, modelled to predict segment shifts, and then explained with SHAP to investigate the findings. To ensure everything is systematic and repeatable, the CRISP-DM framework is used throughout the process.

2.3 Prediction

Prediction is the process of estimating the value or class of an unobserved event based on patterns learned from historical data. [26] distinguishes predictive modelling from explanatory modelling. Predictive modelling aims to minimise prediction errors on new data, so it must be evaluated with data outside the training sample. This research is predictive in nature: model g is trained to map the behavioural feature vector xi,p of user i in month p to the status of its segment transition in the following month.

$$\hat{y}_{i,p} = g(\mathbf{x}_{i,p}), \quad \hat{y}_{i,p} \in \{\text{Naik,Tetap,Turun}\} \tag{1}$$

The future horizon is the most important component when choosing a forecasting method. The study horizon lasts for one month, from month p to month p + 1, and aligns with the monthly sales and promotion evaluation cycle. The purpose of this horizon is to allow the analysis of prediction results to be conducted before any behavioural changes occur. Predictions are made as supervised multiclass classifications because the output is in the form of categories. [26], the SHAP interpretation in this study explains the patterns of the model being studied rather than testing cause-and-effect relationships.

This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9576
Hal 1541-1551
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

2.4 Data Preprocessing

The research data consists of QRIS PMart transaction logs from a minimarket. The data was obtained from 2025-2026. The study uses the entire population of transactions during that period without sampling. The attributes used are presented in Table 1.

Table 1. QRIS Transaction Data Dictionary

ATTRIBUTE	DESCRIPTION	TYPE	UTILITY
id_transaksi	Unique transaction reference number	Text	Duplicate detection
user_id	Masked payer account identity	Text (hash)	Per-user aggregation key
transaction time	Transaction date and time	Datetime	Recency, period, time features
nominal	Payment amount (IDR)	Numerical	Monetary
status	Successful, failed, or cancelled	Categorical	Filtering of valid transactions
publisher	The paying application or issuing bank	Categorical	Additional features (optional)

Preprocessing consists of five steps. 1) Filtering: Only transactions with a successful status are used; failed, cancelled, and returned transactions. 2) Deduplication: transactions with the same transaction_id are removed. 3) Anonymisation: SHA-256 hash values are converted into user identities before analysis. 4) Outlier handling: the IQR method is used to check for extreme nominal values; values proven to be incorrect are corrected or removed, while valid extreme values are retained and logarithmically transformed at the feature stage. 5) Inclusion criteria: users are evaluated if they conduct transactions for at least one different month. This is because segment changes only apply to users recorded for more than one period. To ensure data tracking, the number of transactions and users before and after each step is reported.

2.5 Dynamic RFM and User Segmentation

The three dimensions of dynamic RFM are defined as follows: Dynamic RFM is recalculated at the end of each month p with a rolling observation window of L months, specifically from month p − L + 1 to p. For user i with a valid transaction set T (i, p) within that window, dynamic RFM is recalculated at the end of each month p [9], [10].

$$R_{i,p} = \tau_p - \max_{j \in T(i,p)} t_j \tag{2}$$

$$F_{i,p} = |T(i,p)| \tag{3}$$

$$M_{i,p} = \sum_{j \in T(i,p)} m_j \tag{4}$$

In the equation, the end date of month p is τp, the transaction date j is tj, and the nominal transaction j is mj. The value of R is shown in days. The length of the main window is set to L = 3 months. Then L = 1 and L = 6 months were used to test its sensitivity. The transformed RFM vector was segmented using K-Means [27] After the transformation of F and M with log (1 + x), z scores are used to standardise the three dimensions. Each observation xi is minimised by K-Means against the centroid μc of its cluster.

$$J = \sum_{c=1}^{k} \sum_{\mathbf{x}_i \in C_c} \|\mathbf{x}_i - \boldsymbol{\mu}_c\|^2 \tag{5}$$

The number of clusters k is chosen based on the Silhouette value [28] [30], considering business readability. The Silhouette value for observation i is calculated as follows.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{6}$$

In that equation, a(i) is the average distance of observation i to its own cluster members, and b(i) is the average distance to the nearest cluster. The K-Means centroid parameters and standardisation are only calculated from the training period, then applied unchanged during the training period to ensure the segment definitions remain consistent from month to month. Without this step, label changes can only be caused by changes in segment boundaries, not changes in user behaviour. Next, the clusters are arranged according to their centroid values into tiered segments, with S1 as the segment with the highest value down to Sk. Users who do not make transactions within the window (F = 0) are placed into a special segment S0 (Inactive).

2.6 Segment Transfer Labelling and Feature Engineering

Each segment is assigned a value level ℓ, with S0 = 0 as the lowest level. The user transition label i in period p is determined by comparing its segment level in month p and month p + 1.

$$y_{i,p} = \begin{cases} \text{Naik}, & \ell_{i,p+1} > \ell_{i,p} \\ \text{Tetap}, & \ell_{i,p+1} = \ell_{i,p} \\ \text{Turun}, & \ell_{i,p+1} < \ell_{i,p} \end{cases} \tag{7}$$

Copyright © 2026 The **Zulham Sitorus**, Page 1544
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9576
Hal 1541-1551
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

The formulation of these three classes—maintaining, encouraging growth, and preventing decline—was chosen because they are directly related to business decisions. For additional analysis, a monthly transition matrix between segments is provided with a description. This shows a more detailed transition pattern [7]. While the label only uses data from month p + 1, the features are built only from data up to τp. Therefore, no future information will leak [29]. Table 2 shows the feature groups.

Table 2. Predictor Feature Group

Group	Feature	Meaning of behavior
RFM now	R, log F, log M in period p	Current behaviour position
RFM lag	R, F, M in the periods p − 1 and p − 2	Short history
Change	ΔR, ΔF, ΔM between p − 1 and p	Direction and rate of behavioural change
Tren	The slope of the regression of monthly spending values within the window	Tendency to rise or fall
Purchasing patterns	Average and standard deviation of transaction intervals (days); average value per transaction (M/F)	The regularity and intensity of shopping
Involvement	Membership duration (months since the first transaction); active month ratio	Depth of relationship with the merchant
Segment status	Current segment (one-hot); duration in the current segment	Stability of segment position
Time (optional)	Proportion of weekend transactions and per hour range	Shopping habits

Features that are highly correlated (|r| > 0.9) are examined, and one of them is removed to prevent the SHAP interpretation from being divided between nearly identical features.

2.7 Modelling

2.7.1 Data Splitting and Class Imbalance

Not randomly, the data is divided by month. In the early months, around 70–80% of the training data is used, and in the later months, the test data is untouched until the final evaluation. Hyperparameter tuning in the training data uses forward-chaining cross-validation with an incrementally expanding window. In other words, training from month 1 to k and validating in month k + 1 [30]. This method mimics real-world usage, where the model is trained to predict past data. The "Stay" class is predicted the most. For XGBoost, sample_weight = balanced, and for SVM, class_weight = balanced, are class weightings. SMOTE [31] is used only on the training folds as a sensitivity test to ensure that synthetic data does not leak into the test or validation data.

2.7.2 Extreme Gradient Boosting (XGBoost)

XGBoost builds an ensemble of decision trees additively. Each new tree fk is trained to correct the errors of the previous tree by minimising the regularised objective function [32].

$$\mathcal{L} = \sum_{i=1}^{n} l\,(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega\,(f_k), \quad \Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2 \tag{8}$$

In that equation, l is the multi-class softmax loss function, T is the number of leaves, and w is the leaf weight. To prevent overfitting, γ and λ control the complexity of the tree. The model uses two multi-objectives: softprob and early stopping for validation data. Tree-based models do not require feature standardisation.

2.7.3 Support Vector Machine (SVM)

SVM searches for a separating hyperplane with the maximum margin between classes [16]. The decision function is formulated as follows.

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i\, y_i K(\mathbf{x}_i, \mathbf{x}) + b\right) \tag{9}$$

In the equation, αi is the Lagrange multiplier, b is the bias, and K is the kernel function. This study uses the Radial Basis Function (RBF) kernel as the main model and the linear kernel as a comparison.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \tag{10}$$

Multiclass classification is performed using a one-vs-one scheme. Features are standardised using parameters from the training data [35], and class probabilities are obtained through Platt scaling for ROC-AUC and SHAP purposes.

Copyright © 2026 The **Zulham Sitorus**, Page 1545
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9576
Hal 1541-1551
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

2.7.4 SHAP Interpretation

SHAP breaks down the model's predictions into the contribution of each feature ϕj based on the Shapley value from cooperative game theory [19].

$$\phi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!\,(|N|-|S|-1)!}{|N|!} [f_x(S \cup \{j\}) - f_x(S)] \quad (11)$$

In that equation, N is the set of all features and fx(S) is the model output when only the features in S are known. The sum of the contributions of all features plus the baseline value ϕ0 equals the model output for that observation.

$$f(\mathbf{x}) = \phi_0 + \sum_{j=1}^{M} \phi_j \quad (12)$$

SHAP is calculated separately for each class, especially for the "Down" class. XGBoost uses the efficient TreeSHAP calculated for the entire test data, while the model-agnostic Kernel SHAP is used for SVM [33]. Background data from the training data is summarised into 100 K-Means centroids, and explanations are calculated for stratified samples from the test data due to high computational costs. The output interpretation consists of three levels. At the local level, a waterfall plot for the user example predicted as "Down" is used as the basis for action recommendations for each user. At the global level, feature rankings are calculated based on the average absolute SHAP values and the beeswarm plot. At the feature relationship level, a dependence plot for the top five features is used.

In addition to the intersection of the five main features measured by the Jaccard index, the Spearman rank correlation between the global feature importance rankings of XGBoost and SVM is used to test the consistency of explanations across models. Because the output scales of the two models differ, the comparison is made on rankings rather than raw values. The belief that the identified factors indicate user behaviour patterns rather than algorithmic artefacts is reinforced by consistent explanations across models.

3. RESULT AND DISCUSSION

3.1 Results

3.1.1 Data Pre-processing Results

The initial data consists of 68,979 transaction log rows from 1,195 user accounts from January 2025 to October 2026, or over a period of twenty-two months. The table shows the results of each preprocessing step.

Table 3. Results of Data Preprocessing

Step	Number of transactions	Explanation
Initial data	68.979	1,195 users
1) Status filtering successful	66.906	2,073 failed/canceled transactions were discarded.
2) Deduplication of transaction_id	66.248	658 duplicates removed
3) SHA-256 Anonymisation	66.248	User identity replaced with hash
4) Handling isolation	66.214	713 extreme nominal; 34 incorrect inputs discarded
5) Inclusion criteria (≥ 2 months active)	66.158	27 users were removed; 1,168 users remain

The extreme nominal limit is obtained by calculating the nominal quartiles of the transactions. Since Q1 = Rp16,500 and Q3 = Rp37,500, IQR = Rp21,000, the extreme limit is Q3 + 3 × IQR = Rp100,500. Of the 713 transactions that exceeded this limit, only 34 had values more than 20 times the median nominal of the same user. As a result, these are considered erroneous inputs and discarded. Legitimate large purchases that are retained are another extreme transaction. Final data shows 66,158 transactions made by 1,168 users.

3.1.2 Results of Dynamic RFM and User Segmentation

RFM is calculated at the end of each month with a three-month window. To ensure that the lag p−2 feature has a full window, observations begin in June 2025. The course runs from June 2025 to May 2026, totalling twelve months; the test runs from June 2026 to September 2026, totalling four months. K-Means was tested on 8,139 RFM vectors from the training period for k = 2–6, and the results are presented in Table 4.

Table 4. Selection of the Number of Clusters

k	Silhouette	Davies–Bouldin	Inersia
2	0,458	0,821	11.080,9
3	0,433	0,799	6.791,5
4	0,409	0,813	4.785,7
5	0,414	0,831	3.925,5
6	0,387	0,814	3.177,3

Copyright © 2026 The **Zulham Sitorus**, Page 1546
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9576
Hal 1541-1551
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Among $k \geq 3$, $k = 3$ has the highest Silhouette (0.433) and the lowest Davies–Bouldin index of all candidates (0.799), but the two segments are too coarse to show up and down movements. Therefore, $k = 3$ is set. The table displays the centroid profile for each segment.

Table 5. Segment Centroid Profile

Segment	Name	R (day)	F (transaction)	M (Rp)
S1	Aktif Bernilai Tinggi	5,5	25,3	729.908
S2	Reguler	13,4	3,9	95.381
S3	Passive	58,1	1,8	40.281
S0	Inactive	–	0	0

Table 5 has returned the centroid values to the original scale. The values of F and M are almost the same as the geometric mean due to their logarithmic transformation. Users in segment S1 average 25.3 transactions in three months, with the last transaction occurring about 5.5 days before the end of the month. Users in segment S3 only transact 1.8 times with a gap of 58.1 days from the last transaction.

3.1.3 Results of Segment Migration Labelling

For active users (S1–S3) in month p, observations were made for 11,217 monthly pairs from 1,163 users, consisting of 8,139 training observations and 3,078 test observations. Figure 2 shows the transition matrix between segments.

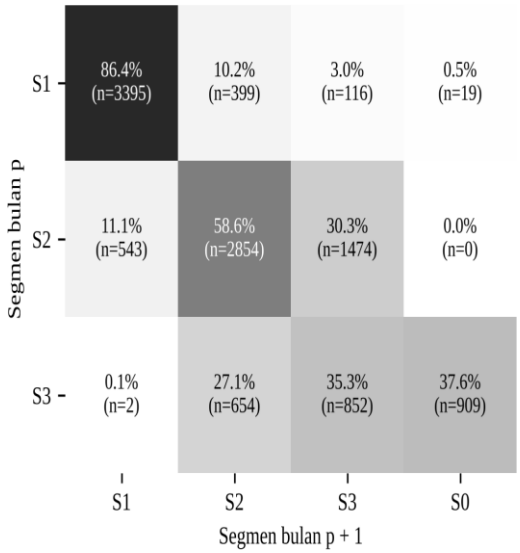


Figure 2. Transition Matrix of Monthly Segments from p to p + 1

From the total observation, 7,101 (63.3%) remained in the same segment, 1,199 (10.7%) increased, and 2,917 (26.0%) decreased. The S1 segment is the most stable, with 86.4 percent of users remaining. The S2 segment is the most dynamic, with an 11.1 percent increase to S1 and a 30.3 percent decrease to S3. In the following month, 37.6% of S3 users became inactive. The table shows the class distribution in the test and training data.

Table 6. Distribution of Segment Transfer Classes

Class	Training data	Test data
Dwon	2.040 (25,1%)	877 (28,5%)
Stull	5.223 (64,2%)	1.878 (61,0%)
Go on	876 (10,8%)	323 (10,5%)
Total	8.139	3.078

The Stable Class dominates, and the Rising Class is a minority. In this situation, the rationale for using class weighting and macro-F1 is reinforced. In the example user, the segment for June 2026 is S1 ($l = 3$). However, because the user did not transact at all from May to July 2026, their segment in July 2026 becomes S0 ($l = 0$). The label for this observation is Low, according to equation (7). Log M has a very high correlation with log F ($|r|$ greater than 0.9), as shown by the correlation check. Therefore, the monetary log features $p-1$ and $p-2$ were removed. The value per transaction, Δlog M, and the shopping trend indicate a constant shopping value. 21 features were used in the final model.

3.1.4 Hyperparameter Tuning Results

Copyright © 2026 The **Zulham Sitorus**, Page 1547
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9576
Hal 1541-1551
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

For each model, one hundred Optuna trials were used for tuning. For forward-chaining validation, the last four months of the training period—from February 2026 to May 2026—were used. The table shows the most suitable hyperparameters.

Table 7. Best Hyperparameters

Model	Best hyperparameter	Macro-F1 validation
XGBoost	max_depth = 10; learning_rate = 0.0824; subsample = 0.707; colsample_bytree = 0.613; min_child_weight = 1; gamma = 0.051; reg_lambda = 2.321; number of trees = 101 (early stopping)	0,6482
SVM	kernel = RBF; C = 26,373; gamma = 0,0159	0,6206

The RBF kernel is considered the best SVM kernel because it is better than the linear kernel. This indicates that the class boundary is not linear. XGBoost achieved a macro-F1 score of 0.6482 during validation, compared to SVM with a score of 0.6206.

3.1.6 Significance and Robustness Test Results

In the test data, n01 = 214 XGBoost predictions that were incorrect but correct by SVM, and n10 = 384 predictions with the opposite condition. Equation (12) shows that 2 = (|214 − 384| − 1)2 / (214 + 384) = 1692 / 598 = 47.76. The error patterns of the two models differ significantly with 1 degree of freedom, with a p-value of 0.001. This indicates that XGBoost reduces classification error.

However, the bootstrap confidence interval per user is 95%, and the macro-F1 difference (XGBoost − SVM) is only 0.0063. The advantage of XGBoost in macro-F1 is not significant because the interval contains zero. In other words, both models are comparable in terms of class performance, although XGBoost is superior overall. The table shows the stability of performance per test month.

Table 8. Macro-F1 per Test Month

Test month	XGBoost	SVM
June 2026	0,6043	0,5820
Juli 2026	0,5710	0,5985
Agustus 2026	0,6365	0,6146
September 2026	0,6447	0,6401

The lowest performance of SVM was in June 2026 (0.5820) and the lowest performance of XGBoost was in July 2026 (0.5710), the semester break month when student shopping patterns changed sharply. The model appeared in August 2026 and September 2026, when academic activities resumed as usual. The table shows the results of the sensitivity test regarding window length and imbalance handling methods.

Table 9. Macro-F1 Sensitivity Test

Skenario	XGBoost	SVM
Main: L = 3, class weighting	0,6178	0,6115
L = 1 month, class weighting	0,5555	0,5786
L = 6 months, class weighting	0,6537	0,6106
L = 3, SMOTE	0,5879	0,6160

Sensitivity test with hyperparameters from the main scenario without resetting. The performance of both models decreased during the one-month window because the segment is highly sensitive to monthly fluctuations. In the six-month window, the macro-F1 of XGBoost increased to 0.6537. However, the longer window also made the segment slower to react to behavioural changes, making it unsuitable for early warning. The performance of SMOTE did not improve; the macro-F1 of XGBoost dropped to 0.5879, while SVM remained relatively stable at 0.6160. As a result, class weighting remains a major issue.

3.1.7 SHAP Interpretation Results

TreeSHAP was calculated for all 3,078 test observations; KernelSHAP was calculated for 210 stratified test observations, 70 per class, with 50 background data centroids, and 300 evaluations per observation. While TreeSHAP was completed in seconds, KernelSHAP compression took 12.3 minutes. Figure 3 shows the global feature importance for the Down class.

Copyright © 2026 The **Zulham Sitorus**, Page 1548
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9576
Hal 1541-1551
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

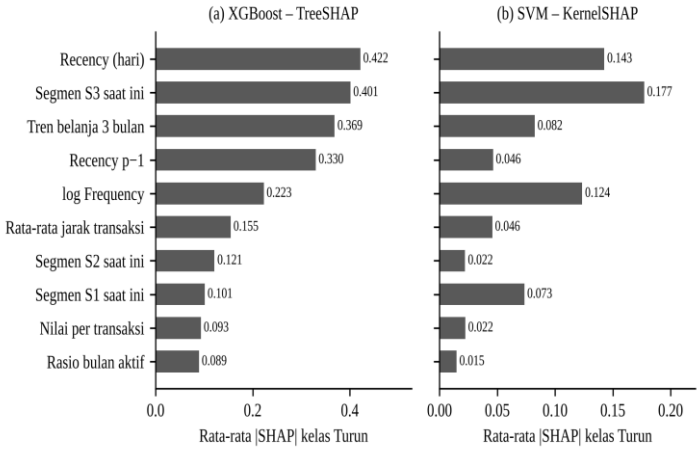


Figure 3. Global Feature Importance for Downstream Class: (a) XGBoost, (b) SVM

Because SHAP XGBoost is expressed in log-odds, while SHAP SVM is expressed in probability, the scale of the second panel's axis is different. Therefore, feature rankings, rather than the main values, are used to compare the two models.

Table 10. Ranking of Feature Importance for Downstream Class

Feature	XGBoost Ranking	SVM Ranking	Direction of influence*
Recency (day)	1	2	Raising
Segment S3 currently	2	1	Lowering
3-month shopping trend	3	4	Lowering
Recency p−1	4	8	Raising
log Frequency	5	3	Lowering
Average transaction distance	6	9	Raising
S2 segment currently	7	12	Raising
S1 segment currently	8	6	Raising

In both models, four out of the five top features are identical: frequency (days), current S3 segment, three-month shopping trend, and log frequency. The Jaccard index for the five main features is 0.667, and the Spearman rank correlation for all features is 0.868 ($p < 0.001$). Therefore, a highly consistent explanation is produced by both models working with different paradigms. Figure 4 shows how each feature affects XGBoost.

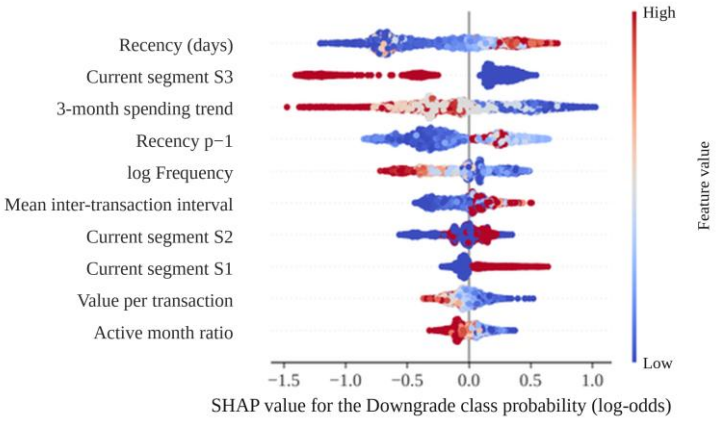


Figure 4. SHAP Value Distribution of XGBoost for the Down Class

Three main patterns are shown in Figure 4. First, the likelihood of segment decline is increased by high frequency (red dots), which means that the longer users do not make transactions, the greater the likelihood of segment decline. Second, the likelihood of segment decline is increased by a negative three-month shopping trend (blue dots), which means that the likelihood of declining shopping value becomes an early signal before users stop. Third, the likelihood of segment decline is reduced by high frequency and the ratio of active months. Status as S3 reduces the likelihood of decline because

Copyright © 2026 The **Zulham Sitorus**, Page 1549
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.9576
Hal 1541-1551
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

the only possibility of decline is becoming completely inactive. Conversely, status as S1 increases the likelihood of decline because S1 users have the most room for decline. Figure 5 shows the local explanation for the user.

3.2 Discussion

Most QRIS users are unstable, according to the research findings. Almost 63 percent of observations remain in the same segment from month to month, and segment S2 has the highest tendency to move in both directions. These results align with [7] that customer segments are moving targets, and with the findings [6] regarding gradually changing purchasing behaviour. Thus, static RFM segmentation calculated once will quickly become outdated [11]. In contrast, dynamic RFM with segment definitions locked during the training period has proven capable of consistently capturing these shifts.

A metric-dependent picture was created by comparing XGBoost and SVM. Accuracy, MCC, ROC-AUC, and the McNemar test favoured XGBoost. These results support the finding that tree-based models tend to excel on tabular data [34] and is consistent with [35], placing XGBoost above SVM. However, the macro-F1 difference is not significant, and SVM excels in balanced precision and recall for the minority class. These second results align with the research conducted Hartanto et al. [36], which found that SVM is the best for determining the segment of at-risk customers. Thus, the differences in prioritised metrics explain the discrepancies in previous research findings: XGBoost is more accurate overall, while SVM is more sensitive to users who churn.

Status, frequency, frequency, and shopping trends are the main drivers of segment migration, according to SHAP interpretation. These results are consistent with those found in [20] [21], which identified that the main factors predicting churn are the number of transactions, transaction value, and their changes over periods. The belief that these components indicate user behaviour patterns rather than algorithmic artefacts is reinforced by the high consistency of explanations between the two models ($\rho = 0.868$). Furthermore, these findings address the criticism that following a black box model is difficult [18], local explanations show why a user is at risk, allowing interventions to be designed according to the user.

4. CONCLUSION

This research found results from the comparison of XGBoost and SVM in predicting the movement of QRIS user behaviour segments based on dynamic RFM with SHAP interpretation on simulated data that mimics campus minimarket transaction patterns. Dynamic RFM with a three-month window and K-Means ($k = 3$) formed segments S1 (High-Value Active), S2 (Regular), and S3 (Passive), along with S0 (Inactive). Segmentation must be updated periodically because 26.0% of observations drop segments and 10.7% rise within a month. SVM achieved a macro-F1 of 0.6115, accuracy of 0.6634, and MCC of 0.4884, while XGBoost achieved a macro-F1 of 0.6178, accuracy of 0.7186, and MCC of 0.4945. Although XGBoost had a lower error rate (McNemar test, $p < 0.001$), the difference in macro-F1 was not significant. In detecting S1 users who will churn, SVM outperformed with 76.3% compared to 60.7%. The main drivers of segment migration are frequency, status, three-month shopping trends, and frequency. The explanations of both models are consistent (Spearman $\rho = 0.868$; Jaccard Five Main Features $\rho = 0.667$).

REFERENCES

[1] Bank Indonesia, *Peraturan Anggota Dewan Gubernur Nomor 21/18/PADG/2019 tentang Implementasi Standar Nasional Quick Response Code untuk Pembayaran*. Jakarta, Indonesia: Bank Indonesia, 2019.

[2] S. Mayasari, "BI Mencatat Nilai Transaksi QRIS Capai Rp 164,48 Triliun pada Januari 2026," *Kontan.co.id*, Mar. 12, 2026. [Online]. Available: https://keuangan.kontan.co.id/news/bi-mencatat-nilai-transaksi-qris-capai-rp-16448-triliun-pada-januari-2026

[3] L. Tesaloni, "Transaksi Digital Tumbuh 39,24%, Penggunaan QRIS Melonjak 75,8%," *Kontan.co.id*, Oct. 23, 2026. [Online]. Available: https://keuangan.kontan.co.id/news/transaksi-digital-tumbuh-3924-penggunaan-qris-melonjak-758

[4] E. H. Muchtar *et al.*, "Quick response code Indonesia standard (QRIS) E-payment adoption: customers perspective," *Cogent Business & Management*, vol. 11, no. 1, Art. no. 2316044, 2024, doi: 10.1080/23311975.2024.2316044.

[5] A. Zakariya and A. A. Arifin, "The Application of Digital Payments (QRIS & E-Wallet) on the Efficiency of MSME Transactions," *Journal of Finance, Economics and Business*, vol. 4, no. 2, pp. 1–19, 2025, doi: 10.59827/jfeb.v4i2.326.

[6] C. P. Blocker and D. J. Flint, "Customer segments as moving targets: Integrating customer value dynamism into segment instability logic," *Industrial Marketing Management*, vol. 36, no. 6, pp. 810–822, 2007, doi: 10.1016/j.indmarman.2006.05.016.

[7] S. H. Ha, S. M. Bae, and S. C. Park, "Customer's time-variant purchase behavior and corresponding marketing strategies: an online retailer's case," *Computers & Industrial Engineering*, vol. 43, no. 4, pp. 801–820, 2002, doi: 10.1016/S0360-8352(02)00141-9.

[8] I. Wicaksono, "Role of Habit and Trust in Driving the Continuance of QR Code Payment Usage: A Study of QRIS Consumers in Indonesia," *Journal of Marketing Innovation (JMI)*, vol. 5, no. 2, 2025, doi: 10.35313/jmi.v5i2.273.

[9] A. M. Hughes, *Strategic Database Marketing*. Chicago, IL, USA: Probus Publishing, 1994.

[10] P. S. Fader, B. G. S. Hardie, and K. L. Lee, "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis," *Journal*

Copyright © 2026 The **Zulham Sitorus**, Page 1550
This Journal is licensed under a Creative Commons Attribution 4.0 International License

of Marketing Research, vol. 42, no. 4, pp. 415–430, 2005, doi: 10.1509/jmkr.2005.42.4.415.

[11] M. A. Gomes and T. Meisen, "A review on customer segmentation methods for personalized customer targeting in e-commerce use cases," *Information Systems and e-Business Management*, vol. 21, no. 3, pp. 527–570, 2023, doi: 10.1007/s10257-023-00640-4.

[12] H. T. Duong and T. D. Pham, "Customer segmentation for marketing and business management in electronic retailing," *International Journal of Business Innovation and Research*, vol. 40, no. 5, pp. 53–71, 2026, doi: 10.1504/ijbir.2026.153810.

[13] Y. R. Rahadian *et al.*, "Customer segmentation based on RFM, deep embedded clustering, and lifetime value analysis for loyalty strategy optimization," *Procedia Computer Science*, vol. 284, pp. 1611–1618, 2026, doi: 10.1016/j.procs.2026.07.154.

[14] Y. Bai, "Research Review on Customer Life Cycle Value Prediction Based on RFM Analysis and Ensemble Learning," in *CPS Digital Library - Series of Conferences*, 2026. [Online]. Available: https://seriesofconference.com/index.php/SCJ/article/download/403/401

[15] T. Herve, "Identification of the Productive Potential of Customers in a Commercial Bank Portfolio: Calculation of the Customer Lifetime Value using Statistical Learning Method," *International Journal of Science and Research (IJSR)*, vol. 10, no. 1, pp. 1435–1444, 2021, doi: 10.21275/sr21123105748.

[16] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/BF00994018.

[17] M. N. H. Mamun, "Advancements in Machine Learning for Customer Retention: A Systematic Literature Review of Predictive Models and Churn Analysis," *Journal of Sustainable Development and Policy*, vol. 1, no. 1, pp. 250–284, 2025, doi: 10.63125/9b316w70.

[18] 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," *Machine Learning and Knowledge Extraction*, vol. 7, no. 3, Art. no. 105, 2025, doi: 10.3390/make7030105.

[19] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 4765–4774.

[20] Y. Li and K. Yan, "Prediction of bank credit customers churn based on machine learning and interpretability analysis," *Data Science in Finance and Economics*, vol. 5, no. 1, pp. 19–34, 2025, doi: 10.3934/dsfe.2025002.

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

[22] L. Li, "An Experimental Study on E-commerce Customer Churn Prediction Based on the Explainable XGBoost Algorithm," *Frontiers in Computing and Intelligent Systems*, vol. 17, no. 1, pp. 68–72, 2026, doi: 10.54097/44rwqq62.

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

[24] F. Provost and T. Fawcett, "Data Science and its Relationship to Big Data and Data-Driven Decision Making," *Big Data*, vol. 1, no. 1, pp. 51–59, 2013, doi: 10.1089/big.2013.1508.

[25] D. Donoho, "50 Years of Data Science," *Journal of Computational and Graphical Statistics*, vol. 26, no. 4, pp. 745–766, 2017, doi: 10.1080/10618600.2017.1384734.

[26] G. Shmueli, "To Explain or to Predict?" *Statistical Science*, vol. 25, no. 3, pp. 289–310, 2010, doi: 10.1214/10-STS330.

[27] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA, USA: University of California Press, 1967, vol. 1, pp. 281–297.

[28] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.

[29] S. Kaufman, S. Rosset, C. Perlich, and O. Stitelman, "Leakage in data mining: Formulation, detection, and avoidance," *ACM Transactions on Knowledge Discovery from Data*, vol. 6, no. 4, pp. 1–21, 2012, doi: 10.1145/2382577.2382579.

[30] C. Bergmeir and J. M. Benítez, "On the use of cross-validation for time series predictor evaluation," *Information Sciences*, vol. 191, pp. 192–213, 2012, doi: 10.1016/j.ins.2011.12.028.

[31] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[32] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[33] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 56–67, 2020, doi: 10.1038/s42256-019-0138-9.

[34] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?" in *Advances in Neural Information Processing Systems 35*, 2022, pp. 507–520, doi: 10.52202/068431-0037.

[35] T. Lan, H. A. Aldowah, and S. A. Shahani, "Machine learning analysis of telecom customer churn in California," *International Journal of Informatics and Communication Technology (IJ-ICT)*, vol. 15, no. 4, pp. 2000–2010, 2026, doi: 10.11591/ijict.v15i4.pp2000-2010.

[36] U. I. Hartanto, I. G. P. A. Buditjahjanto, and W. Yustanti, "Hybrid Clustering and Classification of At-Risk Customer Segments in Network Marketing," *Journal of Information Engineering and Educational Technology*, vol. 9, no. 1, pp. 42–50, 2025, doi: 10.26740/jieet.v9n1.p42-50.

This Journal is licensed under a Creative Commons Attribution 4.0 International License