



MyPertamina Transaction Increase Strategy Through Optimizing Application Use at Gas Stations Based on Machine Learning

Fresly Leo Chandra Hutapea^{1*}

Universitas Padjajaran,
Indonesia

Sunu Widiyanto²

Universitas Padjajaran,
Indonesia

Samidi³

Universitas Padjajaran,
Indonesia

***Corresponding author:**

Fresly Leo Chandra Hutapea,
Universitas Padjajaran, Indonesia.
✉ fresly.leo@gmail.com

Article Info :

Article history:

Received: February 23, 2026

Revised: March 13, 2026

Accepted: March 15, 2026

Keywords:

customer loyalty; digital payment;
gradient boosted trees; machine
learning; mypertamina.

Abstract

Background: The digital transformation in the energy sector is accelerating cashless payment adoption. MyPertamina, PT Pertamina's mobile payment platform, has been deployed across gas stations nationwide. However, only 5.65% of 1.3 million registered users (September 2024) are active, and only 11.5% qualify as loyal users (≥ 4 transactions/month). This performance gap, which exists between registered, active, and loyal users, is the focus of this study, as existing research has overlooked the role of gas station operational governance in shaping transaction behavior.

Objective: This study aims to (1) identify factors influencing MyPertamina usage at DKI Jakarta gas stations, (2) develop a machine learning-based prediction model to classify transaction behavior (MyPertamina vs. cash), and (3) create a G-STIC framework to increase adoption, usage intensity, and loyalty.

Methods: A quantitative case study using the CRISP-DM framework analyzed secondary POS transaction data from 8,000 transactions (5,200 MyPertamina; 2,800 cash) at DKI Jakarta gas stations (2024). Stratified sampling was used, and the models—Decision Tree, Gradient Boosted Trees, and Decision Stump—were evaluated based on accuracy, precision, and recall.

Results: Gradient Boosted Trees achieved the highest accuracy (97.75%). Gas Station Type and Class showed the strongest correlations with MyPertamina usage, suggesting further investigation of the Gas Station Code correlation.

Conclusion: MyPertamina adoption is influenced by operational governance and service standards. The G-STIC framework provides actionable strategies for increasing digital transaction adoption, contributing to both academic literature and managerial practice in the energy retail sector.

To cite this article: Hutapea, F. L. C., Widiyanto, S., & Samidi. (2026). MyPertamina transaction increase strategy through optimizing application use at gas stations based on machine learning. *INKUBIS: Jurnal Ekonomi dan Bisnis*, 8(1), 112–125. <https://doi.org/10.59261/inkubis.v8i1.150>

INTRODUCTION

The rapid proliferation of digital platforms in emerging markets has fundamentally altered consumer-service interactions, particularly within the energy sector where mobile applications serve as primary touchpoints for both transactions and loyalty engagement (Khoiro et al., 2025; Kusuma et al., 2022). In the energy retail sector, digital transformation has similarly reshaped customer engagement dynamics, with cashless payment adoption increasingly studied through digital service quality, user experience, and customer adoption frameworks (Murrar et al., 2025; Pratama & Handayani, 2025). PT Pertamina, one of Indonesia's leading energy companies, has embarked on a digital transformation process since 2017 by integrating digital technology into various business lines, including the launch of MyPertamina, a mobile application

for cashless payments at Pertamina gas stations (Indradewa & Iqbal, 2021).

MyPertamina is a digital solution for consumers of Pertamina products, integrated with various services that provide benefits to consumers (Kustyana et al., 2024). The application boasts numerous superior features, such as electronic payments (e-payments) using an integrated payment system with LinkAja, OVO, GoPay, Mandiri, BNI, BRI, Visa, and Mastercard (Sinurat et al., 2022). Furthermore, every purchase made through the application earns loyalty points that can be collected and redeemed for various rewards and attractive promotions from MyPertamina. MyPertamina is expected to become a strategic tool in maintaining and increasing market share for Pertamina's fuel and non-fuel products (non-fuel business) (Putra & Menorizah, 2025).

DKI Jakarta was selected as the study area because it represents the highest fuel transaction volume and the most mature digital literacy profile among Indonesia's provinces, making it an ideal context for evaluating digital payment adoption at scale. According to September 2024 data, registered MyPertamina users reached 1.3 million. For clarity: registered users are those who have created an account; active users are those who have completed at least one transaction; and loyal users are those making at least four transactions per month. Only 5.65% of registered users are active users, and only 11.5% of active users are categorized as loyal users. This fact indicates a significant gap between the number of registered users, active users, and loyal users. In other words, although the application has succeeded in attracting new users, the level of engagement and frequency of use remains relatively low (Prastiyo & Lahindah, 2024).

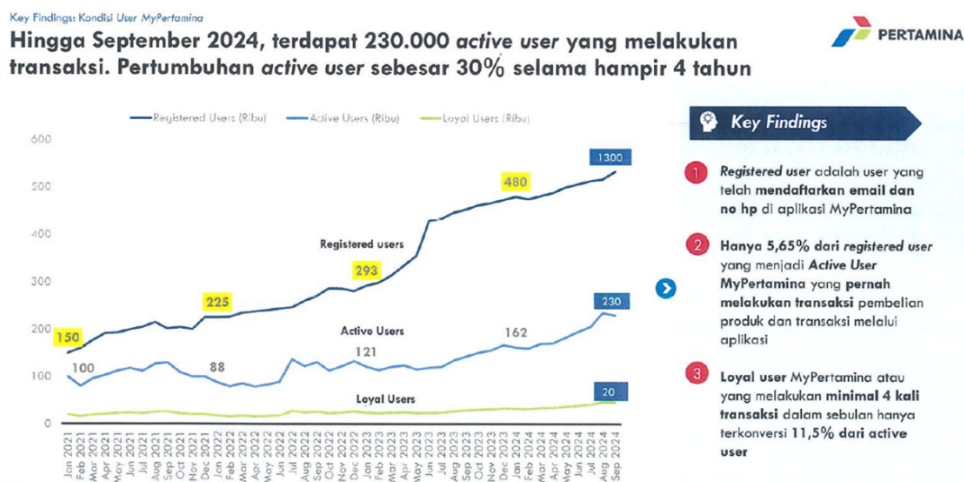


Figure 1. MyPertamina Application User Data 2024
Source: PT Pertamina Patra Niaga 2024 Annual Report

Based on a 2023 survey of MyPertamina users, more than 50% of active users employ the app to pay for fuel purchases, while 22% utilize the Right Subsidy Program feature. However, only 8% access the loyalty points feature. This underutilization has direct revenue and retention implications: low feature engagement limits Pertamina's ability to capture customer lifetime value, reduces cross-selling opportunities for non-fuel products, and weakens the loyalty loop that could otherwise reduce customer churn and stabilize market share. The low utilization of these additional features indicates that the MyPertamina app has not yet fully delivered a value-added experience capable of building long-term engagement with consumers.

This situation has several critical implications. First, the low utilization of features other than payments indicates that the MyPertamina app is still perceived as merely a transaction tool, rather than an integrated service platform. Research on digital payment adoption in the energy sector highlights that perceived service quality and ease of use are primary antecedents of application adoption; the limited engagement with loyalty and subsidy features in MyPertamina suggests these functionalities have not yet achieved sufficient perceived usefulness among users (Lazuardy et al., 2025; Pratama & Handayani, 2025).

Studies on digital transformation in fuel payment systems further confirm that social influence and facilitating conditions including infrastructure readiness and organizational support significantly moderate user loyalty (Khoiro et al., 2025). Murrar (2025) emphasize that

strategies for driving customer adoption of AI-powered mobile apps require addressing both user-side and supply-side enablers, implying that gas station-level operational support is critical for diffusing MyPertamina usage. Second, if this trend continues, Pertamina could potentially miss a strategic opportunity to build a loyal customer base, even though loyalty is a key factor in maintaining market share amid intense competition in the energy industry (Hastuti et al., 2024).

In the context of consumer data management, the company currently faces the challenge of leveraging big data generated from millions of daily MyPertamina user transactions. The selection of predictive algorithms such as Decision Trees, Gradient Boosted Trees, and Decision Stump in this study was based on the characteristics of transaction data, which is generally tabular and heterogeneous, and requires output that is not only accurate but also explainable. Accordingly, feature importance outputs and decision rules generated by these models are presented in the Results section to fulfill the interpretability requirement, consistent with recent ML studies on MyPertamina (Madyatmadja et al., 2025; Rousyati et al., 2023). With these predictive algorithms, companies can identify application usage patterns, predict transaction trends, and determine key attributes that influence user loyalty.

Previous studies have examined digital payment adoption from demand-side perspectives Dash (2023); Sinurat (2022); Prastiyo (2024), focusing on user attitudes and behavioral intentions. However, supply-side governance factors particularly the role of gas station type, class, and operational standardization remain underexplored as predictors of digital transaction adoption at the point of sale. This study is novel because it is the first to apply a machine learning-based classification approach (CRISP-DM framework) to POS transaction data from a national energy retailer's gas station network, integrating supply-side operational attributes with predictive modeling to derive actionable business strategies. Therefore, this research focuses on strategies to increase MyPertamina transactions through optimizing application usage at gas stations using machine learning.

The objectives of this research are: (1) to identify and analyze factors influencing the level of MyPertamina application usage by consumers at gas stations in DKI Jakarta Province; (2) to develop and evaluate machine learning-based prediction models to classify transaction behavior (MyPertamina vs. cash); and (3) to formulate a business strategy using the G-STIC framework to sustainably increase the level of adoption, usage intensity, and consumer loyalty to the MyPertamina application. This research is expected to contribute in two main aspects. First, academically, this research adds to the literature on the use of analytical technology in managing digital applications in the energy sector. Second, practically, the research results can provide concrete recommendations for Pertamina in optimizing application features, improving user experience, strengthening customer loyalty, and supporting the company's competitiveness in the era of digitalization and energy transition.

METHOD

This research employed a quantitative descriptive case study design to analyze MyPertamina transaction behavior at gas stations in DKI Jakarta Province. The primary data source was internal subsidized fuel sales transaction data extracted from the Point of Sale (POS) system of Pertamina gas stations in DKI Jakarta Province for the year 2024. The secondary data used was sourced from fuel sales transaction records through the Point of Sale (POS) system at all gas stations in DKI Jakarta Province in 2024. This data included information on transaction time, type of fuel product purchased, purchase volume, and frequency of MyPertamina application use by consumers. The data on the number of active MyPertamina users up to September 2024 and the results of the 2023 user survey were used only in a limited capacity to support the compilation of the background section and were not processed in the data analysis stage. The data required for this research consisted of secondary data sourced from the company's internal systems, specifically those related to fuel sales transactions through the MyPertamina application, as well as supporting data used in compiling the background.

The primary data source was internal subsidized fuel sales transaction data extracted from the Point of Sale (POS) system of Pertamina gas stations in DKI Jakarta Province for the year 2024. The transaction data used for modeling was internal operational (primary-sourced) data obtained from the company's POS systems, not derived from literature. Supporting contextual data (active user counts and the 2023 survey) were used only for the background section and

were not included in the modeling analysis. This data included information on transaction time, type of fuel product purchased, purchase volume, and frequency of MyPertamina application use by consumers.

The data analysis followed the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework in the following sequential stages (Azeroual et al., 2025): (1) Business Understanding defining research objectives and analytical questions; (2) Data Understanding collecting and exploring the POS dataset to assess data quality and attribute characteristics; (3) Data Preparation including data selection, cleaning, transformation, and integration; (4) Modeling building and comparing Decision Tree, Gradient Boosted Trees, and Decision Stump classification models in RapidMiner; (5) Evaluation comparing model performance using accuracy, precision, recall, F1-score, and AUC metrics; and (6) Deployment translating modeling results into the G-STIC business strategy framework.

The total data population was 7,473,523 transactions; for computational efficiency, stratified sampling was conducted, yielding a sample of 8,000 transactions consisting of 5,200 MyPertamina transactions and 2,800 cash transactions. A 90/10 training-testing split (7,200 training / 800 testing) was applied. Due to computational resource constraints, k-fold cross-validation was not applied in this study. This represents a methodological limitation that may introduce optimistic bias in accuracy estimates. Future research is recommended to validate the model using k-fold cross-validation or out-of-sample testing on data from other provinces to assess generalizability.

RESULT AND DISCUSSION

Result

Analysis of Factors Influencing the Level of Use of the MyPertamina Application by Consumers at Gas Stations

In evaluating the most optimal model for predicting the behavior of consumers using the MyPertamina application at gas stations in the Jakarta Province area, a modeling stage was carried out using a machine learning-based classification method.

1. Business Understanding. The central business questions guiding this analysis are: (a) Which gas station and transaction attributes most strongly predict whether a consumer will pay via MyPertamina or cash? (b) Which classification model best separates these two behaviors? (c) How can model outputs be translated into a governance and service strategy to increase MyPertamina adoption, usage intensity, and loyalty? MyPertamina serves as Pertamina's primary digital retention tool; however, persistent loyalty gaps-5.65% active rate, 11.5% loyal rate-indicate that the application's value proposition has not yet been operationally embedded across the gas station network (Prastiyo & Lahindah, 2024).
2. Data Understanding. In this stage, data was collected from downloaded detailed data on subsidized fuel sales transactions at gas stations in the DKI Jakarta Province in 2024, available on the Pertamina dashboard and sourced from the PC POS of each Pertamina gas station in Excel format. For this study, the author only collected data relevant to the data mining objectives, including variables such as gas station code, city, product, volume sold, transaction type, transaction time, gas station class, and gas station type. Regarding data governance: the dataset contains no personal customer identifiers. Transaction records are anonymized at the aggregate POS level, with gas station codes serving as station-level identifiers rather than individual consumer identifiers. No customer names, IDs, or personal data were included in the analysis dataset, ensuring compliance with applicable data privacy standards.
3. Data Preparation. Given the limited computing resources and the need for efficiency, a stratified sampling strategy was implemented to select 8,000 transaction data points (5,200 MyPertamina; 2,800 cash) from the total population of 7,473,523. The stratification was based on the original class proportions in the population (approximately 65% MyPertamina, 35% cash), intentionally preserving the natural imbalance to maintain representativeness and ecological validity. This proportional stratification means the model's performance metrics reflect real-world class distribution; however, it also means results may be less generalizable to gas stations with markedly different adoption profiles outside DKI Jakarta.

Data preparation involved the following sub-stages: 1) Data Selection: Variables used: Gas Station Code, City, Product, Time_Category, Gas Station Type, Gas Station Class, Volume Sold.

Transaction Type is designated as the label/target variable and was excluded from the feature set to prevent data leakage. 2) Data Cleaning: Steps included eliminating duplicates, checking for missing values, standardizing variable types for RapidMiner compatibility, and validating data units and outlier detection.

Table 1. Final Dataset

SPBU_Code	City	Product	Sold_Volume	Time_Category	SPBU_Type	SPBU_Class	Transaction_Type
3113730	Central Jakarta	PERTALITE	8.0974 86785	evening	COCO	SPBU Pasti Prima Level	My Pertamina
3413810	North Jakarta	PERTALITE	7.6552 71506	morning	DODO	SPBU Pasti Prima Level	My Pertamina
3489636	East Jakarta	PERTALITE	16.389 59037	morning	DODO	SPBU Pasti Prima Level	My Pertamina
3115318	Central Jakarta	PERTALITE	4.1187 6849	morning	COCO	SPBU Pasti Prima Level	My Pertamina
3113830	North Jakarta	PERTALITE	30.622 78027	morning	COCO	SPBU Pasti Prima Level	My Pertamina
3411706	West Jakarta	PERTALITE	10	evening	DODO	Basic	Cash
3410504	Central Jakarta	PERTALITE	20	morning	DODO	Basic	Cash
3116612	East Jakarta	PERTALITE	30.025 24864	evening	COCO	Basic	My Pertamina
3413903	East Jakarta	PERTALITE	3.5009 99928	morning	DODO	Basic	Cash
3414306	North Jakarta	BIO_SOLAR	19.118 00003	afternoon	DODO	Basic	Cash

Source: Researcher (2025)

- a) Data Transformation: Normalization and standardization were applied to numerical variables (Volume Sold). Categorical variables were numerically encoded for correlation analysis. The dataset falls under supervised learning, with Transaction Type as the classification label.

▼ Kode_SPBU	Nominal	0	Least 3499975 (1)	Most 3411706 (158)	Value: 3411706 (158), 3414203 (74), ...[5181 more]
▼ City	Nominal	0	Least Kota Jak [...] at (1010)	Most Kota Jak [...] at (3049)	Value: Kota Jakarta Pusat (3049), Kota Jakarta Utara (1407), ...[3 more]
▼ Produk	Binominal	0	Negative PERTALITE	Positive BIO_SOLAR	Value: PERTALITE (7280), BIO_SOLAR (720)
▼ Volume_Terjua	Real	0	Min 0.077	Max 191.176	Average 8.415
▼ Waktu_kategor	Nominal	0	Least siang (1649)	Most sore (2244)	Value: sore (2244), malam (2168), ...[2 more]
▼ Tipe_SPBU	Nominal	0	Least CODO (212)	Most DODO (4869)	Value: DODO (4869), COCO (2919), ...[1 more]
▼ Kelas_SPBU	Nominal	0	Least Pasti Pa [...] ent (770)	Most SPBU Pas [...] 2 (3647)	Value: SPBU Pasti Prima Level 2 (3647), Basic (2533), ...[2 more]
▼ Jenis_Transaks	Binominal	0	Negative My Pertamina	Positive Cash	Value: My Pertamina (5200), Cash (2800)

Figure 2. Data Statistics

Source: Researcher (2025)

Figures 1 and 2 illustrate the final dataset used in the study and a summary of its statistics and characteristics after the cleaning and transformation stages. Figure 1 shows an example of a transaction data structure containing key attributes such as Gas Station Code, City, Product, Sold Volume, Category Time, Gas Station Type, Gas Station Class, and Transaction Type as labels (MyPertamina or Cash). Next, Figure 2 displays statistics for each attribute, including data type (nominal/binominal/real), minimum–maximum and average values for the numeric variable (Sold Volume), as well as the category distribution for nominal variables such as transaction time (morning/afternoon/evening/night), gas station class and type, and class composition within Transaction Type. These visualizations confirm that the dataset is ready to be used for classification modeling to predict payment methods and identify factors related to MyPertamina usage.

Correlation Matrix

Correlations involving nominal variables (City, Product, Gas Station Code) require careful interpretation. For this study, nominal variables were numerically encoded prior to Pearson correlation analysis. It is acknowledged that this approach does not capture the full relational structure of categorical data. A more robust complementary approach such as chi-square tests or Cramér's V for nominal-nominal associations is recommended in future work. The strong negative correlation of Gas Station Code (-0.665) is interpreted as an ID effect reflecting inter-station variability in adoption rather than a causal factor; model-based feature importance (e.g., SHAP values) would provide a more defensible measure of this attribute's contribution. The correlation analysis was conducted in RapidMiner following the process illustrated in Figure 3.

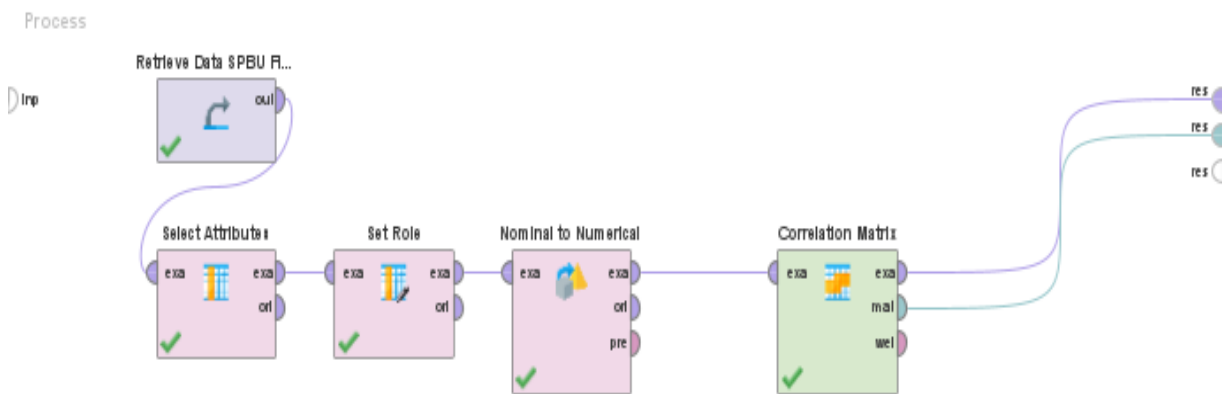


Figure 3. Matrix Correlation Test in RapidMiner
Source: Researcher (2025)

The execution and results are displayed as follows.

Table 2. Correlation Matrix Test Results

No.	Attribute	Correlation Coefficient	Direction of Relationship
1	Gas Station Type	+0.525	Strong Positive
2	Gas Station Class	+0.489	Strong Positive
3	City	+0.315	Moderate Positive
4	Volume Sold	+0.281	Weak Positive
5	Product	+0.226	Weak Positive
6	Time Category	-0.044	Weak Negative
7	Gas Station Code	-0.665	Strong Negative (ID effect*)

Source: Researcher (2025)

Table 3. Results of the Correlation Matrix Test on Transaction Types

No	Attributes	Correlation Coefficient	Direction of Relationship
1	Gas Station Type	+0.525	Strong Positive
2	Gas Station Class	+0.489	Strong Positive
3	City	+0.315	Weak Positive
4	Volume Sold	+0.281	Weak Positive
5	Product	+0.226	Weak Positive
6	Time Category	-0.044	Weak Negative
7	Gas Station Code	-0.665	Strong Negative

Table 3 above illustrates the results of the correlation analysis between factors influencing consumer usage of the MyPertamina application at gas stations. A value closer to 1 indicates a strong positive correlation between the two variables. Conversely, a value closer to -1 indicates a strong negative correlation between the two variables. Meanwhile, a value closer to 0 indicates a weak correlation between the two variables. The figure above reveals the following correlations between factors:

1. Gas station type has a strong positive correlation with MyPertamina transaction usage. This indicates that consumers are more likely to use MyPertamina at gas stations with certain types of ownership and management, which have more standardized operational governance.
2. Gas station class has a strong positive correlation with MyPertamina transaction usage. This indicates that gas stations with higher service classes tend to have better MyPertamina adoption rates, along with more optimal service quality and compliance with operational standards.
3. City has a weak to moderate positive correlation with MyPertamina transaction usage. This indicates that regional characteristics, such as urbanization levels and digital literacy, also influence consumer preferences for digital transactions, although the effect is not dominant.
4. Volume Sold has a weak positive correlation with MyPertamina transaction usage. This indicates that consumers with larger fuel purchase volumes tend to use MyPertamina slightly more frequently than consumers with smaller purchase volumes.
5. Product has a weak positive correlation with MyPertamina transaction usage. This indicates that the type of fuel purchased is not a major factor driving MyPertamina usage, as the app is used relatively evenly across various product types.
6. Time Category has a weak negative correlation with MyPertamina transaction usage. This indicates that transaction time, whether morning, afternoon, evening, or night, does not significantly influence consumers' decisions to use MyPertamina.
7. Gas Station Code has a strong negative correlation with MyPertamina transaction usage. This indicates that there are significant differences in MyPertamina usage levels across gas stations, with certain gas stations still dominated by cash payments compared to digital transactions using MyPertamina.

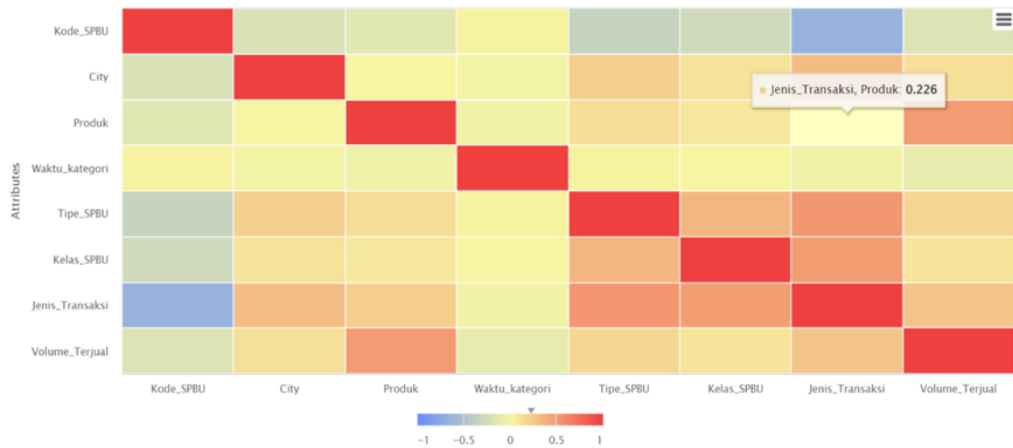


Figure 4. Heatmap Visualization of Correlation Matrix Test Results
Source: Researcher (2025)

The correlation results confirm that MyPertamina application usage is primarily driven by gas station governance characteristics Gas Station Type and Gas Station Class rather than transaction-level attributes. Gas stations with more standardized ownership types and higher service classes demonstrate higher MyPertamina adoption, consistent with the operational readiness and SOP compliance literature (Fähndrich, 2023). These findings align with recent empirical studies on MyPertamina adoption that highlight the centrality of physical infrastructure and organizational readiness in enabling digital transaction behavior Pratama (2025); Aditya (2025), suggesting that the quality and governance of gas station operations directly enable or constrain MyPertamina usage (Kusuma et al., 2022).

Modeling

Within the CRISP-DM framework, the modeling phase applies three classification algorithms Decision Tree, Gradient Boosted Trees, and Decision Stump selected for their ability to handle tabular, heterogeneous transaction data and produce interpretable outputs (Cai et al., 2022; Khataei & Bazargan, 2025; Shimaoka et al., 2024).

The modeling process in RapidMiner was executed in five core steps: (1) Data ingestion importing the prepared .csv dataset and configuring attribute roles (Transaction Type as label; remaining attributes as features); (2) Data splitting applying a 90/10 stratified split (7,200 training / 800 testing); (3) Model training running Decision Tree, Gradient Boosted Trees, and Decision Stump operators on the training partition; (4) Model application applying each trained model to the test partition using the 'Apply Model' operator; and (5) Performance evaluation measuring accuracy, precision, recall, F1-score, and AUC for each model using the 'Performance (Classification)' operator.

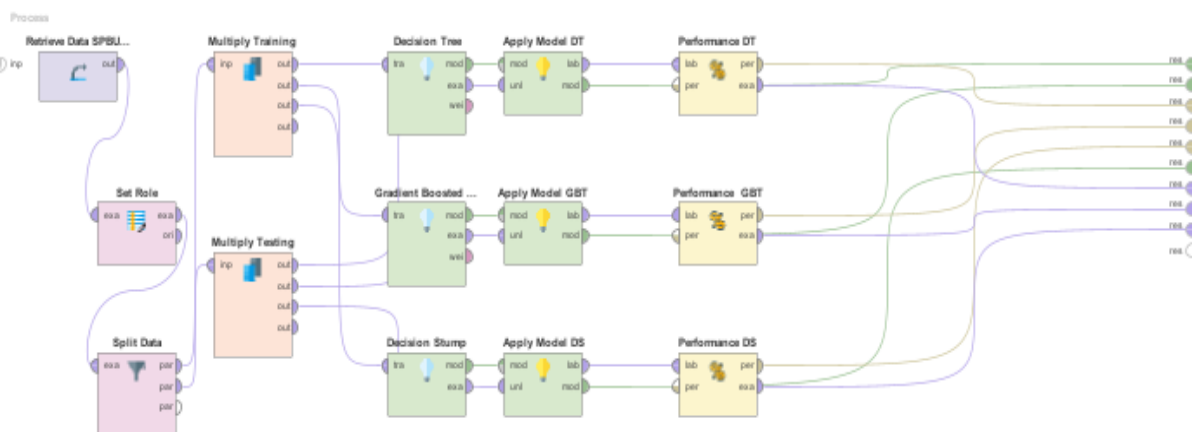


Figure 5. Compiling the Classification Method Algorithm Model in RapidMiner
Source: Researcher (2025)

Decision Tree Model Performance Results

Decision Tree Model: Accuracy 96.00% — 504 MyPertamina and 264 Cash transactions correctly classified, with 16 misclassifications in each class. Precision and recall for MyPertamina: 96.92%; for Cash: 94.29%. F1-score (MyPertamina): 96.92%; F1-score (Cash): 94.29%. AUC: 0.971, indicating excellent discriminative ability. The Decision Tree offers strong interpretability through hierarchical decision rules, though its generalization may be limited compared to ensemble methods.

accuracy: 96.00%

	true My Pertamina	true Cash	class precision
pred. My Pertamina	504	16	96.92%
pred. Cash	16	264	94.29%
class recall	96.92%	94.29%	

Figure 6. Performance Model Decision Tree
Source: Researcher (2025)

Gradient Boosted Trees Model: Accuracy 97.75% — 509 MyPertamina and 273 Cash transactions correctly classified. Precision/recall for MyPertamina: 98.64%/97.88%; for Cash: 96.13%/97.50%. F1-score (MyPertamina): 98.26%; F1-score (Cash): 96.81%. AUC: 0.993, reflecting near-perfect class separation. The GBT model excels at capturing nonlinear patterns and interaction effects between gas station attributes.

Performance Results of the Gradient Boosted Trees Model

accuracy: 97.75%

	true My Pertamina	true Cash	class precision
pred. My Pertamina	509	7	98.64%
pred. Cash	11	273	96.13%
class recall	97.88%	97.50%	

Figure 7. Performance of the Gradient Boosted Trees Model
Source: Researcher (2025)

Decision Stump Model: Accuracy 91.00% limited to a single decision rule. While the model achieves 100% precision for MyPertamina and 100% recall for Cash, the recall for MyPertamina (86.15%) and precision for Cash (79.55%) reveal significant class imbalance in performance. F1-score (MyPertamina): 92.59%; F1-score (Cash): 88.37%. AUC: 0.931. This model serves as a useful baseline but is unsuitable for deployment.

Decision Stump Model Performance Results

accuracy: 91.00%

	true My Pertamina	true Cash	class precision
pred. My Pertamina	448	0	100.00%
pred. Cash	72	280	79.55%
class recall	86.15%	100.00%	

Figure 8. Performance Model Decision Stump
Source: Researcher (2025)

Table 4. Expanded Model Performance Comparison (with F1-Score and AUC)

Model	Accuracy	Prec. MP	Recall MP	Prec. Cash	Recall Cash	F1 MP / Cash AUC
Decision Stump	91.00%	100.00%	86.15%	79.55%	100.00%	92.59% / 88.37% 0.931
Decision Tree	96.00%	96.92%	96.92%	94.29%	94.29%	96.92% / 94.29% 0.971
Gradient Boosted Trees	97.75%	98.64%	97.88%	96.13%	97.50%	98.26% / 96.81% 0.993

MP = MyPertamina

Source: Researcher (2025)

Implications for Digital Transaction Adoption Strategy: The superiority of Gradient Boosted Trees translates directly to strategic value: the model's high recall for MyPertamina (97.88%) minimizes 'missed conversions'-cash users who should have been identified as MyPertamina-convertible-while its high precision (98.64%) reduces misdirected interventions. In terms of error cost asymmetry, Type II errors (MyPertamina predicted as Cash) are considered more costly for business purposes, as they represent missed revenue and loyalty-building opportunities; the GBT model's performance is favorable on this dimension. The model outputs a probability score per transaction that can serve as a conversion risk score for operational targeting (threshold $\hat{\alpha}$ 0.65 recommended for high-confidence MyPertamina-convertible identification).

Model Evaluation

The evaluation phase of this project focused on answering the second problem formulation in the capstone document, namely determining the most optimal machine learning-based prediction model for predicting MyPertamina app usage behavior compared to cash transactions at gas stations in the DKI Jakarta Province. Based on the confusion matrix in the appendix, the total test data consisted of 800 observations, consisting of 520 MyPertamina transactions and 280 cash transactions. The evaluation, available in the appendix, covers accuracy, class precision, and class recall for two classes (MyPertamina and Cash).

1. Decision Tree. The Decision Tree model demonstrated excellent performance with an accuracy of 96.00%, as well as high and relatively balanced precision and recall values for both transaction classes (MyPertamina and cash). This model was able to correctly classify most transactions and had a low error rate. Furthermore, the Decision Tree offers the advantage of interpretability, as its structure can depict the influence of key attributes on MyPertamina usage hierarchically. However, compared to more complex models, the Decision Tree's generalization capabilities can still be improved.
2. Gradient Boosted Trees (GBT). The Gradient Boosted Trees model produced the best performance among the three models, with an accuracy of 97.75%, as well as very high precision and recall values that were more balanced across both classes. This model was able to minimize misclassification of both MyPertamina and cash transactions. The advantage of Gradient Boosted Trees lies in its ability to capture nonlinear patterns and interactions between variables by gradually combining multiple decision trees, resulting in a model with better predictive power and generalization than a single Decision Tree.
3. Decision Stump. The Decision Stump model had the lowest accuracy (91.00%) compared to the other two models. Although it achieved 100% precision in the MyPertamina class and 100% recall in the cash class, this model exhibited imbalances in performance between classes. The lower recall value for MyPertamina indicates that some MyPertamina transactions were not detected and classified as cash transactions. This reflects Decision Stump's limitations in capturing the complexity of relationships between variables because it only uses a single decision rule. Therefore, this model is more appropriate to use as a baseline model.

Table 5. Comparison of Classification Algorithm Model Performance

Model	Accuracy	Precision MyPertamina	Recall MyPertamina	Precision Cash	Recall Cash
Decision	91.00%	100.00%	86.15%	79.55%	100.00%

Stump					
Decision Tree	96.00%	96.92%	96.92%	94.29%	94.29%
Gradient Boosted Trees	97.75%	98.64%	97.88%	96.13%	97.50%

Source: Researcher (2025)

Based on the criteria available in the appendix (accuracy, precision, recall), the optimal machine learning-based prediction model in predicting the behavior of MyPertamina application usage at gas stations in the DKI Jakarta Province is Gradient Boosted Trees because it has the highest accuracy compared to other models, an excellent balance of precision and recall in both transaction classes thus minimizing classification bias and stronger generalization capabilities, especially in capturing complex patterns and nonlinear relationships between attributes that influence MyPertamina usage, thereby reducing two main risks: (1) Cash that "escapes" the conversion target (Cash predicted as MyPertamina) and (2) MyPertamina users who are mistakenly treated as Cash (MyPertamina predicted as Cash). These results are in accordance with literature findings that the Gradient Boosted Trees model tends to excel in binary classification and can be optimized through loss function design to be more robust, including in scenarios that require control over the consequences of misclassification (Mushava & Murray, 2024).

Discussion

Business Strategy Using the G-STIC Method as an Effort to Sustainably Increase the Level of Adoption, Intensity of Use, and Consumer Loyalty to the Mypertamina Application

At the Strategy stage, the highest correlations between gas station types and classes should be used as a basis for governance differentiation. COCO tends to be easier to standardize because operational control rests with the corporation. Therefore, the strategy is to make COCO, especially the Pasti Prima class, the benchmark for end-to-end digital service processes (queue layout, digital lanes, device readiness, training, and communication standards). Conversely, for CODO and DODO, success is heavily influenced by incentive alignment and implementation discipline at the operator/owner level; therefore, the strategy needs to combine capability development, strengthened compliance controls, and a KPI-based incentive scheme. Gas station class differentiation also determines the design of the service experience: for Pasti Prima, the strategy focuses on service excellence and reliability (strengthening the premium experience), while for Pasti Pas/Basic, the strategy must close gaps in device, network, and competency readiness to prevent digital transactions from increasing service friction. Therefore, the strategy is not based on marketing segmentation, but rather on an operational control architecture aligned with ownership structure and service standards. Based on the classification modeling results, Gradient Boosted Trees was determined as the best model with 97.75% accuracy and balanced precision–recall for both MyPertamina and cash classes, thus used as the basis for determining targets and priorities for the G-STIC strategy. This model produces a MyPertamina usage probability score for each combination of Gas Station Type, Gas Station Class, City, Volume Sold, Product, Time Category, and Gas Station Code attributes, which are then utilized in the Goal stage to formulate targets for increasing the portion of MyPertamina transactions, liters sold via the application, and reducing the dominance of cash transactions at high-risk gas stations. In the Strategy stage, the prediction score directs the differentiation of governance and service experience: the Pasti Prima class COCO serves as an end-to-end digital benchmark, while CODO/DODO and lower-middle class gas stations with high cash probability are focused on policy harmonization, incentive schemes, service leveling, and strengthening digital infrastructure. The Tactics Implementation–Control phase is implemented through a package of concrete programs gas station incentives, volume-based rewards and personalized notifications, location campaigns in cities with low adoption, time-based incentives and fast lanes during peak hours, and the SPBU Champion program which are implemented in stages (pilot–scale–standardize) and monitored through a dashboard that combines model scores with operational KPIs. This ensures that the strategy to increase MyPertamina adoption, usage intensity, and loyalty is more focused, measurable, and sustainable because it relies on the most accurate model predictions.

The Tactics Implementation Control phase should be viewed as an inseparable package. Strong tactics but weak implementation will result in variations in execution quality between gas

stations; conversely, implementation without control will degrade process discipline after the initial phase of the program. Therefore, tactics need to be concrete: simplifying digital service SOPs (including failure and fallback handling), appointing digital champions during busy shifts, visual management at service points, and integrating promotions that do not disrupt queue flow. Implementation must be gradual (pilot-scale-standardize) with a clear PIC and a realistic timeline, while control is conducted through a dashboard that monitors myPertamina transaction share, success rate, service time, and compliance indicators. Control management studies show that digitalization requires adaptation of control instruments, competencies, and governance so that controls can quickly read operational data and drive corrective actions, not just reporting (Fähndrich, 2023). On the demand side, digital payment adoption is not homogeneous; differences between adopters and non-adopters require appropriate on-site enablement and communication tactics to increase the diffusion of myPertamina usage without compromising the service experience (Dash et al., 2023).

A structured implementation plan to translate the G-STIC strategy into concrete operational steps across seven focus variables, complete with a division of PIC roles and a quarterly implementation schedule for 2025–2026. For the Gas Station Type variable, implementation is directed at developing and ratifying MyPertamina performance incentive policies for each gas station type, developing uniform digital service SOPs across types, and providing outreach and training to gas station owners/managers. This will be coordinated by the RFS, HC, and Hiswana Migas functions, with phases starting in Q1 to Q4 2026. For the Gas Station Class variable, the work plan focuses on establishing minimum standards for MyPertamina facilities and equipment, a baseline audit of equipment and service readiness, and a facility improvement program at mid-range/low-end gas stations. This will be implemented in stages by the RFS, Hiswana Migas, the gas station, and independent auditors, leading to a re-audit in Q4 2026.

For the City variable, implementation includes identifying cities with the lowest adoption based on transaction data, developing local campaigns, and providing network support (Wi-Fi/signal boosters) at gas stations with connectivity issues. RFS, RBS, and Hiswana Migas will act as PICs, with the implementation schedule running from Q1 to Q4 2026. For the Volume Sold variable, implementation steps use the Gradient Boosted Trees predictive model to map gas stations and high-risk cash customer segments, design a volume-based rewards program, and activate personalized notifications for high-volume customers, implemented by RFS and Hiswana Migas from Q1 to Q4 2026. For the Product variable, the work plan emphasizes determining priority products that are still predominantly cash-based, developing specific promotional packages, and adjusting the application's appearance to ensure easy access to quota information and purchase history. RFS and RBS will act as PICs, overseeing the design, launch, and evaluation stages of the promotion throughout 2026.

For the Time Category variable, implementation focuses on determining peak and off-peak hour categories per gas station, developing dedicated MyPertamina queue lines during peak hours, preparing digital companions, and implementing happy hour promotions during off-peak hours. This is led by RFS through the time slot and queue scheme design phase (Q1), trials at pilot gas stations (Q2), and rollout with improved mechanisms (Q3–Q4 2026). Finally, for the Gas Station Code variable, the implementation plan includes the development of a performance dashboard per Gas Station Code, classification of Gas Stations into champion, developing, and lagging, as well as the implementation of the MyPertamina Champion Gas Station program in the form of awards and mentoring, which is coordinated by RFS and RBS with the stages of dashboard development (Q1), program implementation and coaching (Q2–Q3), as well as review of results and adjustment of targets for the following year in Q4 2026.

CONCLUSION

This study demonstrates that MyPertamina application adoption at gas stations in DKI Jakarta is primarily driven by gas station operational governance particularly Gas Station Type ($r = +0.525$) and Gas Station Class ($r = +0.489$) rather than transaction-level attributes such as product type or time of day. Gradient Boosted Trees emerged as the optimal classification model (accuracy: 97.75%; F1-score MyPertamina: 98.26%; AUC: 0.993), confirming the feasibility of machine learning-based behavioral prediction in energy retail contexts, consistent with recent ML

applications on MyPertamina data. The G-STIC strategic framework translates these findings into a five-tier governance differentiation model, linking model-derived risk scores to targeted programs across gas station segments from end-to-end digital standardization at COCO/Pasti Prima stations to infrastructure gap-filling at DODO/Basic stations. Academically, this study contributes to literature on digital service quality and technology adoption in the energy sector by demonstrating the centrality of supply-side facilitating conditions in driving consumer digital transaction behavior, consistent with recent frameworks on digital adoption in fuel retail.

This study has several limitations that future research should address. First, the 90/10 split without cross-validation may introduce optimistic bias; replication with k-fold cross-validation is recommended, as applied in similar ML-based MyPertamina studies. Second, the single-city scope (DKI Jakarta) limits generalizability; external validation in provinces with lower digital literacy or different gas station network compositions is necessary. Third, sampling bias may exist as stratified proportional sampling preserves the DKI Jakarta class distribution, which may not reflect national patterns. Future research should incorporate multi-province datasets, explore deep learning architectures, and examine longitudinal adoption trajectories particularly in light of Indonesia's evolving fuel subsidy digitalization policy.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to all participants who contributed valuable data for this study. Special thanks to PT Pertamina (Persero) for providing the necessary secondary data and for their support throughout the research. We also acknowledge the contribution of the Universitas Padjajaran, which provided the necessary academic resources and facilities for the successful completion of this project.

AUTHOR CONTRIBUTION STATEMENT

Fresly Leo Chandra Hutapea conceptualized and designed the research, conducted the data analysis, and was responsible for writing the manuscript. Sunu Widiyanto contributed to the methodology, supervised the data collection, and provided critical revisions to the manuscript. Samidi contributed to the data preparation, performed the machine learning modeling, and assisted in writing the results and discussion sections. All authors approved the final manuscript for submission.

REFERENCES

- Aditya, A., Windasari, N. A., & Mussry, J. S. (2025). Enhancing Customer Satisfaction Through Optimized Fast Track Lines: A Case Study Of Pertamina Fuel Stations. *Jurnal Locus Penelitian Dan Pengabdian*, 4(8), 8003–8020. <https://doi.org/10.58344/locus.v4i8.4260>
- Azeroual, O., Nacheva, R., Nikiforova, A., & Störl, U. (2025). A CRISP-DM and predictive analytics framework for enhanced decision-making in research information management systems. *Informatica*, 49(18).
- Cai, W., Wei, R., Xu, L., & Ding, X. (2022). A method for modelling greenhouse temperature using gradient boost decision tree. *Information Processing in Agriculture*, 9(3), 343–354.
- Dash, G., Sharma, K., & Yadav, N. (2023). The diffusion of mobile payments: Profiling the adopters and non-adopters, Roger's way. *Journal of Retailing and Consumer Services*, 71, 103219. <https://doi.org/10.1016/j.jretconser.2022.103219>
- Fähndrich, J. (2023). A literature review on the impact of digitalisation on management control. *Journal of Management Control*, 34(1), 9–65. <https://doi.org/10.1007/s00187-022-00349-4>
- Hastuti, T. D., Indarto, S. L., & Claudia, G. (2024). Implementation of Financial Employment Governance for the Sustainability of the Jem Spotting Business. *Proceeding International Conference of Community Service*, 2(1). <https://doi.org/10.18196/iccs.v2i1.304>
- Indradewa, R., & Iqbal, M. A. (2021). Knowledge Shifting Process in Energy Sector Company (Case Study PERTAMINA an Indonesian Energy Company). *American International Journal of Business Management*, 4(06), 8–17.
- Khataei, A., & Bazargan, K. (2025). Treelut: An efficient alternative to deep neural networks for inference acceleration using gradient boosted decision trees. *Proceedings of the 2025 ACM/SIGDA International Symposium on Field Programmable Gate Arrays*, 14–24.
- Khoiro, A., Ahmadi, H., Wijaya, L., Imran, M. A., Mirdad, A. J., & Sangkereng, I. (2025). Digital

- Transformation in Fuel Payment and Its Impact on User Experience and Loyalty. *2025 4th International Conference on Creative Communication and Innovative Technology (ICCIT)*, 1–7. <https://doi.org/10.1109/ICCIT65724.2025.11167054>
- Kustyana, E., Prasetyo, E., Nurfauziah, S., Azahra, W. A., & Nanda, Z. (2024). Sentiment Analysis of MyPertamina on Google Play Store Using Naïve Bayes for Sustainable Policies. *Jurnal Dialektika: Jurnal Ilmu Sosial*, 22(3), 297–311. <https://doi.org/10.63309/dialektika.v22i3.371>
- Kusuma, S. S., Suryoko, S., & Pradhanawati, A. (2022). Mapping the public-private partnership in digitalization of PT. Pertamina Gas Station in Indonesia. *International Conference on Public Organization (ICONPO 2021)*, 343–349. <https://doi.org/10.2991/aebmr.k.220209.046>
- Lazuardy, M. M., Harahap, E. F., & Habyba, A. N. (2025). User satisfaction analysis of an Innovation on Supply Loss (IOS) system in the commercial and trading sector using the PIECES-usability framework and binary logistic regression. *Journal Industrial Servicess*, 11(2). <https://doi.org/10.62870/jiss.v11i2.32837>
- Madyatmadja, E. D., Kosasih, R., Aulia, A., Sutanto, V., & Joe, R. J. (2025). Indonesia Digital Voice: MyPertamina Sentiment Analysis with Machine Learning. *2025 11th International HCI and UX Conference in Indonesia (CHIuXiD)*, 7–12. <https://doi.org/10.1109/CHIuXiD68326.2025.11323971>
- Murrar, A., Paz, V., Batra, M., & Yerger, D. (2025). Strategies for driving customer adoption of AI-powered mobile apps: insights from structural equation modeling in the water sector. *Journal of Systems and Information Technology*, 27(3), 343–364. <https://doi.org/10.1108/JSIT-12-2023-0335>
- Mushava, J., & Murray, M. (2024). Flexible loss functions for binary classification in gradient-boosted decision trees: An application to credit scoring. *Expert Systems with Applications*, 238, 121876.
- Prastiyo, R. H., & Lahindah, L. (2024). Customer Requirements on Digital Application in Energy Sectors for B2B Customers (a Part of Quality Function Deployment Analysis on My Pertamina for Business). *Interdisciplinary Journal and Hummanity (INJURITY)*, 3(1), 31–48.
- Pratama, R., & Handayani, W. (2025). Analysis of mypertamina service design to improve service quality. *JHSS (JOURNAL OF HUMANITIES AND SOCIAL STUDIES)*, 9(1), 65–71. <https://doi.org/10.33751/jhss.v9i1.49>
- Putra, J. M., & Menorizah, M. (2025). An Analysis of the Use of the MyPertamina Application in Reducing the Misuse of Government-Subsidised Peralite (A Literature Review). *JERIT: Journal of Educational Research and Innovation Technology*, 2(2), 62–72.
- Rousyati, R., Prاتمanto, D., Ardiansyah, A., & Aji, S. (2023). Sentiment Analysis on Fuel Purchase Policy Through MyPertamina Application Using NB and SVM Methods Optimized by PSO as Weight Optimization. *Journal of Applied Informatics and Computing*, 7(2), 240–245. <https://doi.org/10.30871/jaic.v7i2.5131>
- Shimaoka, A. M., Ferreira, R. C., & Goldman, A. (2024). The evolution of CRISP-DM for data science: Methods, processes and frameworks. *SBC Computing Reviews*, 4(1), 28–43.
- Sinurat, A. A., Hendriyani, C., & Damayanti, F. (2022). MyPertamina Application To Increase Consumer Engagement. *The International Journal of Business Review (The Jobs Review)*, 5(1), 65–74.