



Improving Vehicle Payment Method Classification Using XGBoost with SMOTE and SHAP Interpretation

Dedi Trisnawarman¹, Reza Mahendra^{2*}

^{1,2}Department of Information System, Faculty of Information Technology, Universitas Tarumanagara, Jakarta, Indonesia

¹dedit@fti.untar.ac.id, ²reza.825210004@stu.untar.ac.id,

Abstract

*Class imbalance in vehicle payment method classification can cause predictive models to become biased toward the majority. This study aims to build a classification model for automotive consumer payment methods using Extreme Gradient Boosting (XGBoost), with class balancing handled through the Synthetic Minority Over-sampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN), and model interpretability performed using SHAP (SHapley Additive Explanations). The dataset consisted of 11,011 records and 13 attributes derived from Toyota vehicle delivery order transactions. Results show that the XGBoost model without balancing achieved 67.37% accuracy but only 0.24 recall for the Cash class. After applying SMOTE, the recall for the Cash class improved to 0.58, while ADASYN produced a similar improvement at 0.59, with overall accuracy maintained at around 61–62% and a stable ROC-AUC of 0.65. Feature importance and SHAP analysis identified *c_vehicle_model* and *c_city* as the most influential factors in predicting the payment method. From a business perspective, the improved ability to detect cash customers reduces the risk of misclassification and enables dealers to better segment customer payment preference. This supports more effective marketing campaigns, sales strategies, and financing risk management. The combination of XGBoost, SMOTE, ADASYN, and SHAP has proven effective in handling imbalanced data while offering transparent interpretability of predictions, making it a practical foundation for data-driven decision-making in the automotive industry.*

Keywords: explainable ai; payment method classification; SHAP; SMOTE; Xgboost

How to Cite: Dedi Trisnawarman and Reza Mahendra, “Improving Vehicle Payment Method Classification Using XGBoost with SMOTE and SHAP Interpretation”, *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 10, no. 1, pp. 11 - 19, Feb. 2026.

Permalink/DOI: <https://doi.org/10.29207/resti.v10i1.6935>

Received: July 9, 2025

Accepted: January 28, 2026

Available Online: February 15, 2026

*This is an open-access article under the CC BY 4.0 License
Published by Ikatan Ahli Informatika Indonesia*

1. Introduction

The automotive industry is one of the major economic sectors that has experienced significant development from 2021 to 2023. Based on data from the Association of Indonesia Automotive Industries (Gaikindo), motor vehicle sales in Indonesia reached more than 1,005,802 units in 2023, a decrease of 4% from 1,048,040 units in 2022, but this data reflects a large and competitive market potential. One of the crucial aspects of the vehicle sales process is the payment method chosen by consumers, whether in cash or through financing facilities (credit). According to Brown and Mues (2012), early identification of consumer payment preferences can provide strategic advantages in credit risk management, targeted leasing offers, and more accurate market segmentation.

In practice, the selection of payment methods is closely related to factors such as demographic characteristics, vehicle type, and the geographical location of consumers [1]. Therefore, payment method prediction can be formulated as a binary classification problem in the field of machine learning.

One of the main challenges in payment method classification is class imbalance, where credit transactions significantly outnumber cash transactions. This imbalance often causes machine learning models to prioritize majority-class accuracy while underperforming in minority-class detection [2- 4]

Previous studies have shown that imbalanced datasets can bias classification models toward the majority class, leading to poor detection of minority-class instances [2], [5]. In the automotive context, this bias may result

in missed opportunities to identify cash customers with strategic business value.

To address this challenge, advanced machine learning algorithms and data balancing techniques have been widely explored. Extreme Gradient Boosting (XGBoost) has demonstrated strong performance in structured tabular data, while oversampling methods such as SMOTE are commonly used to improve minority-class representation. In addition, explainable AI techniques like SHAP enable transparent interpretation of model predictions [4], [6-9].

Wang et al. (2023) demonstrated that tree-based ensemble models, particularly XGBoost, achieve high predictive performance in credit-related classification tasks. Similarly, Zhang (2024) highlighted the robustness of XGBoost in handling imbalanced payment behavior datasets. However, most existing studies focus primarily on predictive performance without sufficiently addressing model interpretability and comparative balancing strategies [10], [11].

Despite the growing use of machine learning for payment behavior prediction, limited research has simultaneously examined data balancing strategies and model interpretability in the context of automotive payment method classification. Unlike previous studies that primarily emphasize predictive accuracy, this research integrates Extreme Gradient Boosting (XGBoost) with two oversampling techniques (SMOTE and ADASYN) and SHAP-based interpretation to balance predictive performance and explainability [12], [13]. Therefore, this study aims to develop an interpretable classification model for vehicle consumer payment methods using XGBoost, address class imbalance through oversampling techniques, and analyze feature contributions using SHAP based on real-world automotive transaction data.

2. Methods

This study adopts a quantitative research approach using supervised machine learning. The overall research design and experimental workflow are illustrated in Figure 1.

Extreme Gradient Boosting (XGBoost) is employed as the primary classification algorithm. Model performance is evaluated using accuracy, recall, F1-score, and ROC-AUC, with class imbalance handled through Synthetic Minority Over-sampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN). Model interpretability is achieved using SHapley Additive Explanations (SHAP).

The dataset used in this study consists of 11,011 delivery order transaction records of Toyota vehicles obtained from a dealer in Indonesia. It contains 13 attributes representing vehicle characteristics, geographical information, and payment-related variables. The classification task aims to predict the consumer payment method, which is defined as a binary target variable: Cash (0) and Credit (1). The target

variable `payment_method` is derived from the credit attribute, where transactions with a non-empty credit value are labeled as Credit, and those with an empty value are labeled as Cash [1], [12].

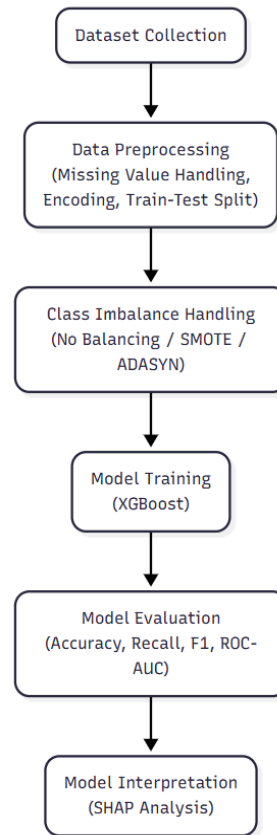


Figure 1. Research design and experimental workflow.

The initial data distribution shows class imbalance, with a proportion of 65.0% for credit and 35.0% for cash. Several attributes have missing values, including.

1. `c_sales_name`: 359 missing.
2. `c_province`: 48 missing.
3. `c_city`: 31 missing.

The target classification attribute is `payment_method`, which is derived from the credit attribute. The value of `payment_method` is set to 'Credit' if the value in the credit column is not empty and 'Cash' if it is empty.

Data preprocessing was conducted to ensure data quality and model compatibility. The process included handling missing values, encoding categorical features using label encoding, and applying a stratified train-test split to preserve class distribution. Records with missing values in essential attributes were removed prior to modeling [14], [15]. Less than 2% of records contained missing values and were removed to ensure data consistency. The proportion of removed data was considered small and did not significantly affect the class distribution. Attributes such as `credit` and `n_tenor` were not treated as conventional missing values, as empty credit entries were used to define the target variable (Cash), and `n_tenor` was only applicable for credit transactions.

To address the class imbalance in the training data, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. SMOTE generates synthetic samples for the minority class by interpolating between existing instances, thereby improving class distribution during model training [2], [16]. Prior studies have shown SMOTE to be effective in increasing model sensitivity to minority classes in financial risk detection and consumer behavior modeling [17- 19].

SMOTE is an oversampling method to address class imbalance by creating synthetic data in the minority class through interpolation between similar data [20], [21]. The SMOTE interpolation formula is shown in Equation 1.

$$x_{new} = x\delta \times (x_{NN} - x) \quad (1)$$

The SMOTE interpolation process is illustrated in Equation 1, where δ represents a random value in the interval [0,1]. In this study, SMOTE was applied only to the training dataset to avoid information leakage, while the test dataset remained unchanged.

The main algorithm used is Extreme Gradient Boosting (XGBoost) is an ensemble learning algorithm based on gradient-boosted decision trees that iteratively improves model performance by minimizing loss functions while incorporating regularization to reduce overfitting. It iteratively improves model performance by minimizing loss functions while incorporating regularization to reduce overfitting [6], [22], [23]. It has been widely adopted in machine learning competitions and applications due to its speed and accuracy. In this study, hyperparameter tuning is performed using GridSearchCV with 5-fold cross-validation, focusing on macro F1-Score to account for class imbalance. Testing the following parameters includes learning_rate [0.01, 0.1, 0.3], max_depth [3, 5, 7], and n_estimators [100, 200]. Evaluation is conducted for both data without SMOTE and data after balancing with SMOTE.

Model performance was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC. Given the imbalanced nature of the dataset, recall for the minority class (Cash) was emphasized as a primary evaluation metric.

Model evaluation was conducted under two conditions: without class balancing and with class balancing applied to the training data [24].

SHAP (SHapley Additive Explanations) was employed to interpret the predictions of the XGBoost model by quantifying the contribution of each feature to the output [8], [4]. Previous studies have shown SHAP to be effective in interpreting predictions in financial and automotive applications [12], [25].

Google Colab, which is based on Python 3.10 and contains the following essential libraries to ensure that the research findings may be reproduced by others, was used to conduct the entire experimental process. Use xgboost (v1.7.6) to model the XGBoost approach; use imbalanced-learn (v0.11.0) to implement SMOTE; and

use shap (v0.43.0) to interpret the model using SHAP. Make use of scikit-learn (v1.2.2) to preprocess, assess, and validate the model.

Table 1 offers a thorough explanation of the data structure and variables, even though the dataset is confidential and originates from internal transactions at auto dealers.

The scikit-learn pipeline API and a validation technique called StratifiedKFold have been used to redesign the modeling pipeline. This makes it easier to replicate by maintaining the same class proportions in the test data. On request, the author can email the source code.

Table 1. Data Profile

Attribute	Data Type	Missing
c_vehicle_model	object	0
c_color_name	object	0
c_province	object	48
c_city	object	31
c_segmentation	object	12
c_class	object	0
credit	object	3854
cash	Float64	11011
c_leasing_name	object	3854
c_sales_name	object	359
d_so_date	datetime	0
n_tenor	float64	7041
c_district	object	32
payment_method	target	0

3. Results and Discussions

3.1 Exploratory Data Analysis and Class Distribution

Initial analysis was conducted on the dataset to understand the characteristics of class distribution and data completeness. The dataset consists of 11,011 records and 13 attributes, including information on vehicle type, color, consumer location, and payment method.

The initial distribution of the target variable payment_method shows a significant class imbalance, where in Table 2, the proportion of credit reaches 65.0%, while cash only accounts for 35.0%. This distribution is visualized in Figure 2.

Table 2. Distribution of Proportion

Payment Method	Distribution
Cash	35.0%
Credit	65.0%

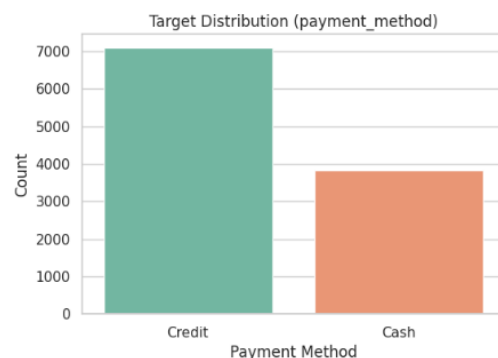


Figure 2. Target Distribution

This imbalance has the potential to cause a model prediction bias toward the majority class, so balancing strategies such as SMOTE are needed during the modeling stage. To ensure data quality before the model training process, we conducted a missing value analysis. Table 3 presents the results of the examination of the missing values in each column.

Table 3. Count of Missing Value

Column	Missing
c_vehicle_model	0
c_color_name	0
c_province	0
c_city	0
c_segmentation	0
c_class	0
payment_method	0

After preprocessing and data cleaning, no missing values remain in the variables used for classification, as missing entries were handled during preprocessing.

Next, we carried out an analysis of the correlation between numerical features. The correlation heatmap results are shown in Figure 3, where the features c_segmentation and c_class have a relatively high correlation of 0.74.

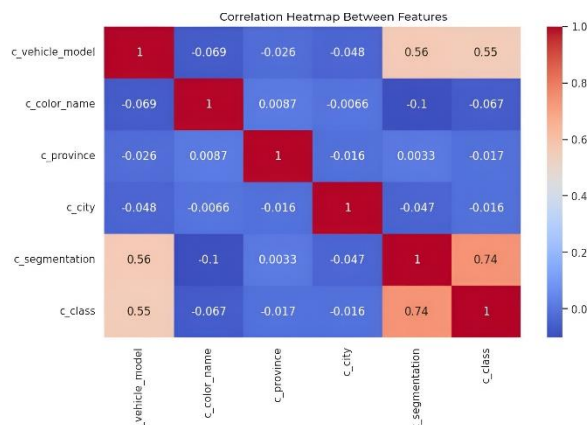


Figure 3. Correlation Between Features

3.2 Evaluation Without SMOTE

Model evaluation was conducted by training three classification algorithms, namely Extreme Gradient Boosting (XGBoost), Random Forest, and Logistic Regression, using a dataset without class balancing.

The purpose of this comparison is to determine the performance of each model under imbalanced data conditions, as well as to identify the best baseline model to be used in the next stage after applying the SMOTE balancing technique. The performance results of the three models are presented in Figure 4.

Based on the results, the XGBoost model has the highest accuracy and ROC-AUC, which are 67.37% and 0.65, respectively. However, the recall for the Cash class is still low at 0.24, indicating that the model tends to be biased toward the majority class, Credit.

The Random Forest model shows similar performance, with an accuracy of 65.95% and a recall for the Cash

class of 0.28, slightly better than XGBoost in detecting Cash but still not optimal.

In contrast, logistic regression exhibits the lowest performance, with an accuracy of 65.08% and a recall for the Cash class of only 0.01, indicating that this model is highly biased towards the majority class. This is due to the characteristic of logistic regression being sensitive to imbalanced class distributions [24].

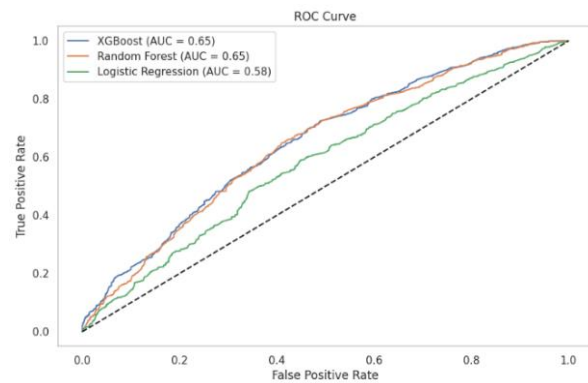


Figure 4. ROC 3 Model

The confusion matrix of the baseline XGBoost model is shown in Figure 5, while the ROC curve and precision-recall curve are presented in Figures 6 and 7, respectively.

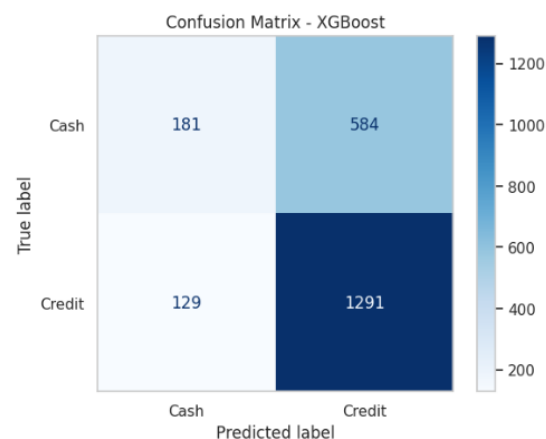


Figure 5. Confusion Matrix - XGBoost

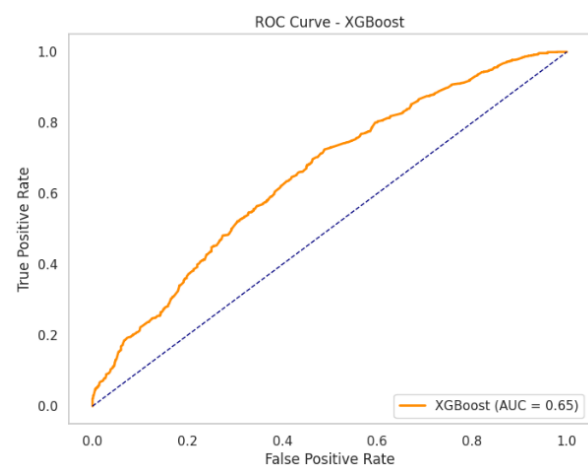


Figure 6. ROC Curve - XGBoost

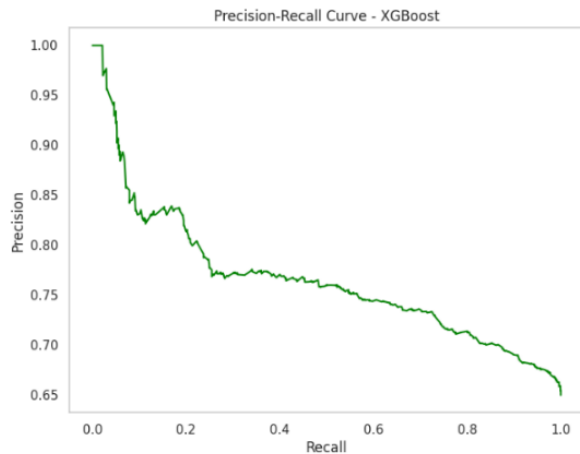


Figure 7. Precision-Recall Curve - XGBoost

Based on this assessment, it can be said that while XGBoost performs better overall than the other two models, data imbalance prevents any of the three algorithms from detecting the Cash class as effectively. To improve class distribution in the following stage, the SMOTE approach was implemented [2].

The hyperparameter tuning method produced the optimal settings with a learning_rate of 0.1, max_depth of 5, and n_estimators of 200 in addition to model evaluation.

The XGBoost model sans SMOTE obtained a Best Cross-Validation Macro F1-Score of 0.5518 with the set of parameters. This finding suggests that the model's capacity to manage unbalanced data remains restricted, particularly with regard to the minority class.

3.3 Evaluation With SMOTE

After identifying that the baseline XGBoost model was biased toward the majority class Credit, SMOTE was applied to balance the training [26], [2], resulting in an equal class distribution during model training.

The XGBoost model was retrained using the SMOTE-balanced training data and evaluated on the original test set with an unaltered class distribution. The results are presented in Table 4.

The evaluation results show a substantial improvement in detecting the minority class. Specifically, recall for Cash increased from 0.24 (baseline XGBoost) to 0.58 with SMOTE, meaning the model can correctly identify more than twice as many Cash customers. However, this improvement comes with a trade-off: overall accuracy decreased slightly from 67.37% to 61.74%.

Table 4. Model Performance + SMOTE

	Precision	Recall	F1-Score	Support
Cash	0.46	0.58	0.52	765
Credit	0.74	0.64	0.68	1420
Accuracy			0.62	2185

From a business perspective, this trade-off is acceptable because the main goal is to capture minority-class consumers (cash payments) more effectively, which is crucial for targeted marketing and financial risk

management. The re-tuning process produced the best hyperparameters (learning_rate = 0.1, max_depth = 5, n_estimators = 100), yielding a cross-validation macro F1-score of 0.5937, higher than the baseline model.

The confusion matrix of the SMOTE-based model is presented in Figure 8, while Figures 9 and 10 illustrate the ROC and precision-recall curves. These results confirm that applying SMOTE improves the model's ability to detect the minority Cash class while maintaining balanced performance across both classes.

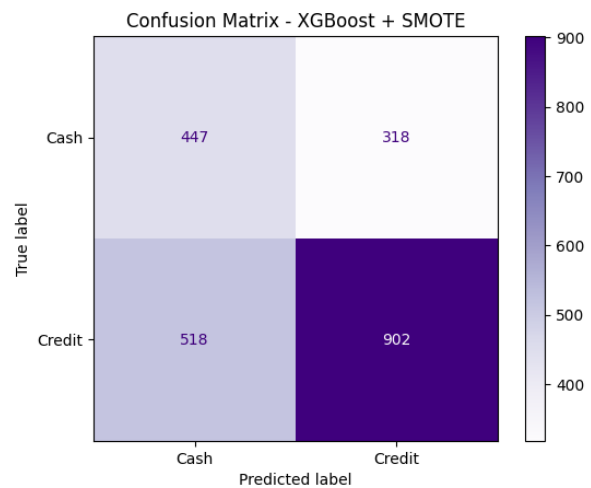


Figure 8. Confusion Matrix – XGBoost + SMOTE

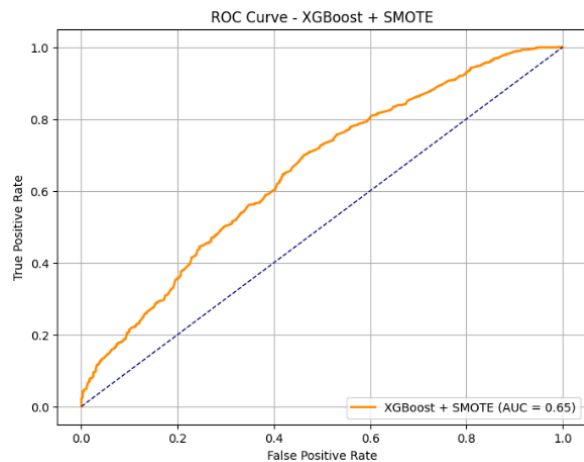


Figure 9. ROC Curve – XGBoost + SMOTE

It should be noted that the ROC-AUC value of the model after SMOTE remains stable at 0.65, indicating that data balancing does not significantly affect the trade-off between true positive rate (TPR) and false positive rate (FPR). However, the improvement in recall for the minority class is the main benefit derived from the implementation of XGBoost + SMOTE.

These results agree with what Chawla et al. [2] found, which says that SMOTE helps the model pay more attention to the minority class without greatly reducing the ROC-AUC.

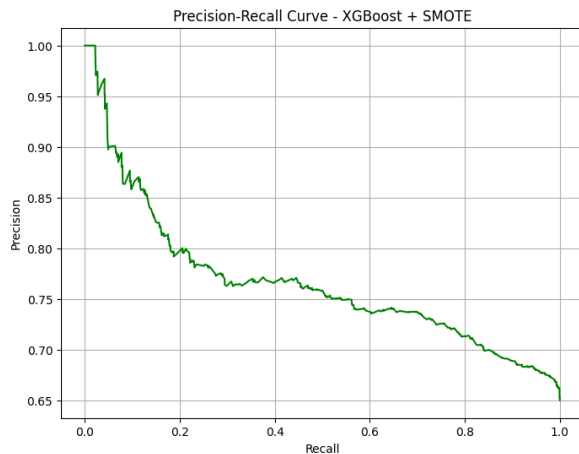


Figure 10. Precision-Recall Curve - XGBoost + SMOTE

3.4 Evaluation with ADASYN

In addition to SMOTE, the Adaptive Synthetic Sampling (ADASYN) technique was also applied to address the class imbalance problem. Unlike SMOTE, which generates synthetic samples uniformly, ADASYN focuses on creating more synthetic data in regions where the minority class is harder to learn, thereby adaptively shifting the decision boundary toward these challenging areas [3].

The XGBoost model was retrained on the ADASYN-balanced dataset, and evaluation was conducted on the same original test set. The results are shown in Table 5.

Table 5. Performance Comparison (XGBoost + ADASYN)

	Precision	Recall	F1-Score	Support
Cash	0.46	0.59	0.52	765
Credit	0.74	0.62	0.68	1420
Accuracy			0.61	2185

The results demonstrate that ADASYN produces performance comparable to SMOTE. Recall for the Cash class improved significantly from the baseline (0.24) to 0.59, slightly higher than SMOTE (0.58). Accuracy was 61.2%, which is very close to the SMOTE result (61.7%).

This suggests that both balancing methods are effective in improving minority-class detection, though the improvement in Cash recall with ADASYN is marginally better. From a business perspective, this improvement indicates that ADASYN may help dealers identify more cash customers compared to SMOTE, albeit with nearly identical accuracy trade-offs.

Hyperparameter tuning for ADASYN also resulted in the same optimal configuration (learning_rate = 0.1, max_depth = 5, n_estimators = 100), with a cross-validation macro F1-score of 0.5913, which is very close to the SMOTE-tuned model.

The evaluation results for ADASYN are summarized in Table 5. Since the performance is similar to SMOTE, the detailed confusion matrix and curves are omitted for brevity. These results confirm that ADASYN is a viable alternative to SMOTE in addressing class imbalance,

offering a small but consistent gain in minority-class recall.

In addition to the tabular results, a direct comparison between SMOTE and ADASYN across key metrics is presented in Figure 11. The figure shows that the two balancing techniques perform very similarly, with ADASYN yielding slightly higher accuracy and macro F1-score, while SMOTE produces a marginally higher recall for the Cash class.

This visual comparison reinforces the conclusion that both balancing methods are effective, with the choice between them depending on business priorities. For example, if improving Cash recall is the main objective, SMOTE may be preferred, whereas ADASYN offers a slightly better balance across metrics.

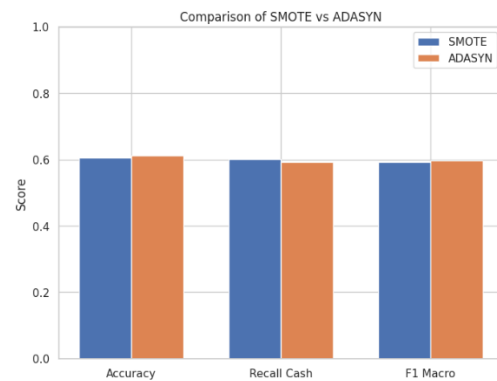


Figure 11. Comparison of SMOTE vs ADASYN

3.5 Feature Importance

Following the implementation of SMOTE, a feature importance analysis of the XGBoost model was carried out in order to comprehend the relative contribution of each feature in forecasting customer payment methods. The degree to which a feature aids in impurity reduction at each node of the XGBoost ensemble tree is determined by this analysis. Figure 12 displays the feature significance visualization's findings.

The most significant characteristic, according to the results, is c_vehicle_model, which contributes about 23% to the model's predictions. With an importance of around 21%, c_city comes next, followed by c_class at 17%, c_color_name and c_segmentation at roughly 16% each, and c_province at roughly 7%.

Hierarchically, the primary determinants in the model are the attributes that are directly associated with the consumer's geographic location (c_city) and car type (c_vehicle_model). This makes sense given that the type of vehicle is frequently associated with market groups and preferred payment methods, but the consumer's residence may have an impact on local preferences or dealer policies.

These results are consistent with Burez and Van den Poel's research [1], which found that consumer demographics and product features play a significant role in financial behaviour prediction models. SHAP-

based explainable AI study further supports the findings of this feature relevance [8], [27].

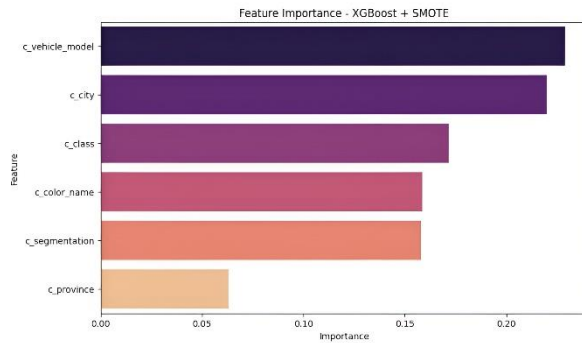


Figure 12. Feature Importance – XGBoost + SMOTE

3.6 SHAP Result

The contribution of each feature to the prediction output of the XGBoost model following data balancing using SMOTE is measured in this study using the SHAP (SHapley Additive Explanations) approach as the model interpretation step. The SHAP value theory in cooperative game theory, which determines the average marginal contribution of each characteristic to the prediction output, serves as the foundation for this approach [8], [19].

According to Figure 13, the features that contribute most to the model's prediction output are c_city and c_vehicle_model. The large range of both positive and negative values for the SHAP value of c_city suggests that, depending on its particular value, it can have an impact on the likelihood of correctly identifying the credit or cash class.

While c_province has the least impact, other characteristics like c_color_name, c_class, and c_segmentation contribute moderately.

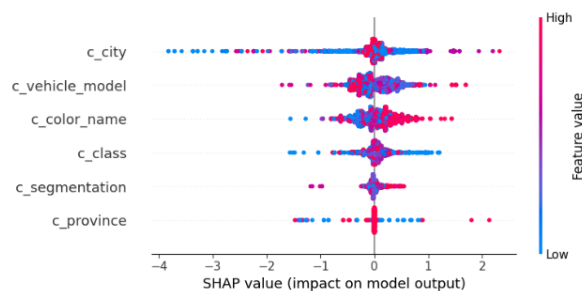


Figure 13. SHAP Summary Plot

A waterfall plot for a single prediction with a model expectation value of $E[f(x)] = -0.003$ is displayed in Figure 14. In this instance: (1).The primary factor driving the prediction in the direction of the credit class is the feature c_city, which has a value of +0.15. (2). The characteristic c_color_name has a negative contribution of -0.14, which lowers the likelihood of a credit prediction. (3). With corresponding contributions of -0.09, -0.08, and -0.06, the features c_class, c_vehicle_model, and c_segmentation all have a negative impact. (4). The prediction is barely influenced by c_province.

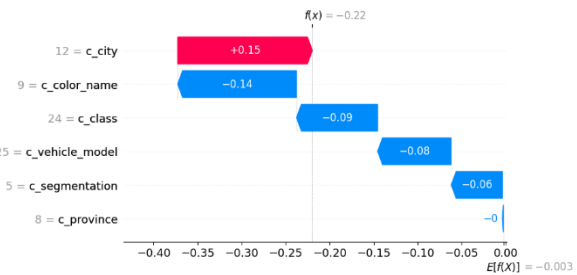


Figure 14. SHAP Waterfall Plot

This illustration demonstrates how the model makes additive predictions, with each feature having the capacity to raise or lower the final prediction probability. In addition to indicating which aspects are significant, this SHAP-based interpretation also indicates which way those features affect the prediction result.

The SHAP analysis's findings corroborate the earlier feature importance findings, which show that c_vehicle_model and c_city are the model's primary determining factors. Furthermore, SHAP offers detailed interpretations at the individual prediction level, which can be applied to data-driven business decision-making that is more accountable and transparent [27].

3.7 Discussion

The experimental results indicate that the class distribution imbalance in the initial dataset causes the three baseline models, particularly XGBoost, Random Forest, and Logistic Regression, to tend to be biased towards the majority class, Credit. This bias is evidenced by the very low recall value for the Cash class, especially in logistic regression with a recall of only 0.01, while XGBoost only reaches 0.24.

The application of SMOTE as a balancing method successfully improved the model's performance on the minority class. The recall of the Cash class in the XGBoost model significantly increased from 0.24 to 0.58, although accompanied by a decrease in accuracy from 67.37% to 61.74%. Although the application of SMOTE slightly decreased accuracy, this trade-off is considered reasonable in imbalanced classification [24], [13], and is prioritized if recall for the minority class is deemed more valuable from a business perspective. This trade-off is still acceptable in the context of business decision support, where the priority is more on the ability to detect as many cash customers as possible rather than maintaining overall accuracy [2], [5]. The stable ROC-AUC value of 0.65 both before and after SMOTE indicates that balancing does not significantly reduce the trade-off between True Positive Rate and False Positive Rate.

To further evaluate balancing strategies, ADASYN was also tested. The results show that ADASYN produces performance very similar to SMOTE, with recall for Cash reaching 0.59 and accuracy of 61.2%. Compared to SMOTE, ADASYN provides slightly higher overall accuracy and macro F1, while SMOTE yields

marginally higher recall for the Cash class. The summary comparison of SMOTE and ADASYN (Table 5 and Figure 11) illustrates that the two methods are nearly equivalent, and the choice between them may depend on business priorities: if detecting more Cash customers is critical, SMOTE offers a minor advantage, whereas ADASYN provides a slightly better balance across metrics.

From the perspective of model interpretation, the results of feature importance and SHAP consistently show that the attributes `c_vehicle_model` and `c_city` are the main determinants in influencing consumers' payment method decisions. These results reinforce the findings of Burez and Van den Poel [1], which state that product characteristics and customer demographics play a significant role in payment behavior in the financial sector. The application of explainable AI through SHAP also provides strategic advantages for automotive companies, as it reveals how well the model performs and helps understand the key factors influencing the model's predictions. Designing more effective sales strategies based on vehicle segmentation and consumer geographical location can benefit from this [8], [28], [29], [30].

From the results of this research, it can be concluded that the combination of XGBoost, balancing methods (SMOTE and ADASYN), and SHAP interpretation is an effective and transparent approach for building a prediction system for automotive consumer payment methods based on historical transaction data [20], [21], [31], [32], [33]. However, this research has several limitations, including the limited number of features and the relatively modest accuracy despite improvements in recall. Further research is expected to explore variations in balancing methods such as SMOTE-ENN or SMOTE-Tomek [26], advanced feature engineering, and ensemble techniques to achieve a better balance between accuracy and recall. In addition, testing across datasets from multiple automotive brands would enhance the generalizability of the findings.

4. Conclusions

This study examined the classification of vehicle payment methods using XGBoost combined with balancing techniques and model interpretability. The baseline XGBoost model achieved good overall accuracy (67.37%) but performed poorly in detecting the minority Cash class (recall = 0.24), confirming that the imbalance problem led to strong bias toward the majority Credit class.

To address this issue, two balancing methods were applied. With SMOTE, the recall of Cash customers improved to 0.58, while ADASYN further increased recall slightly to 0.59, both accompanied by a reduction in overall accuracy to around 61–62%. Although accuracy decreased compared to the baseline, this trade-off is acceptable in the context of imbalanced classification, especially when the business priority is

to identify as many cash customers as possible. From a practical perspective, these results provide valuable insights for automotive dealers, enabling more accurate segmentation of payment preferences and supporting targeted marketing strategies and risk management.

The contribution of this research lies in demonstrating that combining XGBoost, balancing techniques (SMOTE and ADASYN), and explainable AI (SHAP) yields an effective and transparent prediction system for automotive consumer payment methods. The use of SHAP provides interpretability by showing which features, such as `c_vehicle_model` and `c_city`, influence model predictions, thereby supporting more informed and data-driven business decisions.

Acknowledgements

The author declares no conflict of interest. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

- [1] J. Burez and D. Van den Poel, "Handling class imbalance in customer churn prediction," *Expert Syst. Appl.*, vol. 36, no. 3, Part 1, pp. 4626–4636, 2009, doi: <https://doi.org/10.1016/j.eswa.2008.05.027>.
- [2] 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: <https://doi.org/10.1613/jair.953>.
- [3] G. M. Weiss, "Mining with rarity: a unifying framework," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 7–19, Jun. 2004, doi: <https://doi.org/10.1145/1007730.1007734>.
- [4] M. Altalhan, A. Algarni, and M. Turki-Hadj Alouane, "Imbalanced Data Problem in Machine Learning: A Review," *IEEE Access*, vol. 13, pp. 13686–13699, 2025, doi: [10.1109/ACCESS.2025.3531662](https://doi.org/10.1109/ACCESS.2025.3531662).
- [5] F. Sağlam and M. A. Cengiz, "A novel SMOTE-based resampling technique through noise detection and the boosting procedure," *Expert Syst. Appl.*, vol. 200, no. 9, p. 117023, Aug. 2022, doi: [10.1016/j.eswa.2022.117023](https://doi.org/10.1016/j.eswa.2022.117023).
- [6] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13–17-August-2016, pp. 785–794, Aug. 2016, doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).
- [7] P. Zhang, Y. Jia, and Y. Shang, "Research and application of XGBoost in imbalanced data," *Int. J. Distrib. Sens. Netw.*, vol. 18, no. 6, pp. 1897–1906, Jun. 2022, doi: [10.1177/15501329221106935](https://doi.org/10.1177/15501329221106935).
- [8] Z. Li, "Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost," *Comput. Environ. Urban Syst.*, vol. 96, Sep. 2022, doi: [10.1016/j.compenvurbysys.2022.101845](https://doi.org/10.1016/j.compenvurbysys.2022.101845).
- [9] A. Jude and J. Uddin, "Explainable Software Defects Classification Using SMOTE and Machine Learning," *Annals of Emerging Technologies in Computing*, vol. 8, no. 1, pp. 35–49, 2024, doi: [10.33166/AETIC.2024.01.00](https://doi.org/10.33166/AETIC.2024.01.00).
- [10] C. Wang, H. Jiang, X. Jin, and Z. Zhou, "Customer credit rating by machine learning," *BCP Business & Management*, vol. 36, pp. 122–130, 2023, doi: <https://doi.org/10.54691/bcpbm.v36i.3490>.
- [11] X. Zhang, "Machine learning insights into digital payment behaviors and fraud prediction," *Applied and Computational Engineering*, vol. 77, pp. 55–66, 2024, doi: [10.54254/2755-2721/77/2024MA0066](https://doi.org/10.54254/2755-2721/77/2024MA0066).
- [12] Y. Kong, Y. Wang, S. Sun, and J. Wang, "XGB and SHAP credit scoring model based on Bayesian optimization,"

- Journal of Computing and Electronic Information Management*, vol. 10, no. 1, pp. 46–53, Feb. 2023, doi: [10.54097/jceim.v10i1.5368](https://doi.org/10.54097/jceim.v10i1.5368).
- [13] J. West and M. Bhattacharya, “Intelligent financial fraud detection: A comprehensive review,” *Comput. Secur.*, vol. 57, pp. 47–66, 2016, doi: <https://doi.org/10.1016/j.cose.2015.09.005>.
- [14] V. Çetin and O. Yıldız, “A comprehensive review on data preprocessing techniques in data analysis,” *Pamukkale University Journal of Engineering Sciences*, vol. 28, no. 2, pp. 299–312, 2022, doi: [10.5505/pajes.2021.62687](https://doi.org/10.5505/pajes.2021.62687).
- [15] B. J. Jansen, K. K. Aldous, J. Salminen, H. Almerikhi, and S. Jung, “Data Preprocessing,” 2024, pp. 65–75. doi: [10.1007/978-3-031-41933-1_6](https://doi.org/10.1007/978-3-031-41933-1_6).
- [16] Q. Zheng, C. Yu, J. Cao, Y. Xu, Q. Xing, and Y. Jin, “Advanced Payment Security System: XGBoost, LightGBM and SMOTE Integrated,” Jun. 2024, [Online]. Available: <http://arxiv.org/abs/2406.04658>
- [17] Z. Zhao and T. Bai, “Financial Fraud Detection and Prediction in Listed Companies Using SMOTE and Machine Learning Algorithms,” *Entropy*, vol. 24, no. 8, p. 978, Aug. 2022, doi: [10.3390/e24081157](https://doi.org/10.3390/e24081157).
- [18] I. Aruleba and Y. Sun, “Effective Credit Risk Prediction Using Ensemble Classifiers With Model Explanation,” *IEEE Access*, vol. 12, no. 2, pp. 115015–115025, 2024, doi: [10.1109/ACCESS.2024.3445308](https://doi.org/10.1109/ACCESS.2024.3445308).
- [19] P. Thanathamathée, S. Sawangarereerak, S. Chantamunee, and D. N. M. Nizam, “SHAP-Instance Weighted and Anchor Explainable AI: Enhancing XGBoost for Financial Fraud Detection,” *Emerging Science Journal*, vol. 8, no. 6, pp. 2404–2430, Dec. 2024, doi: [10.28991/ESJ-2024-08-06-016](https://doi.org/10.28991/ESJ-2024-08-06-016).
- [20] Y. Han, Z. Wei, and G. Huang, “An imbalance data quality monitoring based on SMOTE-XGBOOST supported by edge computing,” *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: [10.1038/s41598-024-60600-x](https://doi.org/10.1038/s41598-024-60600-x).
- [21] S. Wang, Y. Dai, J. Shen, and J. Xuan, “Research on expansion and classification of imbalanced data based on SMOTE algorithm,” *Sci. Rep.*, vol. 11, no. 1, Dec. 2021, doi: [10.1038/s41598-021-03430-5](https://doi.org/10.1038/s41598-021-03430-5).
- [22] J. Brownlee, “How to configure XGBoost for imbalanced classification,” *Machine Learning Mastery*, Aug. 2020, Accessed: Jun. 16, 2025. [Online]. Available: <https://machinelearningmastery.com>
- [23] A. Maurya, A. Kumar, and S. Prakash, “Credit Card Fraud Detection using XGBoost Classifier with a Threshold Value,” Jun. 2022, doi: [10.21203/rs.3.rs-1722294/v1](https://doi.org/10.21203/rs.3.rs-1722294/v1).
- [24] T. Saito and M. Rehmsmeier, “The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets,” *PLoS One*, vol. 10, no. 3, Mar. 2015, doi: [10.1371/journal.pone.0118432](https://doi.org/10.1371/journal.pone.0118432).
- [25] R. Rodríguez-Pérez and J. Bajorath, “Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions,” *J. Comput. Aided. Mol. Des.*, vol. 34, no. 10, pp. 1013–1026, Oct. 2020, doi: [10.1007/s10822-020-00314-0](https://doi.org/10.1007/s10822-020-00314-0).
- [26] P. C. Y. Cheah, Y. Yang, and B. G. Lee, “Enhancing Financial Fraud Detection through Addressing Class Imbalance Using Hybrid SMOTE-GAN Techniques,” *International Journal of Financial Studies*, vol. 11, no. 3, p. 12, Sep. 2023, doi: [10.3390/ijfs11030110](https://doi.org/10.3390/ijfs11030110).
- [27] Y. Chen, R. Calabrese, and B. Martin-Barragan, “Interpretable machine learning for imbalanced credit scoring datasets,” *Eur. J. Oper. Res.*, vol. 312, no. 1, pp. 357–372, Jan. 2024, doi: [10.1016/j.ejor.2023.06.036](https://doi.org/10.1016/j.ejor.2023.06.036).
- [28] S. M. Lundberg, P. G. Allen, and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017, doi: [10.48550/arXiv.1705.07874](https://doi.org/10.48550/arXiv.1705.07874).
- [29] Y. Wang, Y. Jia, Y. Zhong, J. Huang, and J. Xiao, “Balanced incremental deep reinforcement learning based on variational autoencoder data augmentation for customer credit scoring,” *Eng. Appl. Artif. Intell.*, vol. 122, p. 106056, Jun. 2023, doi: [10.1016/j.engappai.2023.106056](https://doi.org/10.1016/j.engappai.2023.106056).
- [30] D. Vijayanand and G. S. Smrithy, “Explainable AI - enhanced ensemble learning for financial fraud detection in mobile money transactions,” *Intelligent Decision Technologies*, vol. 19, no. 1, pp. 52–67, Jan. 2025, doi: [10.1177/18724981241289751](https://doi.org/10.1177/18724981241289751).
- [31] K. Handayani et al., “Comparison of XGboost, Extra Trees, and LightGBM with SMOTE for Fetal Health Classification,” 2024, doi: <https://doi.org/10.32520/stmsi.v13i3.3646>.
- [32] M. Imani, A. Beikmohammadi, and H. R. Arabnia, “Comprehensive Analysis of Random Forest and XGBoost Performance with SMOTE, ADASYN, and GNUS Under Varying Imbalance Levels,” *Technologies (Basel)*, vol. 13, no. 3, Mar. 2025, doi: [10.3390/technologies13030088](https://doi.org/10.3390/technologies13030088).
- [33] M. Yusoff, Y. Mahmud, P. A. R. Azmi, and M. T. M. Sallehud-din, “The improvement of SMOTE-ENN-XGBoost through Yeo Johnson strategy on Dissolved Gas Analysis dataset,” *Energy Reports*, vol. 13, pp. 6281–6290, Jun. 2025, doi: [10.1016/j.egyr.2025.05.013](https://doi.org/10.1016/j.egyr.2025.05.013).