International Journal Of Humanities Education And Social Sciences ***(IJHESS)*** *E-ISSN:* **2808-1765**
Volume 5, Number 6, June 2026, *Page. 3105 – 3115*
Email : editorijhess@gmail.com

Comparative Analysis of XGBoost and LightGBM Algorithms in Credit Card Fraud Detection: Cost Sensitivity and Computational Complexity

Nefzah Atirah Qalby[1)], Achmad Fitro[2)], Achmad Kautsar[3)], Riska Dhenabayu[4)]
[1,2,3,4)]Digital Business/Faculty Of Economics And Business, State University of Surabaya, Indonesia

*Corresponding Author
Email: nefzah.22059@mhd.unesa.ac.id

Abstract

Cybercrime in credit card transactions inflicts severe financial damage, with extreme class imbalances often biasing conventional models toward high false positive rates. This study compares and optimizes tree-ensemble algorithms XGBoost and LightGBM to develop a Fraud Detection model that is statistically accurate, computationally efficient, and minimizes real banking financial losses. Utilizing a dataset of 23,769 transactions with a 270:1 imbalance ratio, both models were optimized via the Tree-structured Parzen Estimator and validated using 5-Fold Stratified Cross-Validation. Performance was evaluated through classification metrics, computational efficiency, and the Expected Cost of Misclassification, while Explainable AI via SHAP values ensured model transparency. Results demonstrate LightGBM's superiority, achieving perfect precision (1.000) and an F1-Score of 0.9714, effectively minimizing financial losses to Rp5,000,000. Although XGBoost trained faster, LightGBM's 60-millisecond latency meets real-time standards, providing a robust, transparent risk mitigation system for banking operations. The implementation of this architecture significantly enhances the competitiveness of IT efficiency and banking risk governance in the digital era.

Keywords: Fraud Detection, XGBoost, LightGBM, Expected Cost of Misclassification, Explainable AI.

INTRODUCTION

According to Widjaja, (2026), digitalization of payment systems has become a modern economy pillar, though Lawson & Nancy, (2024) warn it opens new cybercrime gaps. Ahsan et al., (2024) note electronic transaction surges correlate with $39.7 billion in projected global fraud losses, while Widjaja, (2026) emphasizes information asymmetry weakens consumer security literacy. In the Indonesian context, prominent financial crises like the BSI ransomware incident CNBC Indonesia, (2023) and the Rp204 billion dormant account exploitation (CNN Indonesia, 2025; PPATK, 2025), underscore the critical need for automated detection to mitigate manual validation failures. Traditional rulebased infrastructures prove inadequate, generating 80% false positives that disrupt legitimate transactions (Lawson & Nancy, 2024), making Machine Learning (ML) essential for real time anomaly detection (Forough & Momtazi, 2021)

Practitioners often implement Deep Learning, though Ahsan et al., (2024) note neural networks struggle with complex fraud patterns. Shwartz-Ziv & Armon, (2022) demonstrate Deep Learning underperforms on tabular banking data due to high computational costs. Tree-ensemble algorithms, by contrast, offer superior performance-to-cost ratios. Forough & Momtazi, (2021) explain Gradient Boosting effectively minimizes overfitting while handling non-linear relationships. Tayebi & El Kafhali, (2025) validate XGBoost's capability in high-dimensional fraud detection.

Ensemble algorithms falter against extreme class imbalance. Kurniawan et al., (2025) highlight fraudulent transactions average only 0.17%, biasing models toward false positives. While SMOTE synthesizes minority samples, Kurniawan et al., (2025) warn its increased sensitivity causes overfitting and reduced precision. Forough & Momtazi, (2021) note overly-sensitive models block legitimate transactions, triggeringtr customer churn.

This study adopts algorithm-level cost-sensitive approaches such as scale_pos_weight in XGBoost and class_weight='balanced' in LightGBM. Fernández et al., (2018) demonstrate this approach stabilizes performance for extreme datasets without data duplication. Darmawan et al., (2026) support practical banking implementation. Yet Abdelghafour et al., (2024), Xiao et al., (2025), and Xu et al., (2023) show conventional ensemble optimization approaches neglect the 1:100 cost-sensitivity disparity in real financial losses. The asymmetry is stark as Widjaja, (2026) emphasizes undetected fraud causes losses up to millions of rupiah, while false alarms cost merely IDR 50,000 in administrative fees. This disparity drives Expected Cost of Misclassification (ECM) approaches. Xiao et al., (2025) explain ECM imposes penalties proportional to financial consequences, and Höppner et al., (2022) stress banking fraud must be treated as instance-dependent cost-sensitive classification.

Vanderschueren et al., (2022) frame this as predict-and-optimize, adjusting probabilities for rupiah-weighted losses. Ariza-Garzón et al., (2024) confirm cost-configured XGBoost maximizes bank ROI through selective blocking. However, XGBoost's computational costs increase with data volume. LightGBM offers complementary efficiency with sub-50ms latency, creating a strong architectural foundation for production fraud detection.

Both algorithms remain vulnerable to concept drift from evolving consumer behavior and fraud tactics. Kurniawan et al., (2025) identify this phenomenon's impact and advocate continual learning frameworks for real-time parameter updates, preventing catastrophic forgetting. Banking IT investments thus remain viable despite strict transparency regulations. Liu et al., (2026) recommend Explainable AI (SHAP) for decision transparency, while Tao et al., (2024) highlight efficient Tree-structured Parzen Estimator (TPE) hyperparameter optimization.

Recent literature emphasizes comparative machine learning, where Nugroho et al., (2026) show ensemble combinations outperform single algorithms; Zanah et al., (2025) demonstrate ensemble importance for imbalanced data; Husin et al., (2025) validate standardized CRISP-DM pipelines for anomaly detection. This study rejects blind statistical accuracy pursuit, instead comparing XGBoost and LightGBM across three integrated dimensions: classification performance, computational efficiency, and financial loss minimization (ECM). Multidisciplinary comparison strengthened by XAI accountability yields practical solutions enhancing IT efficiency and banking risk governance.

The Concept of Financial Fraud and Cost-Sensitive Learning Conceptually

Credit card fraud involves highly deviant patterns causing severe financial harm (Hajek et al., 2026). Conventional evaluations assume equal risk; however, Statistical Decision Theory (Wald, 1950) reveals real-world decisions carry asymmetric losses (Elkan & Charles, (2001). Cost Sensitive Learning integrates loss matrices into model evaluation, underpinning the Expected Cost of Misclassification (ECM) metric, which imposes heavier penalties on False Negatives than False Positives to minimize aggregate material loss. Integrating cost matrices with gradient boosting consistently outperforms data-level resampling in fraud detection detection (Fernández et al., 2018; Vanderschueren et al., 2022).

Gradient Boosting and Greedy Function Approximation Architectures

XGBoost (v3.2.0) and LightGBM (v4.6.0) are Gradient Boosting Decision Trees (GBDT) rooted in Friedman, (2001) Greedy Function Approximation. They sequentially build models by combining shallow decision trees to incrementally minimize residual errors via gradient descent (Chen & Guestrin, 2016 dan Ke et al., 2017). This mechanism rapidly explores nonlinear relationships in tabular data (Velarde et al., 2023), an architecture dominating global machine learning competitions (Grinsztajn et al., 2022).

Explainable AI and Cooperative Game Theory

The black-box nature of ensemble trees complicates banking forensic audits and transparency regulations (Xu et al., 2023). This study utilizes the SHapley Additive exPlanations (SHAP) framework (Lundberg & Lee, 2017), rooted in Shapley, (1952) Cooperative Game Theory. SHAP has become the de facto standard for explainable AI in finance, calculating marginal feature contributions to transparently decipher transaction-blocking decisions (Liu et al., 2026). Mathematically, the contribution of the ith feature (denoted as ϕ_i) is calculated using the following Shapley value formula:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(M-|S|-1)!}{M!} [f(S \cup \{i\}) - f(S)] \ldots . (1)$$

Where N is the set of all features, M is the total number of features, S is the subset of contributing features, and $f(S)$is the model prediction for the subset S.

RESEARCH METHODS

This research applies a quantitative experimental approach Sugiyono, (2023) to compare and optimize machine learning algorithms for financial anomaly detection. The study implements and optimizes XGBoost and LightGBM algorithms within a rigorously controlled Google Colab environment (Python 3.12, 12.7 GB virtual RAM). The experimental architecture integrates systematic operational methods with the theoretical foundations previously discussed. Operational stages are presented in Figure 1

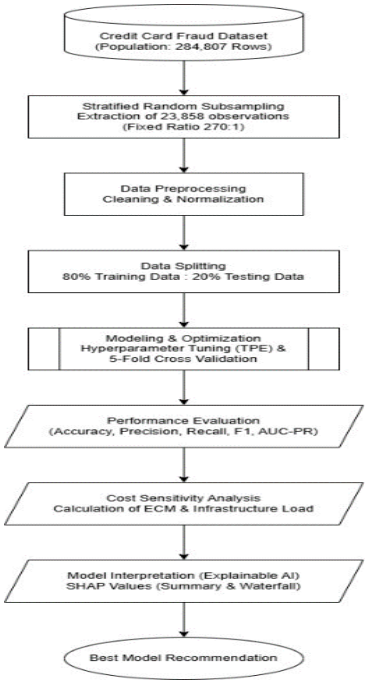

Figure 1 System Overview

The European Credit Card Fraud Detection dataset (284,807 observations; 30 features) serves as a global benchmark; its PCA-transformed variables (V1–V28) ensure privacy while preserving universal fraud distributions and predictive integrity (Pozzolo et al., 2015). To manage the extreme computational load of TPE optimization and 5-Fold Cross-Validation, Stratified Random Subsampling extracted 23,858 rows, retaining the 270:1 imbalance ratio to prevent

minority class distortion (Haixiang et al., 2017). Removing 89 invalid rows (9 missing, 80 duplicates) to prevent data leakage Lemaˆıtre et al., (2017) yielded a final skewed dataset of 23,769 transactions (23,681 legitimate, 99.63%; 88 fraudulent, 0.37%). Finally, the Time and Amount features were normalized via StandardScaler to complete data preparation.

Data Partitioning and Cross-Validation (5-Fold Stratified CV)

The final dataset was partitioned using Stratified Data Splitting into 80% training data (19,015 rows; 70 frauds) and 20% test data (4,754 rows; 18 frauds). This stratification ensured the class inequality ratio remained identical across both partitions (Abdelghafour et al., 2024). To ensure model robustness and prevent overfitting, the training data was rigorously evaluated using 5-Fold Stratified Cross-Validation to obtain the mean and standard deviation estimates of its statistical performance (Chicco & Jurman, 2020).

Bayesian-Based Hyperparameter Optimization (TPE)

To optimize parameters efficiently, automated hyperparameter search was executed using the Tree structured Parzen Estimator (TPE) via the Hyperopt library (Bergstra et al., 2013). Unlike accuracy driven methods, the TPE objective function minimized the Expected Cost of Misclassification (ECM) (Tao et al., 2024). Configured for 50 iterations based on Bergstra et al., (2013) recommendation of 10 evaluations per dimension, this setup achieves convergence comparable to a 500 combination Grid Search with 10× faster time efficiency (Bergstra et al., 2013)

Asymmetric Cost Evaluation (ECM) Testing and Valuation Scenarios

The best final model of the XGBoost and LightGBM architectures optimized by TPE was then tested head-to-head on the test data (4,754 rows) through three integrated evaluation scenarios:

(1) Statistical Classification Metrics: Accuracy, Precision, Recall, F1-Score, AUC-ROC, and AUC-PR (Velarde et al., 2023). Given extreme class imbalance, F1-Score was prioritized:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad \ldots. \ (2)$$

(2) Infrastructure Computational Complexity: Memory usage (Space Complexity in KB) and execution speed (Time Complexity in seconds), converted to banking cloud server operational cost estimates.

(3) Operational Cost Calculation (ECM): Prediction results from the confusion matrix are converted directly into nominal rupiah values. Total losses are calculated using the following Expected Cost of Misclassification (ECM) equation:

$$ECM = (FP \times C_{FP}) + (FN \times C_{FN}) \ldots. (3)$$

Symmetric metrics are inadequate for financial analytics; thus, this study adopts Cost Sensitive Learning reflecting asymmetric real-world losses (Ariza-Garzón et al., 2024; Elkan & Charles, 2001; Vanderschueren et al., 2022). True Positives (*TP*) and True Negatives (*TN* incur zero penalty (C_{TP} = 0 and C_{TN}= 0). The False Positive Cost (C_{FP}) is estimated at IDR 50,000 to account for customer service operational costs triggered when legitimate transactions are incorrectly blocked as false alarms. Meanwhile, the False Negative Cost (C_{FN}) is set at a severe IDR 5,000,000, representing the principal limit loss the bank must replace when fraudulent transactions escape the system. This 1:100 asymmetric ratio, validated by Indonesian macro data (IDR 1.4 trillion from Indonesia Anti Scam Center and IDR 2 trillion from OJK) (Husin et al., 2025; Zanah et al., 2025), confirms that minimizing accumulated ECM through heavy False Negative penalties is the optimal business solution.

RESULTS AND DISCUSSION

Cross-Validation Results

a. Empirical Result

Before conducting final testing on the test data set, the XGBoost and LightGBM architectures, optimized using the Tree-structured Parzen Estimator (TPE), were first evaluated on the training data using a 5-Fold Stratified Cross-Validation scenario. The empirical results from the five test folds, including the mean and standard deviation (std), are summarized in Table 1.

Evaluation Metrics	**XGBoost (Mean ± Std)**	**LightGBM (Mean ± Std)**
Akurasi	0.9993 ± 0.0001	0.9992 ± 0.0002
Presisi	0.9129 ± 0.0518	0.9176 ± 0.0244
Recall	0.8974 ± 0.0424	0.8745 ± 0.0660
F1-Score	0.9028 ± 0.0128	0.8938 ± 0.0339
AUC-ROC	0.9973 ± 0.0027	0.9985 ± 0.0015

Table 1 5-Fold Stratified Cross-Validation results are here. Fill in the table with the mean ± standard deviation data from Colab)

b. Discussion

Based on the empirical data in Table 1, both ensemble algorithms demonstrated robust performance with extremely small standard deviations across all metrics. The standard deviations, which are significantly below 0.05 for both models, theoretically indicate that the TPE optimization successfully established stable generalization capabilities. This strictly prevents indications of overfitting or underfitting even when the models are trained on different data subsamples (Abdelghafour et al., 2024). Furthermore, the ROC-AUC values in the validation phase consistently approached perfection (>0.99), proving that the sequential tree-building architecture successfully mapped the distribution patterns of the minority (fraud) class despite the severe 270:1 imbalance.

Classification Performance Comparison

a. Empirical Result

Final testing was conducted on a test data dimension that had never been encountered by the models before (consisting of 4,736 legitimate transactions and 18 fraudulent transactions). The classification performance metrics are presented in Table 2, supported by the Confusion Matrix in Figure 2 and the ROC/PR Curves in Figure 3.

Metrik Evaluasi	**XGBoost**	**LightGBM**
Akurasi	0.9994	0.9998
Presisi	0.8947	1

Recall	0.9444	0.9444
F1-Score	0.9189	0.9714
AUC-ROC	0.9783	0.9982
AUC-PR	0.9419	0.9503

Table 2 Classification Performance Comparison on Test Data is here. Populate the table with the metrics Accuracy, Precision, Recall, F1-Score, AUC-ROC, and AUC-PR.

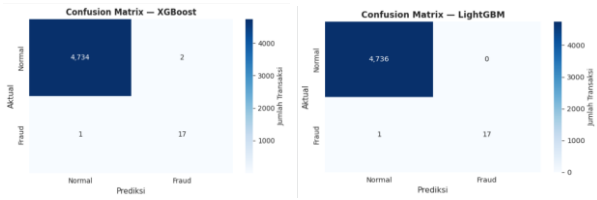

Figure 2 Confusion Matrix XGBoost and LightGBM

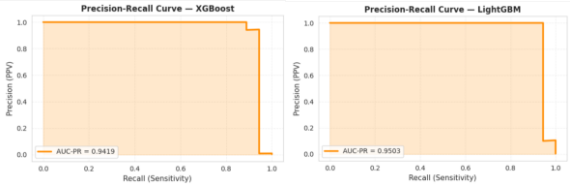

Figure 3 Kurva ROC-AUC and Kurva AUC-PR

The empirical results show that LightGBM outperformed XGBoost across nearly all key metrics. Both achieved an identical Recall of 0.9444 (capturing 17 out of 18 frauds). However, LightGBM achieved perfect Precision (1.0000) with zero false positives (TN=4,736, FP=0), while XGBoost recorded two false positives (FP=2). To ensure this was not random chance, a McNemar statistical test was performed, yielding a chi-square value of 2.00 with a p-value of 0.157 (alpha = 0.05).

b. Discussion

Statistically, the McNemar test implies that on a sample of 4,754 transactions, the two models have comparable fundamental detection capabilities ($p > 0.05$), although LightGBM shows a marginal but critical advantage in exact precision (Raschka, 2018). Delving deeper into the theoretical mechanics, this superiority can be traced directly to its growth architecture. XGBoost grows trees level-wise (widening), which sometimes retains less relevant branches. Conversely, LightGBM utilizes a leaf-wise growth strategy (depth-wise). TPE optimization successfully found the optimal maximum depth for LightGBM to precisely isolate anomalies. From a practical banking perspective, the absolute absence of false positives in LightGBM makes it an ideal architecture, ensuring there are no service frictions that could trigger customer churn.

Infrastructure Computational Efficiency Evaluation

a. Empirical Result

The operational feasibility of the anomaly detection system is evaluated through time complexity and space complexity (converted into estimated cloud server operational costs), as presented in Table 3.

Parameter	**XGBoost**	**LightGBM**
Waktu Latih (Training)	0.5441 detik	1.0507 detik

Waktu Prediksi (Inferensi)	0.0098 detik	0.0608 detik
Memori Model	4.98 KB	8.31 KB
Estimasi Biaya Cloud	Rp 120,91	Rp 233,50

Table 3 Computational Efficiency Metrics and Infrastructure Cost Estimates are here. Fill in the Training Time, Prediction Time, Model Memory, and Infra Cost from Colab.

Technically, XGBoost demonstrated superior time and memory efficiency. XGBoost completed the training phase in just 0.5441 seconds, nearly twice as fast as LightGBM (1.0507 seconds). Its inference speed for 4,754 rows was 0.0098 seconds. Consequently, the estimated cloud infrastructure cost for XGBoost (IDR 120.91) is lower than LightGBM (IDR 233.50).

b. Discussion

When evaluated within a broader business context, this efficiency vs. precision trade-off significantly favors LightGBM. At the scale of modern banking IT architectures, LightGBM's inference latency of 60 milliseconds strictly meets the Service Level Agreement (SLA) standards. As established by Tao et al., (2024), the maximum latency threshold for digital payment systems is 100 milliseconds per transaction batch. Thus, LightGBM operates 40% below the critical threshold. From a Return on Investment (ROI) standpoint, the minor additional infrastructure cost of IDR 112.59 for LightGBM is vastly offset by the operational savings generated from zero false positives.

Cost Sensitivity Analysis (Expected Cost of Misclassification / ECM)

a. Empirical Result

Converting prediction weaknesses into direct financial losses using the ECM approach reveals the true operational disparity. The total loss calculations are summarized in Table 4.

ECM Matrix	**Cost per Case**	**XGBoost**	**LightGBM**
False Positive (False Alarm)	Rp50.000	2 (Rp 100.000)	0 (Rp 0)
False Negative (The Fraudster Gets Through)	Rp 5.000.000	1 (Rp 5.000.000)	1 (Rp 5.000.000)
Total Loss (ECM)	**-**	**Rp 5.100.000**	**Rp 5.000.000**

Table 4 Calculate the Cost Matrix and Total ECM here. Compare the FP, FN, and total ECM costs between XGBoost and LightGBM.

Both models recorded one False Negative (failure to detect a fraud), resulting in a principal loss penalty of IDR 5,000,000 each. However, due to generating two false alarms (False Positives), XGBoost triggered an additional administrative fee of IDR 100,000, increasing its total ECM loss to IDR 5,100,000. LightGBM successfully reduced the False Positive rate to zero, keeping the total ECM loss at the absolute minimum (IDR 5,000,000).

b. Discussion

Translating these findings into macroeconomic implications reveals the vital importance of Cost-Sensitive Learning in modern banking. The IDR 100,000 difference may appear minor on a test sample of 4,754 data points. However, assuming a bank processes millions of transactions daily, extrapolated losses from false alarms grow exponentially. LightGBM's absolute reliability as the most cost-effective model will essentially save billions of rupiah in operational inefficiencies for the national banking sector while completely preventing the loss of customer loyalty caused by inappropriate card blocking.

Model Interpretation Using Explainable AI (SHAP)

a. Empirical Result

To transparently unpack the predictive logic (black box) of the models, SHAP values were extracted. Based on the SHAP Summary Plot (Figure 4), anonymous features such as V14, V4, and V17, along with the non-PCA Amount feature, emerged as the strongest marginal contributors. The Waterfall Plot (Figure 5) further provides a local forensic breakdown for a specific transaction case, explicitly showing how the combination of the Amount and V14 features pushed the prediction probability beyond the detection threshold.

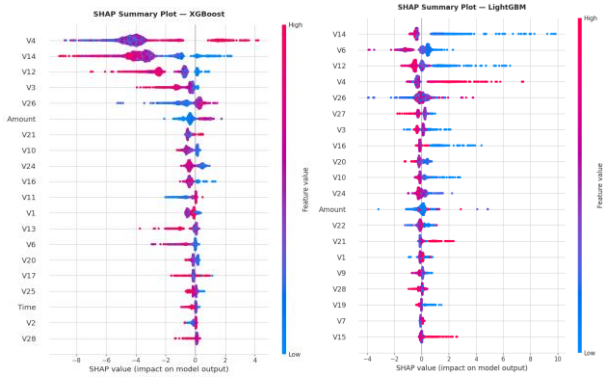

Figure 4 SHAP Summary Plot of XGBoost and LightGBM

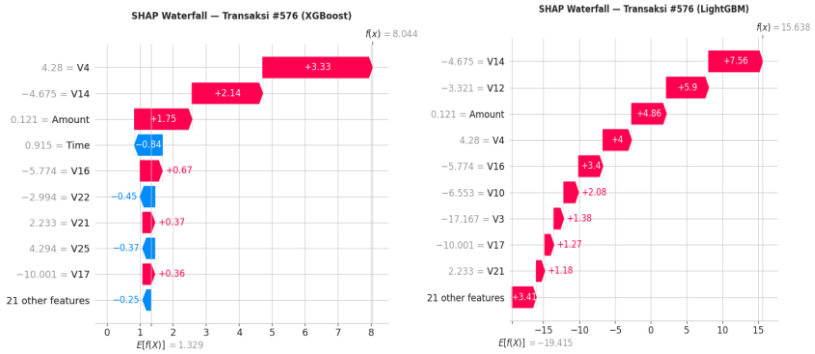

Figure 5 SHAP Waterfall Plot for specific Fraud transaction prediction

b. Discussion

Interpreting these visualizations theoretically, the strong negative correlation of V14 (where lower numerical values aggressively encourage fraud prediction) and extreme Amount values indicate a significant deviation in transaction behavior from historical customer baseline patterns. This aligns with standard anomaly principles (Lundberg & Lee, 2017). Furthermore, from a regulatory compliance standpoint, the granular transparency provided by the SHAP Waterfall Plot ensures the system is fully accountable during banking risk management audits, perfectly satisfying the stringent requirements of Explainable AI regulations in the financial sector (Liu et al., 2026).

CONCLUSION

This study successfully optimized tree-ensemble algorithms to develop a cost-sensitive fraud detection model. TPE evaluation demonstrated LightGBM's superiority over XGBoost in handling extreme financial data, achieving perfect precision (1.0000), an F1-Score of 0.9714, and an AUC-PR of 0.9503. While XGBoost trains faster, LightGBM’s 60-millisecond inference meets real-time banking SLA standards. Crucially, LightGBM proved to be the most cost-effective model, minimizing ECM operational losses to IDR 5,000,000 (versus XGBoost's IDR

5,100,000). SHAP integration further ensures audit-compliant transparency by identifying V14 and Amount as primary prediction drivers, significantly enhancing banking risk governance.

A primary limitation is the reliance on a European dataset, which restricts geographic generalizability and risks concept drift within the Indonesian financial ecosystem (Kurniawan et al., 2025). Future research must validate the winning algorithm using local banking datasets with distinct demographics and implement a Dynamic ECM tailored to actual transaction amounts for precise loss estimation. Finally, developers should transition this empirical model into production through an API or UI/UX dashboard, enabling risk teams to monitor fraud alerts and SHAP interpretations in real time.

REFERENCES

Abdelghafour, E. B., Mohamed, C., Noura, A., & Abdelhamid, B. (2024). Enhancing Credit Card Fraud Detection Using a Stacking Model Approach and Hyperparameter Optimization. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 15, Number 10). www.ijacsa.thesai.org

Ahsan, M. M., Ali, M. S., & Siddique, Z. (2024). Enhancing and improving the performance of imbalanced class data using novel GBO and SSG: A comparative analysis. *Neural Networks*, *173*. https://doi.org/10.1016/j.neunet.2024.106157

Ariza-Garzón, M. J., Arroyo, J., Segovia-Vargas, M. J., & Caparrini, A. (2024). Profit-sensitive machine learning classification with explanations in credit risk: The case of small businesses in peer-to-peer lending. *Electronic Commerce Research and Applications*, *67*. https://doi.org/10.1016/j.elerap.2024.101428

Bergstra, J., Yamins, D., & Cox, D. D. (2013). *Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures* (Vol. 28).

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*, 785–794. https://doi.org/10.1145/2939672.2939785

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1). https://doi.org/10.1186/s12864-019-6413-7

CNBC Indonesia. (2023, May 12). *Diserang Ransomware, Mobile Banking BSI Error Berhari-hari*. https://www.cnbcindonesia.com/news/20230512134929-8-436887/diserang-ransomware-mobile-banking-bsi-error-berhari-hari

CNN Indonesia. (2025, September 26). *Fakta-fakta Pembobolan Rekening Dormant Rp204 Miliar*. https://www.cnnindonesia.com/nasional/20250926070515-12-1277851/fakta-fakta-pembobolan-rekening-dormant-rp204-miliar

Darmawan, R. A., Musyafa, A., & Handayani, M. (2026). Optimization of RNN and Tree-Based Models with Imbalance Handling for Fraud Detection in Digital Banking Transactions. *Jurnal Ilmiah Multidisiplin Indonesia (JIM-ID)*, *5*(02), 347–366.

Elkan, & Charles. (2001). *The Foundations of Cost-Sensitive Learning*.

Fernández, A., Garcia, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). *Learning from Imbalanced Data Sets*.

Forough, J., & Momtazi, S. (2021). Ensemble of deep sequential models for credit card fraud detection. *Applied Soft Computing*, *99*. https://doi.org/10.1016/j.asoc.2020.106883

Friedman, J. H. (2001). GREEDY FUNCTION APPROXIMATION: A GRADIENT BOOSTING MACHINE 1. In *The Annals of Statistics* (Vol. 29, Number 5).

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). *Why do tree based models still outperform deep learning on tabular data*.

Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. In *Expert Systems with Applications* (Vol. 73, pp. 220–239). Elsevier Ltd. https://doi.org/10.1016/j.eswa.2016.12.035

Hajek, P., Novotny, J., & Munk, M. (2026). Financial statement fraud detection using topic-driven financial sentiment analysis. *Decision Support Systems*, *203*, 114615. https://doi.org/10.1016/j.dss.2026.114615

Höppner, S., Baesens, B., Verbeke, W., & Verdonck, T. (2022). Instance-dependent cost-sensitive learning for detecting transfer fraud. *European Journal of Operational Research*, *297*(1), 291–300. https://doi.org/10.1016/j.ejor.2021.05.028

Husin, L. S. S., Darmayanti, E. F., & Nusantoro, J. (2025). *Implementasi Model Pendekatan Machine Learning untuk Deteksi Fraud pada Transaksi Pembayaran Digital Paper.Id*.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*. https://github.com/Microsoft/LightGBM.

Kurniawan, M., Putra, H., Mintaraga, C. A., & Hidayaturrahman. (2025). Sequential Oversampling for Fraud Detection: Leveraging Generative Adversarial Networks and Continual Learning Approach in Imbalanced Data Streams. *Procedia Computer Science*, *269*, 485–501. https://doi.org/10.1016/j.procs.2025.08.301

Lawson, R., & Nancy, J. (2024). *Detecting First-Party Fraud in Online Lending Using Machine Learning Models*.

Lemaˆıtre, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. In *Journal of Machine Learning Research* (Vol. 18). http://jmlr.org/papers/v18/16-365.html.

Liu, W., Han, Y., Lan, X., Yu, Z., Xia, M., Lin, S., Pang, C., & Chen, N. (2026). Progressive gradient boosted trees for imbalanced financial distress prediction. *Expert Systems with Applications*, *321*. https://doi.org/10.1016/j.eswa.2026.132187

Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*. http://arxiv.org/abs/1705.07874

Nugroho, L. P., Saputro, R. E., & Utomo, F. S. (2026). Performance Comparison Of Xgboost Lightgbm And Lstm For E-Commerce Repeat Buyer Prediction. *Jurnal Teknik Informatika (JUTIF)*, *7*(1). https://doi.org/10.52436/1.jutif.2026.7.1.5746

Pozzolo, A. D., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). Calibrating probability with undersampling for unbalanced classification. *Proceedings - 2015 IEEE Symposium Series on Computational Intelligence, SSCI 2015*, 159–166. https://doi.org/10.1109/SSCI.2015.33

PPATK. (2025, September 25). *Bareskrim Polri Ungkap Kasus Pembobolan Rekening Dorman Bank BUMN Rp204 Miliar Terkait Kejahatan Siber dan Pencucian Uang*. https://www.ppatk.go.id/news/read/1529/bareskrim-polri-ungkap-kasus-pembobolan-rekening-dorman-bank-bumn-rp204-miliar-terkait-kejahatan-siber-dan-pencucian-uang.html

Raschka, S. (2018). *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*.

Shapley, L. S. (1952). *A Value for N-Person Games*.

Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, *81*, 84–90. https://doi.org/10.1016/j.inffus.2021.11.011

Sugiyono. (2023). *METODE PENELITIAN KUANTITATIF, KUALITATIF, DAN R&D*. Alvabeta. www.cvalfabeta.com

Tao, S., Peng, P., Li, Y., Sun, H., Li, Q., & Wang, H. (2024). Supervised contrastive representation learning with tree-structured parzen estimator Bayesian optimization for

imbalanced tabular data. *Expert Systems with Applications*, *237*. https://doi.org/10.1016/j.eswa.2023.121294

Tayebi, M., & El Kafhali, S. (2025). A novel approach based on XGBoost classifier and Bayesian optimization for credit card fraud detection. *Cyber Security and Applications*, *3*. https://doi.org/10.1016/j.csa.2025.100093

Vanderschueren, T., Verdonck, T., Baesens, B., & Verbeke, W. (2022). Predict-then-optimize or predict-and-optimize? An empirical evaluation of cost-sensitive learning strategies. *Information Sciences*, *594*, 400–415. https://doi.org/10.1016/j.ins.2022.02.021

Velarde, G., Sudhir, A., Deshmane, S., Deshmunkh, A., Sharma, K., & Joshi, V. (2023). *Evaluating XGBoost for Balanced and Imbalanced Data: Application to Fraud Detection*. http://arxiv.org/abs/2303.15218

Wald, A. (1950). *Statistical Decision Functions*.

Widjaja, G. (2026). Digitalisasi Sistem Pembayaran Dan Risiko Hukum Ai Dalam Deteksi Fraud: Kepastian Hukum Bagi Pelaku Usaha Dan Konsumen Dalam Ekonomi Finansial Teknologi. *Jurnal Ekonomi Dan Bisnis (Jebi)*, *3*(11).

Xiao, Y., Tan, L., & Liu, J. (2025). *Application of Machine Learning Model in Fraud Identification: A Comparative Study of CatBoost, XGBoost and LightGBM*. https://doi.org/10.20944/preprints202503.1199.v1

Xu, B., Wang, Y., Liao, X., & Wang, K. (2023). Efficient fraud detection using deep boosting decision trees. *Decision Support Systems*, *175*. https://doi.org/10.1016/j.dss.2023.114037

Zanah, A. S., Calista, B., Dewi, W. N., & Sokibi, P. (2025). Deteksi Dini Fraud pada Layanan Keuangan Digital Menggunakan Metode Random Forest. In *Indonesian Journal on Data Science* (Vol. 3, Number 2). https://ejournal.unjaya.ac.id/index.php/ijds