Techno-Science Research Journal V14 (Special Issue) (2026) P 40-47

Content list available at ITC

Techno-Science Research Journal

Journal Homepage: http://techno-srj.itc.edu.kh/

Analysis on Machine Learning Models for Imbalanced Data Problem in Payment Fraud Detection

Siuphing SEUN*, Sokkhey PHAUK, Kimlong NGIN, Pakrigna LONG

Department of Applied Mathematics and Statistic, Institute of Technology of Cambodia, Russian Federation Blvd., P.O. Box 86, Phnom Penh, Cambodia

Received: 20 June 2025; Revised: 16 September 2025; Accepted: 18 September 2025; Available online: 31 July 2026

Abstract: *Payment card fraud losses worldwide reached $33.83 billion in 2023 (Nilson Report, 2025). Alarmingly, Deputy Prime Minister and Minister of Interior Sar Sokha (2024) stated that in the first semester of 2024, Cambodians lost nearly $40 million to online and digital fraud. To prevent these significant financial losses, it's crucial to identify the predictive models that can more accurately detect the anomalies in transactions. This study investigates which predictive models work best in predicting the anomalies in the payment transaction. The dataset contains types of online transactions, the amount of the transactions, names of the sender and receiver, the sender's balance of account balance before and after the transaction, and the receiver's account balance before and after receiving money. Autoencoder, LightGBM, Neural Network, Logistic Regression, Random Forest, and CatBoost were built as prediction models within this research. Each of these algorithms is employed to create prediction models, which are meticulously fine-tuned to yield the most accurate prediction of the fraud payment in the system involving optimizing hyperparameters and selecting the best features to enhance the models' prediction power. Common performance metrics such as Precision, Recall, F1-Score, and AUC-ROC were used to test each model's performance. Extensive experimentation data shows the best performance of CatBoost (AUC-ROC of 0.895 for Oversampled and 0.999 for other kinds of datasets) and Random Forest model (AUC-ROC of 0.99), which consistently outperforms other machine learning methods in accurately predicting payment fraud, whether training with an imbalanced or balanced dataset. These results highlight the model's ability to handle complex, non-linear relationships within the data and its effectiveness in generalizing across different scenarios and conditions. In conclusion, the findings of this study will be beneficial mostly to the banking system as they can apply the models we found in their system to prevent any fraudulent activities. Spotting the fraudulent activities in the early stage plays an immense role in preventing the loss.*

Keywords: Payment Card Fraud; Machine Learning Models; Fraud Prediction; Predictive Modeling; Class Imbalance

1. INTRODUCTION

According to the World Payments Report 2025 by Capgemini Research Institute (2023), global non-cash transaction volumes reached 1.411 trillion in 2023 and are projected to rise to 1.65 trillion in 2024. With this frequent transaction will attract malicious actors seeking to exploit vulnerabilities in the payment system.

The term "Fraud" was given the meaning of the crime of getting money by deceiving people (Cambridge, 2023). Payment card fraud losses worldwide reached $33.83 billion in 2023 (Nilson Report, 2025). Alarmingly, Deputy Prime Minister of Interior Sar Sokha (2024) stated that in the first semester of 2024, Cambodians lost nearly $40 million to online fraud and digital fraud. With this enormous amount of money lost due to the fraudulent payment, it's needed to find the predictive models that can better predict the anomalies in transactions.

Our research applied various machine learning (ML) such as Autoencoder, LightGBM, Logistic Regression, Neural Network, Random Forest, and CatBoost to compare the prediction performance of the abnormalities in the transaction.

Bhattacharyya et al. (2011) compare Logistic Regression and Support Vector machines, and Random Forest to see which one yields the best outcome. Moreover, Oza (2018) used Logistic Regression, Linear Kernel, and Radial Basis Function Kernel to predict fraud payment. Likewise, Oğuz

* Corresponding author: Siuphing SEUN
E-mail: siuphing15@gmail.com;

and Layth (2020) applied RF, NB, SVM, and KNN to study the comparison of the proposed model and found out that RF is the most effective model for classifying credit card fraud transactions as fraudulent or genuine. Similarly, Minjun (2023) proposed 3 ML known as SVM, LR, and Decision Tree which obtained the result in Decision Tree as the best model compared to SVM and LR for its high accuracy. Last but not least, Feng and Kim (2024) selected five machine learning models to perform fraud payment prediction, such as Random Forest with adaptive Boosting, Gradient Boosted Decision Tree, K-Nearest Neighbor, Convolutional Neural Network, and Support Vector Machine.

While previous studies have compared the performance of many proposed models in their studies as well as identified factors that lead to fraudulent payment, there is still a gap in effectively using different approaches that can better predict fraudulent activity with both balanced and imbalanced datasets. For instance, Neural Network can be a candidate in predicting fraud for their ability to capture complex relationships in data, especially when there's a lot of nonlinear and high-dimensional information. Likewise, Catboost is the ML that shouldn't be left out in applying to train the model to detect legitimate activity. Although it hasn't been applied by many, Catboost is designed for faster and more accurate results, especially when working with complex datasets. Our study addresses this by applying some ML that haven't been selected to analyze as well as using resampling techniques to balance the data. This research aims to validate the effectiveness of ML in predicting fraud activities. The findings of this study will be beneficial mostly to the banking system as they can apply the models we found into their system to prevent any fraudulent activities. Spotting the fraudulent activities in the early stage plays an immense role in preventing the loss. Nevertheless, predicting fraudulent payments aligns closely with the goals of SDG9 and SDG16. By leveraging innovative technologies and robust infrastructure (SDG9), we can enhance payment systems to detect and prevent fraud, fostering economic growth and safeguarding resources. Furthermore, ensuring integrity in financial transactions supports peaceful and just societies by combating corruption and fostering trust in institutions (SDG16). Fraud detection not only protects individuals and businesses but also promotes global stability and fairness, making it a valuable tool for advancing the human race.

Table 1. Summary of Previous works related to Fraud detection

Reference	Title	Dataset	Algorithms	Result
Bhattacharyya et al. (2010)	Data mining for credit card fraud: A comparative study	Internationally operating financial institution	LR, SVM, RF	Random Forest and SVM perform better than Logistic Regression
Oza (2018)	Fraud Detection using Machine Learning	Paysim Dataset	LR, Linear SVM, and SVM with RBF Kernel	The research demonstrates that SVM based methods are more effective
Oğuz and Layth (2020)	Comparative Analysis of Different Distributions Dataset by Using Data Mining Techniques on Credit Card Fraud Detection	European credit card transaction	RF, NB, SVM, and KNN	The study found that RF is the most effective model for classifying credit card fraud transaction as fraudulent or genuine.
Minjun (2023)	Multiple Machine Learning Models on Credit Card Fraud Detection	Paysim Dataset	SVM, LR, and Decision Tree	Decision Tree is chosen to be the best model compare to SVM and LR for its high accuracy score.
Feng and Kim (2024)	Novel Machine Learning Based Credit Card Fraud Detection Systems	European credit card holders dataset	RF with Adaptive Boosting, GB with Decision Tree, KNN, CNN, and SVM	Random Forest with Adaptive Boosting perform the best compare to the other four models.

2. METHODOLOGY

2.1 Dataset

Due to the lack of access to the real payment transaction from the local bank as those are a very sensitive data which couldn't be leak or share otherwise one has to be responsible with the consequence, the analysis is based on a large dataset gathered from Kaggle called "Online Fraud Payment Detection" which is a synthesis of datasets generated by the PaySim mobile money simulator. Although PaySim dataset is just a synthesis data, it has the pattern very likely to the real-world transaction behavior making it highly suitable to be used within this research. The dataset comprises information from 6,362,620 transactions with five different types of payment known as Cash In, Cash Out, Debit, Transfer, and Payment. Each transaction contains 10 features, including the unit of time, the transaction's type, the transaction's amount, the sender's and receiver's accounts, the balance before and after the transaction of the sender and receiver with the target variables (isFraud). A first examination found that the dataset is imbalanced, with many fewer non-fraud cases than fraud cases. This is a typical issue in fraud payment datasets. Still, it's not really a problem with our study as we want to compare the performance of the selected ML to see which one performs better with both Imbalanced and Balanced datasets used for the training.

To ensure the dataset's reliability and efficiency in training predictive models, thorough cleaning methods were applied. The procedures involved resolving missing values, normalizing continuous attributes, and encoding categorical variables. Each record was thoroughly analyzed for completeness and consistency to ensure that the dataset was reliable and appropriate for testing various predictive algorithms. Additionally, we conducted exploratory data analysis to identify trends and correlations in the dataset, which is essential for developing precise predictive models to improve the prediction of fraud activity in payment transactions. The study also addressed class imbalance by utilizing resampling techniques such as oversampling, under sampling, and SMOTE to balance the dataset and enhance model accuracy. While for the case training the selected ML with imbalance, we applied various techniques such as sample weight, scale pos weight, etc.

Table 2. Description of Variables in the Online Transaction Dataset

Variable's Name	Description
step	A unit of time where 1 step equals to 1 hour
type	Type of online transaction
amount	The amount of the transaction
nameOrig	Customer starting the transaction
oldblanceOrg	Balance before the transaction
newbalanceOrig	Balance after the transaction
nameDest	Recipient of the transaction
oldbalanceDest	Initial balance of recipient before the transaction
newbalanceDest	The new balance of recipient after the transaction
isFraud	Indicates whether the transaction is fraud. The 0 value indicates non-fraud while 1 value indicates fraud

2.2 Data Preprocessing

Data Preprocessing is a crucial step in training the model. We follow the four steps in order to preprocess the data such as missing value handling, data cleaning, data standardization, and encoding technique.

2.3 Feature Selection

Cramér's V and T-Test is applied to see the association between another variable and target variable, isFraud.

Table 3. Cramér's V Result

Feature Name	Cramér's V
type	0.058912
nameDest	0.670693
nameOrig	0.999146

Table 4. T-Test Result

Feature Name	P-Value (T-Test)
step	0.00
amount	0.00
newbalanceOrig	3.50e-94
oldbalanceOrg	2.69e-194
newbalanceDest	0.204
oldbalanceDest	5.42e-51

Table 3. and Table 4. shows the result of Cramer's V and T-Test between every variable compared to the isFraud variable, so-called target variable, to have a better understanding of which variables are strongly related to the target variable and which variables are less related to the target variable. According to the result, we can see that the variable "type" is very less related to the target variable which can be dropped from the dataset when training the model.

2.4 Resampling Methods

To address the dataset's predicted class imbalance, we used three resampling strategies: oversampling, under sampling, and the Synthetic Minority Over-sampling Technique. Each strategy was chosen based on its capacity to improve class balance, hence improving predictive model performance and generalizability.

Oversampling: Using this method, we expanded the size of the minority class ('isFraud = 1') by randomly reproducing instances until the number of fraud cases equaled the number of non-fraud cases. Oversampling ensures that the model is not biased towards the majority class and can learn from the intricacies found in the minority class.

Under sampling: In contrast to oversampling, under sampling includes removing instances at random from the majority class ('isFraud = 0') in order to equal the number of fraud instances. This technique is one of the several techniques used by data scientist to extract more accurate information from originally imbalanced dataset. Nevertheless, the drawback is that important information could be lost.

Synthetic Minority Over-sampling Technique (SMOTE): SMOTE generates new synthetic data points for underrepresented classes through interpolation rather than duplication. This more nuanced representation of the minority class aids models in learning a broader range of features associated with fraud, which can lead to improved prediction of fraudulent transactions. SMOTE has the advantage of creating more diverse synthetic samples, which can enhance the model's ability to generalize to new, unseen data. Even if SMOTE raise several potential risks such as overfitting on synthetic data as well as blurring of class boundaries, it was applied within cross-validation pipeline on the training folds to minimize the risk.

By resolving the class imbalance using these resampling strategies, we saw considerable increases in the performance of our prediction models. Models trained on resampled datasets have higher precision, recall, F1-Score, AUC-ROC, and AUC-PR than those trained on the original imbalanced dataset. This suggests that the models performed better in recognizing both fraud and non-fraud cases, resulting in more trustworthy and generalizable predictions.

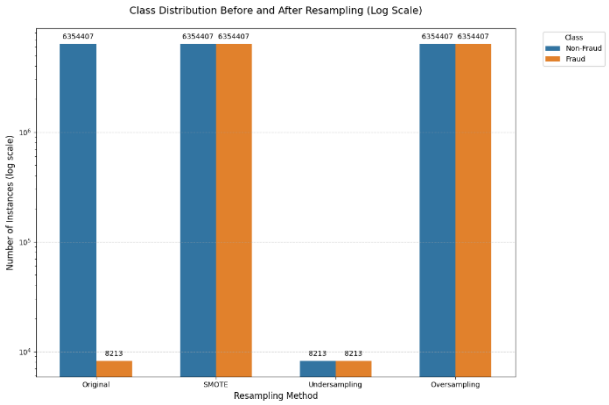

Fig. 1. Class distribution before and after resampling technique

2.5 Proposed Models

Selecting predictive modelling is a crucial phase in the research process as it affects the performance and interpretability of the results. In this study, a variety of algorithms known as Autoencoder, LightGBM, Neural Network, Logistic Regression, Random Forest, and CatBoost are purposefully selected.

Autoencoder are suitable for fraud detection because of their ability to learn compressed representations of normal transaction patterns and identify anomalies that differ significantly from these patterns. As fraud cases are rare compared to legitimate transactions, autoencoders excel in this imbalanced learning scenario by reconstructing normal transactions with low error while producing high reconstruction errors for fraudulent ones. Zhou and Paffenroth (2017) stated that this approach is unsupervised, requiring only normal transaction data for training, which is advantageous given the scarcity of labeled fraud cases. Four-layer architecture consisting of two encoder and two decoder layers were utilized during the implementation stage. It starts with an input layer of the encoder, followed by a dense hidden layer of 64 neurons using ReLU activation and L2 regularization (λ=2.08e-4). Then dropout rate of 0.3793 was applied before compressing into a bottleneck layer of between 16 and 32 neurons, optimized using Optuna, with ReLU activation. On the other hand, the decoder uses the same structure as the encoder with a 64-neuron ReLU layer and concludes with a linear-activated output layer for feature reconstruction. It is trained by applying the Adam optimizer of 1.069e-4 learning rate with MSE loss function. A fixed batch size of 512 was selected from optimization trials of 256/512 for 20 epochs during the final training, applying 5 epochs during hyperparameter tuning, with a throughout applied of dropout and L2 regularization. The data is standardized with StandardScaler before the training, and a split of 20% validation is used to monitor reconstruction performance.

Additionally, LightGBM known as Light Gradient Boosting Machine is particularly effective for fraud detection

for its capability to handle imbalanced datasets, high-dimensional features, and complex decision boundaries inherent in financial transactions. It optimized decision trees for speed and accuracy, making it well-suited for real-time fraud detection systems as a gradient-boosting framework. Ke et al. (2017) initiated that LightGBM's histogram-based algorithm and leaf-wise tree growth enable faster training and lower memory usage compared to traditional boosting methods, which is critical for processing large-scale transaction data in real-time. Gradient boosting with several key hyperparameters optimized through Bayesian Optimization were employed during the training process. The number of leaves (num_leaves), ranging from 20 to 100, is included to control tree complexity along with a learning rate between 0.01 to 0.3 to adjust optimization step sizes. The num_leaves parameter indirectly influences model depth as LightGBM grows trees' leaf-wise rather than depth-wise. Class weighting through scale_pos_weight, adjusting the importance of the positive class, is incorporated to handle class imbalance. The training also applied additional regularization via L1/L2 penalties, so-called reg-alpha and reg_lamba, both ranging from 0 to 1 in order to prevent overfitting. At the same time, the subsampling techniques known as feature_fraction and bagging_fraction, ranging from 0.5 to 1, introduce randomness by selecting subsets of features and data per iteration. A minimum number of samples per leaf, min_child_samples, ranging from 20 to 100, is also enforced to ensure robust splits. During the training, hyperparameter tuning is performed by applying stratified 3-fold cross-validation, optimizing for AUC-PR, average_precision.

Moreover, Neural Networks are highly effective for fraud detection due to their ability to model complex, non-linear relationships in transactional data while adapting to evolving fraud patterns. According to Zhou and Paffenroth (2017), Neural Networks' deep learning architectures such as autoencoders for unsupervised anomaly detection and recurrent networks for sequential pattern analysis, excel at identifying subtle fraudulent activities that traditional rule-based systems miss. The Neural Network architecture employed in this research consists of 3 fully connected layers, such as an input layer with a tunable number of neurons ranging from 64 to 128 using ReLU activation, followed by a dropout layer ranging from 10 to 30 percent for regularization, a hidden layer ranging from 32 to 64 neurons using ReLU activation with another dropout layer ranging from 10 to 20 percent, and an output layer with a single neuron and sigmoid activation for binary classification. By applying Optuna Hyperparameter, Adam optimizer with a dynamically tuned learning rate ranging from 1e-4 to 1e-2 and optimized for binary cross-entropy loss is compiled. A batch size of either 512 or 1024 is employed through the training and runs for 100 to 150 epochs, with early stopping triggered if AUC-PR fails to improve for 5 consecutive epochs.

Also, Logistic Regression remains a widely used method for fraud detection due to its interpretability, computational efficiency, and effectiveness in binary classification tasks. Unlike complex models, logistic regression provides clear probabilistic outputs (between 0 and 1) making it easier to set risk thresholds for fraud alerts (Hosmer et al., 2013). With the inverse regularization strength parameter C being tuned across a logarithmic range from 10^{-4} to 10^{4} during hyperparameter optimization, L2 regularization was applied by default via scikit-learn's built-in penalty term during the implementation. Two different solver's such as 'lbfgs' and 'liblinear', both supporting L2 penalties, were explored. In order to handle class imbalance, both class weighting, balanced option, and various resampling techniques, including SMOTE, random undersampling, and oversampling, were tested. Selecting the best-performing configuration of solver and regularization strength, the randomized search method optimized for average precision (AUC-PR) while evaluating different combinations of these parameters via cross-validation.

Furthermore, Random Forest is exceptionally effective in detecting fraud due to its ensemble learning approach, which combines multiple decision trees to improve accuracy and robustness. Breiman (2001) stated that Random Forest method excels at handling high-dimensional data and identifying complex, non-linear patterns typical in fraudulent transactions. Using several key hyperparameters to optimize performance on imbalanced data, the models are tuned during the implementation stage. The number of trees, n_estimators, is tested, ranging from 50 to 100. At the same time, to balance complexity and generalization, the maximum depth of each tree, max_depth, is explored, ranging from 5 to 10 levels. Also, to control overfitting, the minimum samples required to split a node, min_samples_split, is evaluated between 2 and 10. Multiple strategies are employed in order to address class imbalance. class_weight = 'balanced' to automatically adjust weights based on class frequencies as the baseline model in company with manual weighting schemes such as {0:1, 1:5} and {0:1, 1:10} to emphasize the minority class. Evaluating 5 random combinations per model using 3-folds cross-validation and prioritizes the average_precision metric, RandomizedSearchCV is efficiently conducted during the hyperparameter optimization method.

Last but not least, CatBoost is exceptionally effective in detecting fraud for its advanced handling of categorical features, robustness to overfitting, and superior performance on imbalanced datasets. Unlike traditional gradient boosting methods, CatBoost natively reduces processes categorical variables without manual encoding, reducing feature engineering effort while preserving critical fraud signals (Prokhorenkova et al., 2018). Key hyperparameters such as the number of trees, iterations ranging from 700 to 850, maximum depth, depth ranging from 4 to 10, learning rate ranging from 0.038 to 0.052, L2 regularization, l2_leaf_reg ranging from 1.8 to 3.2, were used during the model's

optimization. In addition to that, border-count, ranging from 32 to 255, for feature discretization and random_strength, ranging from 0.7 to 1.2, for noise injection, were included to prevent overfitting. scale_pos_weight at 28.5 is used as a class weighting strategy. The model was trained with a stratified 3-fold cross-validation, and early stopping at 50 rounds, prioritizing AUC-PR performance, and tuned hyperparameters through Optuna.

2.6 Evaluation Metrics

The efficiency of the predictive modelling is measure by Precision, Recall, F1-Score, AUC-ROC, and AUC-PR in this study as the data that we used to train ML is both balanced and imbalanced data.

In imbalanced datasets, accuracy can be misleading because a model can achieve a high accuracy by simply predicting the majority class for most cases, ignoring the minority class entirely. To be precise, in the case of our dataset where fraud cases make up less than 1% of the data, predicting "not fraud" for every instance would result in 99% accuracy, but the model would fail to identify any fraudulent transactions.

By focusing on AUC-ROC, F1-Score, Precision, and Recall, we will get a more accurate reflection of how well the model performs, especially in detecting the minority class, which is often the focus in applications like fraud detection.

Table 5. Confusion Matrix

		Actual	
		Positive	**Negative**
Predicted	**Positive**	True positive (TP)	False positive (FP)
	Negative	False negative (FN)	True negative (TN)

Precision measures the correctness of positive predictions by the model, defined as the number of true positives divided by the total number of positive predictions (true and false). The formula is:

$$Precision = \frac{TP}{TP+FP} \quad \text{(Eq. 1)}$$

Recall, also known as the true positive rate, indicates how well the model identifies positive instances. The formula is: fraudulent transactions.

$$Recall = \frac{TP}{TP+FN} \quad \text{(Eq. 2)}$$

The F1-Score is the harmonic mean of precision and recall, providing a balance between the two. The formula is:

$$F1\ Score = \frac{2 \times TP}{2 \times TP+FP+FN} \quad \text{(Eq. 3)}$$

AUC-ROC is called area under the receiver operating characteristic curve which plots the true positive (TP) rate versus the false positive (FP) rate at different classification thresholds. The formula is:

$$AUC - ROC = \int_0^1 TPR(FPR)\, dFPR \quad \text{(Eq. 4)}$$

3. RESULTS AND DISCUSSION

After the analysis within this research, significant findings were found across the original (imbalanced) dataset, SMOTE, under-sampled, and oversampled modified datasets. The results of different techniques with different datasets with detailed explanations will be provided.

Table 6. Results for fraud prediction using 'Original' dataset

Model Name	**AUC-ROC**	**F1-Score**	**Precision**	**Recall**
Autoencoder	0.888	0.418	0.975	0.266
Neural Network	0.997	0.798	0.963	0.681
Logistic Regression	0.960	0.258	0.160	0.659
LightGBM	0.998	0.158	0.086	0.992
Random Forest	0.998	0.090	0.047	0.987
CatBoost	0.999	0.689	0.537	0.961

Table 6. shows that CatBoost achieved the highest AUC-ROC score of 0.999 indicating near-perfect ability to distinguish between fraud and non-fraud transactions. However, the Neural Network demonstrated the best practical balance with the highest F1-Score of 0.798 combining strong precision of 0.963 and reasonable recall of 0.681. This makes it particularly suitable for real-world application where both minimizing false alarms and detecting fraud are important. LightGBM and Random Forest excel with the AUC-ROC scores of 0.998 but suffered from extremely low precision of 0.086 and 0.047 respectively meaning that they flagged too many legitimate transactions as fraudulent. While their recall was exceptionally high (99.2% and 98.7%), the high number of false positives makes them less practical for deployment. The Autoencoder had the opposite issue—extremely high precision (97.5%) but very low recall (26.6%), meaning it was overly cautious and missed most fraud cases. Logistic Regression performed moderately, with decent recall (65.9%) but poor precision (16.0%), leading to a low F1-Score (0.258).

Table 7. Results for fraud prediction applying 'SMOTE'

Model Name	AUC-ROC	F1-Score	Precision	Recall
Autoencoder	0.906	0.04	0.02	0.657
Neural Network	0.998	0.567	0.404	0.95
Logistic Regression	0.961	0.521	0.486	0.563
LightGBM	0.999	0.669	0.512	0.965
Random Forest	0.996	0.760	0.930	0.642
CatBoost	0.999	0.448	0.292	0.962

Table 7. indicates that after applying SMOTE (Synthetic Minority Over-sampling Technique) reveal distinct performance patterns across the evaluated models. LightGBM and CatBoost achieve optimal discriminative capability with perfect AUC-ROC scores (0.999), though their operational effectiveness differs substantially - LightGBM maintains a balanced F1-score (0.669) through moderate precision (0.512) and exceptional recall (0.965), while CatBoost suffers from precision degradation (0.292) despite high recall (0.962). The Neural Network demonstrates robust performance (AUC-ROC: 0.998) with strong recall (0.95) but compromised precision (0.404), whereas Random Forest exhibits an inverse pattern with outstanding precision (0.93) but limited recall (0.642). Logistic Regression shows intermediate performance across all metrics (AUC-ROC: 0.961, F1: 0.521), while the Autoencoder proves ineffective despite its high recall (0.657) due to critically low precision (0.02). These findings suggest that while SMOTE enhances model sensitivity to minority class detection, it introduces significant precision-recall tradeoffs that vary substantially across algorithmic approaches, with tree-based methods (LightGBM, Random Forest) demonstrating the most favorable balance for practical fraud detection applications.

Table 8. Results for fraud prediction applying 'Under sampling'

Model Name	AUC-ROC	F1-Score	Precision	Recall
Autoencoder	0.775	0.033	0.017	0.335
Neural Network	0.995	0.095	0.05	0.955
Logistic Regression	0.962	0.269	0.169	0.651
LightGBM	0.999	0.24	0.136	0.995
Random Forest	0.994	0.724	0.942	0.588
CatBoost	0.999	0.129	0.069	0.996

We can see in **Table 8.** that the experimental results from the 'Understanding' sampling technique reveal significant variations in model performance for fraud detection. LightGBM and CatBoost demonstrate near-perfect discriminative ability with AUC-ROC scores of 0.999, coupled with exceptional recall values (0.995 and 0.996 respectively), though their practical utility is limited by poor precision (0.136 and 0.069) and consequently low F1-scores (0.24 and 0.129). The Neural Network follows a similar pattern with high AUC-ROC (0.995) and recall (0.955), but suffers from critically low precision (0.05). In contrast, Random Forest achieves the best balance among all models with strong precision (0.942) and moderate recall (0.588), yielding the highest F1-score (0.724). Logistic Regression shows intermediate performance across all metrics, while the Autoencoder proves ineffective with poor scores across all measures. These results suggest that while some models achieve excellent class separation, their practical application in fraud detection remains constrained by fundamental precision-recall trade-offs, with Random Forest emerging as the most operationally viable option when using the 'Understanding' sampling approach.

Table 9. Results for fraud prediction applying 'Oversampling'

Model Name	AUC-ROC	F1-Score	Precision	Recall
Autoencoder	0.909	0.031	0.015	0.698
Neural Network	0.999	0.325	0.194	0.986
Logistic Regression	0.961	0.503	0.447	0.575
LightGBM	0.999	0.856	0.818	0.897
Random Forest	0.998	0.113	0.060	0.989
CatBoost	0.895	0.006	0.005	0.007

According to **Table 9**, The oversampling technique yields particularly interesting performance variations across different fraud detection models. LightGBM emerges as the standout performer, achieving both exceptional discriminative power (AUC-ROC: 0.999) and operational effectiveness with strong precision (0.818), recall (0.897), and the highest F1-score (0.856) among all models. While Neural Network and Random Forest demonstrate similarly impressive AUC-ROC values (0.999 and 0.998 respectively) along with near-perfect recall (0.986 and 0.989), their practical utility is significantly hampered by substantially lower precision (0.194 and 0.060). Logistic Regression maintains consistent, though unremarkable, performance across all metrics (AUC-ROC: 0.961, F1: 0.503). The Autoencoder shows limited effectiveness despite moderate recall (0.698), while CatBoost performs unexpectedly poorly across all measures, including an unusually low AUC-ROC (0.895) for this type of application. These findings suggest that oversampling can produce excellent results for certain algorithms like LightGBM, but may lead to significant precision degradation in others, highlighting the importance of model-specific evaluation when employing this technique for fraud detection tasks.

4. CONCLUSIONS

The study's investigation into machine learning models for predicting fraud activities in payment transactions using historical data of payment transactions to emphasize the exceptional performance of Autoencoder, Light Gradient Boosting Machine (LightGBM), logistic Regression (LR), Neural Network (NN), Random Forest (RF), and CatBoost. Among these, Random Forest and CatBoost are consistently the top-performing models across all techniques. They balance Precision and Recall effectively, achieving near-perfect AUC-ROC scores. On the other hand, Autoencoder, despite its high Precision in the imbalanced dataset, struggles

to deliver balanced performance when sampling techniques are applied. This highlights Random Forest and CatBoost's proficiency in capturing the abnormalities in the payment transactions, providing a solid foundation for optimizing and enhancing payment systems.

Fraudulence in payment transactions is a very sensitive issue that should be tackled and prevented in the early stage, otherwise the loss will be tremendous to the country's economy and growth. By applying the best model among the proposed models in the research to detect the abnormalities in the transaction, the bank can significantly reduce fraudulent activities. Integrating this advanced system will not only enhance security but also improve customer trust and operational efficiency. Additionally, regulatory parties, fintech partners, and customers will benefit from the increased fraud prevention measures, ensuring a safer and more reliable financial ecosystem. Despite challenges, predicting fraudulent payments strongly aligns with the objectives of SDG9 and SDG16. By utilizing advanced technologies and establishing resilient infrastructure (SDG9), payment systems can be improved to identify and prevent fraud, driving economic progress and preserving valuable resources. Moreover, upholding integrity in financial transactions helps build peaceful and just societies by reducing corruption and strengthening trust in institutions (SDG16). Fraud detection safeguards both individuals and businesses while contributing to global stability and fairness, making it an essential instrument for humanity's advancement.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisors, Dr. PHAUK Sokkhey for his expert guidance, unwavering support, and invaluable insights throughout my research. His mentorship has greatly influenced my academic path. Without his support, knowledge, and guidance, this research wouldn't have been possible.

I am also grateful to Mr. NGIN Kimlong, and Mr. LONG Pakrigna for their valuable support and constructive feedback which have greatly contributed to the development of this study. Ther willingness to share their knowledge and provide insightful suggestion has been crucial in refining my work.

Finally, I am deeply grateful to my family and friends for their unwavering support and understanding. Their encouragement and belief in me have been a constant source of motivation.

REFERENCES

Ata, O., & Hazim, L. (2020). Comparative Analysis of Different Distributions Dataset by Using Data Mining Techniques on Credit Card Fraud Detection. *Tehnički Vjesnik*, *27*(2), 618–626. https://doi.org/10.17559/TV-20180427091048

Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, *50*(3), 602–613. https://doi.org/10.1016/j.dss.2010.08.008

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Cambridge Dictionary. (2023). FRAUD | Meaning in the Cambridge English Dictionary. Retrieved from dictionary.cambridge.org website: https://dictionary.cambridge.org/dictionary/english/fraud

Capgemini . (2023, September 14). Global non-cash transaction volumes set to reach 1.3 trillion in 2023. Retrieved from Capgemini website: https://www.capgemini.com/news/press-releases/global-non-cash-transaction-volumes-set-to-reach-1-3-trillion-in-2023/

Dai, M. (2023). Multiple Machine Learning Models on Credit Card Fraud Detection. *BCP Business & Management*, *44*, 334–338. https://doi.org/10.54691/bcpbm.v44i.4839

Feng, X., & Kim, S.-K. (2024). Novel Machine Learning Based Credit Card Fraud Detection Systems. *Mathematics*, *12*(12), 1869–1869. https://doi.org/10.3390/math12121869

Horng, P. (2024, August 20). Interior Minister says Cambodians lost $40m to online fraud in H1 2024 - Khmer Times. Retrieved from Khmer Times - Insight into Cambodia website: https://www.khmertimeskh.com/501544975/interior-minister-says-cambodians-lost-40m-to-online-fraud-in-h1-2024/

Hosmer, D., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. New York, Etc.: John Wiley and Sons, Cop.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., … Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems*, *30*. Retrieved from https://papers.nips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html

Nilson. (2023). Card Fraud Losses Worldwide in 2023. Retrieved from Nilson Report website: https://nilsonreport.com/articles/card-fraud-losses-worldwide-in-2023/

Oza, A. (2018). *Fraud Detection using Machine Learning*.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2019). CatBoost: unbiased boosting with categorical features. *ArXiv:1706.09516 [Cs]*. Retrieved from https://arxiv.org/abs/1706.09516

Zhou, C., & Paffenroth, R. C. (2017). Anomaly Detection with Robust Deep Autoencoders. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/3097983.3098052