UDC 004.77 **DOI: 10.32620/reks.2025.1.06**

Mohamed AKOUHAR[1], Abdallah ABARDA[2], Mohamed EL FATINI[1], Mohamed OUHSSINI[3]

[1] *Lab of Analysis, Geometry, and Applications University ibn Tofail, Kenitra, Morocco*
[2] *Lab of Mathematical Modeling and Economic Calculations Hassan First University, Settat, Morocco*
[3] *Lab SIV, Department of Computer Science University ibn Zohr, Agadir, Morocco*

ENHANCING CREDIT CARD FRAUD DETECTION: THE IMPACT OF OVERSAMPLING RATES AND ENSEMBLE METHODS WITH DIVERSE FEATURE SELECTION

*The **subject matter** of this article is enhancing credit card fraud detection systems by exploring the impact of oversampling rates and ensemble methods with diverse feature selection techniques. Credit card fraud has become a major issue in the financial world, leading to substantial losses for both financial institutions and consumers. As the volume of credit card transactions continues to grow, accurately detecting fraudulent behavior has become increasingly challenging. **The goal** of this study is to enhance credit card fraud detection by analyzing oversampling rates to select the optimal one for the highest-performing models and using ensemble techniques based on diverse feature selection approaches. **The key tasks** undertaken in this study include assessing the models' performance based on accuracy, recall, and AUC scores, analyzing the effect of oversampling using the Synthetic Minority Over-sampling Technique (SMOTE), and proposing an ensemble method that combines the strengths of different feature selection techniques and classifiers. The **methods** used in this research involve applying a range of machine learning techniques, including logistic regression, decision trees, random forests, and gradient boosting, to an imbalanced dataset where legitimate transactions significantly outnumber fraudulent ones. To address the data imbalance, the researchers systematically investigated the impact of varying oversampling rates using SMOTE. Additionally, they developed an ensemble model that integrates seven feature selection methods with the eXtreme Gradient Boosting (XGB) algorithm. The results show that the application of SMOTE significantly improves the performance of the machine learning models, with an optimal oversampling rate of 20% identified. The XGB model stood out for its exceptional performance, with high accuracy, recall, and AUC scores. Furthermore, the proposed ensemble approach, which combines the strengths of the diverse feature selection techniques and the XGB classifier, further enhances the detection accuracy and system performance compared to the traditional methods. The **conclusions** drawn from this research contribute to advancing the field of credit card fraud detection by providing insights into the impact of oversampling and the benefits of ensemble methods with diverse feature selection. These insights can aid in the development of more effective and robust fraud detection systems, helping financial institutions and consumers better protect against the growing threat of credit card fraud.*

Keywords: *credit card fraud; machine learning; fraud detection; ensemble methods; feature selection; SMOTE.*

1. Introduction

The rapid evolution of the global landscape and financial industries has significantly enhanced convenience in individuals' lives, particularly during the COVID-19 pandemic when many transitioned to online platforms. However, this shift has also led to a surge in financial crimes such as credit card fraud. Global losses due to payment fraud have dramatically increased, rising from USD 9.84 billion in 2011 to USD 32.39 billion in 2020, and are projected to reach USD 40.62 billion by 2027 [1]. Credit card fraud continues to be a significant problem in today's financial world, causing substantial losses for both institutions and consumers [2]. As the volume of credit card transactions grows, detecting fraudulent behavior becomes increasingly challenging. Ensuring the security of all transactions, with a focus on fraud detection and prevention, is a critical task [3].

1.1. Motivation

Credit card theft manifests in various forms, from ATM skimming to large-scale data breaches at payment processors. Despite efforts to secure payment systems, enhancing credit card security is an ongoing research topic. Many banks and financial institutions use rule-

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

based systems, where experts design rules based on historical fraud patterns [4]. Transactions that trigger these rules require further investigation, but fraudsters continually develop methods to circumvent these regulations. Credit card fraud can occur through theft, application fraud, use of fake cards, non-receipt of issued cards (NRI), and online fraud, including card-not-present (CNP) fraud, which only requires access to card data, not the cardholder's physical presence [5, 6].

1.2. Objective

The research enhanced credit card fraud detection by developing a novel machine learning approach that shows quantifiable improvements over existing solutions on imbalanced datasets. To address the challenge of class imbalance, this study systematically varies oversampling rates, explaining that this helps optimize model performance without introducing bias or overfitting. By testing all models on real transaction samples, not generated data, the practical applicability of the findings is ensured. The objectives are directly formulated and traced through the results, highlighting the positive effects and quantitative benefits of this approach over known methods in terms of accuracy, recall, and Area Under the Curve (AUC) metrics.

1.3. Approach

This research begins by applying twelve distinct machine-learning models to an imbalanced dataset, where legitimate transactions significantly outnumber fraudulent ones. To address data imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is employed, which artificially augments the minority class of fraudulent transactions to create a more balanced dataset. Various oversampling rates are systematically explored to determine the optimal rate that enhances the model performance. The investigation identified an optimal oversampling rate of 20%, serving as a benchmark for further analysis. Under optimal conditions, the model performance is evaluated, and the top six models are identified based on their metrics. Notably, the eXtreme Gradient Boosting (XGB) model demonstrates exceptional performance when combined with the optimal oversampling rate.

To further enhance the XGB model, seven feature selection methods were employed, integrating them into an ensemble model consisting of seven distinct base learners. Each base learner represents a variant of the XGB model empowered using a unique feature selection method. The ensemble model uses a “hard voting” system, where each base learner’s prediction carries equal weight, collectively determining the authenticity of transactions.

1.4. Structure of the Article

The remainder of this paper is organized as follows: Section 2 provides the related works, while Section 3 describes the dataset used in this study and presents the proposed approach. Section 4 discusses the results. Finally, the paper concludes with the main findings and future research directions.

2. Related Works

The field of credit card fraud detection faces significant challenges due to imbalanced datasets, leading to high false positive and negative rates in existing systems. Addressing these challenges requires innovative approaches to improve the effectiveness of fraud detection mechanisms. The current literature highlights various machine learning techniques and oversampling methods to tackle these issues, but there remains a gap in understanding the impact of different oversampling rates on model performance. As shown in Table 1:

[7]: This study utilizes a credit card dataset from European cardholders and explores under-sampling, SMOTE, and AdaSyn as techniques for handling imbalanced datasets. While it achieves high accuracy, the paper ignores the temporal aspects of fraud, which can be crucial for real-world applications.

[8]: Similar to [7], this study also uses the European cardholder dataset and employs SMOTE for oversampling. Although it demonstrates accurate results, the paper lacks an analysis of the impact of hyperparameter tuning, which can significantly influence model performance.

[9]: This study applies statistical methods for feature selection and SMOTE for oversampling on the European cardholder dataset. It achieves high recall, indicating a good ability to identify positive fraud cases. However, the paper lacks a sufficient explanation of the JAD algorithm, limiting its reproducibility and clarity.

[10]: This paper introduces a novel approach using PCA and a CNN Autoencoder for feature selection, combined with Random Undersampling and SMOTE Tomek for handling imbalanced data. It claims to achieve top performance in credit card fraud detection. However, the authors do not provide comparisons with other methods, and the model's interpretability is limited.

[11]: This study explores a deep autoencoder for feature selection and uses various resampling techniques. While it successfully reduces data dimensionality, the paper lacks comparisons with other approaches and requires further investigation into the model's interpretability.

[12]: Focusing on real credit card transactions, this research emphasizes personalized detection optimization. However, it raises ethical implications that need to

be carefully addressed, especially concerning data privacy and potential biases.

[13]: This paper employs fuzzy c-means clustering for feature selection and SMOTE for oversampling. It demonstrates superior performance compared to other oversampling techniques. However, the study suffers from inadequate feature analysis and high computational complexity.

[14]: Using the IEEE-CIS fraud detection dataset, this study leverages correlation and PCA for feature selection. The key advantage lies in its use of uncertainty quantification to enhance fraud prevention. However, the paper acknowledges the need for more research to validate its findings.

[15]: This study focuses on a financial indicators dataset for listed companies and applies multiple feature selection models along with SMOTE. While it benefits from a large dataset and explores multiple algorithms, the paper lacks industry-specific and temporal financial data, limiting its applicability to real-world scenarios.

[16]: Using PCA for feature selection on the European cardholder dataset, this research achieves high accuracy and F1-score. However, it ignores the important temporal aspect of credit card transactions.

[17]: This study employs correlation-based feature selection and SMOTE for oversampling on the European cardholder dataset. It reports high accuracy and F1-score but lacks comparisons with other methods to demonstrate its superiority.

[18]: This study focuses on enhancing fraud detection using SMOTE for oversampling on a European cardholder transaction dataset. However, it lacks comparisons with other approaches and provides limited discussion on the complexity and scalability.

[19]: This research utilizes a novel CSO algorithm for feature selection on the European cardholder dataset and claims to outperform existing algorithms. However, it lacks comparisons to support this claim.

[20]: This paper applies SVM-RFE for feature selection and SMOTE for oversampling on the IEEE-CIS fraud detection dataset. It identifies an Adaboost + LGBM hybrid model as the best performer. However, the study lacks interpretability and does not address adversarial robustness.

[21]: This study explores SMOTE-ENN for oversampling on a European cardholder transaction dataset and claims to outperform widely used methods. However, it lacks interpretability, making it difficult to understand the model's decision-making process.

[22]: This study utilizes various datasets, including Bank Marketing, Vehicle Insurance, Fraudulent on Cars, Worldline & ULB, and BankSim, and reports significant performance gains. However, it highlights the need to address hyperparameter tuning and computational complexity.

[23]: This study focuses on the European cardholder dataset and claims superior performance compared to other classifiers. However, it lacks specific details about the methods used.

[24]: This research employs a genetic algorithm (GA) for feature selection and SMOTE for oversampling on the European cardholder dataset. It claims to outperform existing systems but lacks interpretability and details about computational efficiency.

[25]: This study combines a neural network ensemble with a hybrid resampling method (SMOTE-ENN) on the European cardholder dataset. However, it overlooks the important aspect of scalability.

[26]: This paper uses PCA for feature selection and SMOTE for oversampling on the European cardholder dataset. It compares the supervised and unsupervised algorithms but lacks interpretability.

[27]: This research explores quantum computing for fast fraud detection using random undersampling on the European cardholder dataset. However, it lacks discussion on scalability, which is crucial for real-world applications of quantum computing.

[28]: This study introduces a novel GNN model for fraud detection on the Sparkov dataset, demonstrating efficient graph processing and improved performance metrics. However, the computational complexity remains a concern.

[29]: This research achieves high AUPRC and AUC on the European cardholder dataset but lacks discussion on scalability.

[30]: This paper employs hybrid undersampling (Tomek links) and oversampling (BCBSMOTE) on the PaySim dataset, resulting in improved F1-score, precision, and AUPRC. However, the computational complexity is a potential drawback.

[31]: This paper introduces an ensemble model (SVM, KNN, Random Forest, Bagging, Boosting) for credit card fraud detection using the Kaggle Credit Card Fraud dataset. Under-sampling and SMOTE addressed the class imbalance. Feature selection was omitted due to anonymized features, a study limitation alongside limited adversarial attack and scalability analysis.

[32]: This study introduces a hybrid ensemble and deep learning approach for the detection of credit card fraud. Using European and Sparkov datasets, it employed oversampling, undersampling, and SMOTE to handle class imbalance, and PCA for feature selection. A key limitation is the performance gap between real-world and synthetic data, raising concerns about the model generalizability tested on synthetic data.

[33]: This article introduces a credit card fraud detection method that combines a neural network with SMOTE to tackle imbalanced datasets. Using a European dataset of 284,807 transactions (0.172% fraudulent), the proposed approach shows improved performance over

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

traditional methods. The limitations include insufficient detail on NN hyperparameters and reliance on a single dataset.

The analysis of credit card fraud detection research papers highlights various methods such as SMOTE and PCA for improving detection accuracy. Common issues include ignoring temporal aspects, lack of method comparisons, poor interpretability, and scalability challenges. Imbalanced data is a significant problem that is often addressed with oversampling techniques, but innovative solutions are needed. Future research should focus on incorporating temporal data, improving method comparisons, enhancing interpretability, ensuring scalability, and developing better strategies for handling imbalanced datasets to create more robust and practical fraud detection systems.

Table 1

Overview of the fraud detection studies

Ref.	Dataset	Feature Selection	Oversampling	Advantages	Limitations
[7]	Credit card dataset from European cardholders	N/A	Undersampling, SMOTE, AdaSyn	High accuracy	Ignores temporal fraud aspects
[8]	Credit card dataset from European cardholders	N/A	SMOTE	Accurate	Ignores hyperparameter tuning impacts
[9]	Credit card dataset from European cardholders	Statistical methods	SMOTE	High recall	Insufficient JAD algorithm explanation
[10]	Credit card dataset from European cardholders	PCA+CNN Autoencoder	Random Undersampling, SMOTE Tomek	Top credit card fraud detection	No comparison, limited interpretability
[11]	Credit card dataset from European cardholders	Deep autoencoder	Resampling techniques	Lower-dimensional data	No comparison, needs interpretability
[12]	Real credit card transactions dataset	N/A	N/A	Personalised detection optimisation	Ethical implications
[13]	Credit card dataset from European cardholders	Fuzzy c-means clustering	SMOTE	Excels over other oversampling	Inadequate feature analysis, computational complexity
[14]	IEEE-CIS fraud detection dataset	Correlation and (PCA)	N/A	Uncertainty quantification enhances fraud prevention	Requires more research for validation
[15]	Financial indicators dataset for listed companies	Multiple feature selection models	SMOTE	Large dataset, multiple algorithms	Missing industry-specific, temporal financial data
[16]	Credit card dataset from European cardholders	PCA	N/A	High accuracy and f1-score	Ignoring temporal aspect

Continuation of Table 1

Ref.	Dataset	Feature Selection	Oversampling	Advantages	Limitations
[17]	Credit card dataset from European cardholders	Correlation	SMOTE	High accuracy and F1-score	Compare methods needed
[18]	Credit card transaction dataset from European cardholders	N/A	SMOTE	Enhanced detection	No comparisons, limited complexity, scalability discussion
[19]	Credit card dataset from European cardholders	CSO	N/A	Outperforms existing algorithms	No comparisons
[20]	IEEE-CIS fraud detection dataset	SVM-RFE	SMOTE	Adaboost + LGBM hybrid model identified as champion	No interpretability, no adversarial robustness
[21]	Credit card transaction dataset from European cardholders	N/A	SMOTE-ENN	Outperforms widely used methods	No interpretability
[22]	Bank Marketing, Vehicle Insurance, Fraudulent on Cars, Worldline & ULB, and BankSim	N/A	N/A	Significant performance gains	Hyperparameter tuning, computational complexity require attention
[23]	Credit card dataset from European cardholders	N/A	N/A	Outperforms other classifiers in terms of performance metrics	Outperforms other classifiers in metrics
[24]	Credit card dataset from European cardholders	GA	SMOTE	Outperforms existing systems	Lack of interpretability, computational efficiency details
[25]	Credit card dataset from European cardholders	N/A	SMOTE-ENN	Neural network ensemble + hybrid resampling method	Scalability overlooked
[26]	Credit card dataset from European cardholders	PCA	SMOTE	Compares supervised/unsupervised algorithms	No interpretability
[27]	Credit card dataset from European cardholders	N/A	Random Undersampling	Quantum computing for fast fraud detection	No scalability discussion

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Continuation of Table 1

Ref.	Dataset	Feature Selection	Oversampling	Advantages	Limitations
[28]	Sparkov dataset	N/A	N/A	Novel GNN model, efficient graph processing, improved performance metrics	computational complexity
[29]	Credit card dataset from European cardholders	N/A	N/A	High AUPRC and AUC	No scalability discussion
[30]	PaySim	N/A	Hybrid undersampling (Tomek links) and oversampling (BCBSMOTE)	Improved F1-score (85.20%) Improved precision (81.27%) Improved AUPRC (72.77%)	Computational complexity
[31]	Credit card dataset from European cardholders	N/A	Undersampling and SMOTE	High accuracy	Scalability analysis
[32]	Credit card dataset from European cardholders and Sparkov datasets	PCA	Oversampling, undersampling, and SMOTE	Improved accuracy	Generalizability
[33]	Credit card dataset from European cardholders	N/A	SMOTE	Improved performance over traditional methods.	Insufficient detail on NN hyperparameters

3. The proposed approach

This study explores the intricate aspects of detecting fraud, particularly focusing on how different oversampling rates affect performance indicators in a range of machine learning models. The initial stage of the approach, depicted in Figure 1, involves introducing 12 diverse models to a dataset where genuine transactions outnumber fraudulent ones. To tackle this imbalance, the SMOTE is used to increase the representation of the minority fraudulent transactions, thereby achieving a more balanced dataset. The core of the study is to methodically test various oversampling rates to find the one that most enhances the models' accuracy, recall, and AUC. Through investigations, we identified 20% as the most effective oversampling rate. Under this optimal condition, we assess the models' performances and highlight the top six models with outstanding performance metrics. The eXtreme Gradient Boosting (XGB) model is particularly noteworthy for its adaptability and efficiency at this oversampling rate. To further improve the XGB model's fraud detection capability, we incorporate seven distinct feature selection methods. These methods were incorporated into an ensemble model comprising seven base learners. Each learner is a variation of the XGB model, enhanced by a different feature selection approach. This ensemble model operates on a "hard voting" mechanism, where each base learner's prediction is equally weighted, collectively determining a transaction's legitimacy. Our study, by delving into the subtle effects of oversampling rates on various machine learning algorithms, seeks to develop robust strategies for combating credit card fraud, thereby advancing the field of predictive accuracy in this essential area.

3.1. Data used

This research uses a well-known credit card fraud detection dataset, which includes 284807 transac-

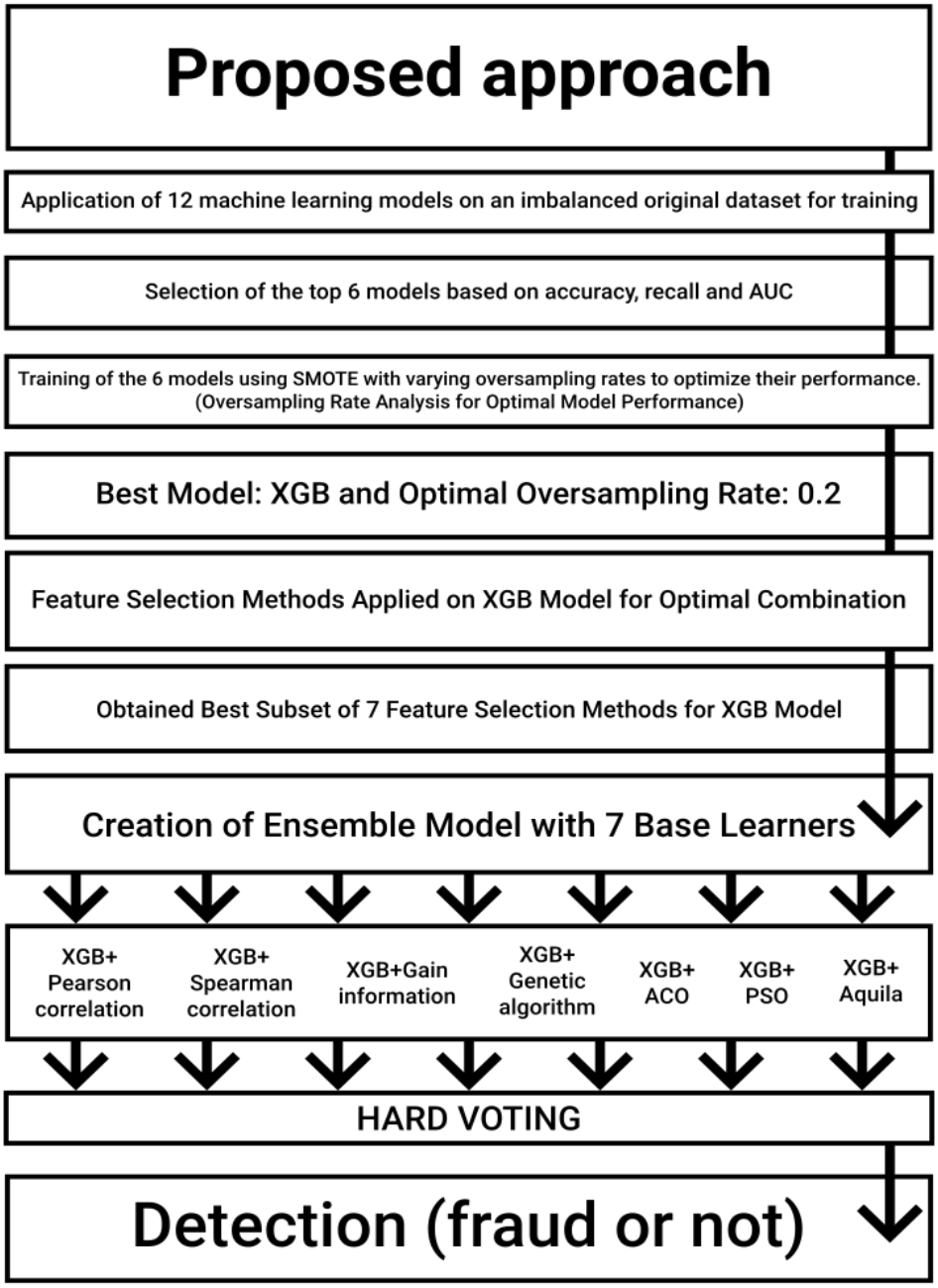

Fig.1. The Proposed approach

tions made by European cardholders over a two-day period in September 2013 [34]. Out of the total transactions, 284315 were legitimate while 492 transactions were fraudulent, accounting for only 0.172% of the dataset. As such, the dataset is severely skewed, posing challenges in developing efficient fraud detection algorithms. The dataset includes 30 numerical features (V1 to V28), as well as the time and amount of each transaction. The last column of the dataset indicates the class of the transaction, with a value of 1 representing a fraudulent transaction and a value of 0 indicating a non-fraudulent transaction. The features from V1 to V28 are not named to ensure security and integrity.

3.2. Data Preprocessing and Splitting

The dataset undergoes several preprocessing steps. Initially, it is cleaned by editing missing information, correcting errors, or removing duplicates. After cleaning, the dataset is normalized using the min-max scaling method, which ensures that all values fall within a specified range. Next, the dataset is split into training and testing sections.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

The training data, comprising 70% of the dataset, is used to train and fine-tune the machine learning models, helping them recognize patterns and relationships. The remaining 30% serves as testing data to evaluate the models' performance by predicting outcomes for new, unseen data. This approach is a standard practice in machine learning for assessing algorithm effectiveness.

3.3. Optimal Models

Twelve models were trained using the training data, and the performance of the system was evaluated using the testing data. These models include GaussanNB, MLPClassifier, XGBClassifier, DecsionTreeClassifier, CatBoostClassifier, RandomForestClassifier, AdaBoostClassifier, LGBMClassifier, GradientBoostingClassifier, KNeighborsClassifier, LogisticRegression, and SVM. Finally, six optimal models were selected based on accuracy, recall, and AUC score. Figure 2 and Algorithm 1 illustrate the methodology employed in selecting the six optimal classifiers

3.4. Resampling

After training the top six classifiers on the original imbalanced dataset, the SMOTE was employed to balance the training data, and the models were subsequently refitted.

For each of the six classifiers, SMOTE oversampling rates ranging from 0.05 to 0.5 in increments of 0.05 were systematically tested. At each oversampling rate, the training data were rebalanced using SMOTE, the classifier was retrained on the oversampled data, and its performance was evaluated. By adjusting the SMOTE oversampling rate for each classifier, the overall model performance was improved by mitigating the class imbalance. After evaluating the performance of each classifier and oversampling rate combination, the one that performed the best on the test set was selected. Algorithm 2 and Figure 3 illustrate the procedure for investigating the impact of the oversampling rate on six classifiers to determine the optimal rate for each.

3.5. Feature Selection

After selecting the best-performing classifier, various feature selection methods were evaluated to identify the ideal subset that enhances the performance of each approach. The goal was to construct a set of base learners, each consisting of the best classifier paired with a different feature selection method. The feature selection techniques we evaluated included the following:

– Pearson correlation - selects features that have a strong linear correlation with the target;

Algorithm 1 Classification Algorithm

Input: Dataset
Output: Best six classifiers

Data preprocessing:

Dataset Cleaning:

- Editing missing information.
- Correcting incorrect data.
- Removing duplicate entries.

Data Normalization:

- The dataset is normalized using the min-max scaling method.
- This normalization ensures all input values lie within a [0,1] range.

Data split: Split the dataset into training and testing sets

Train classifiers: classifiers = [NB(), MLP(), XGB(), DT(), CATBoost(), RF(), AdaBoost(), LGBM(), BG(), KNN(), LR(), SVM()]

Evaluate classifiers: Evaluate the performance of each classifier on the testing set

Select the best six classifiers: Select the six classifiers with the best performance metrics

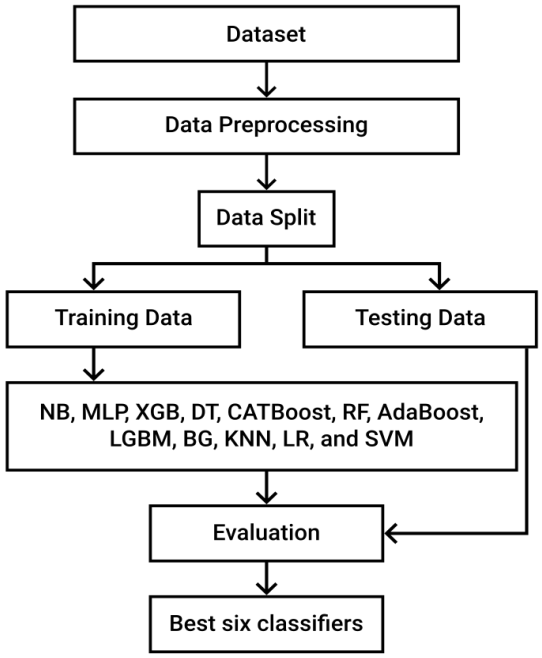

Fig. 2. Machine learning classification process

– Spearman correlation - selects features with a strong monotonic relationship to the target;

– Information gain - selects features that minimize entropy and maximize information about the target;

– Genetic algorithm uses an evolutionary approach to select feature subsets that optimize fitness;

– Particle swarm optimization (PSO) - iteratively searches for optimal features guided by swarm intelligence;

– Ant colony optimization (ACO) - uses artificial ant colonies to select features that maximize pheromone trails;

– Aquila - an optimization algorithm that iteratively adds/removes features to find an optimal subset.

Algorithm 2 Training and Evaluating Classifiers with SMOTE

Input: Original Imbalanced Dataset, Set of Six Highest-Performing Classifiers

Output: Best Classifier and Oversampling Rate Combination

Train classifiers on original imbalanced dataset

Apply SMOTE to rebalance the training data

for *each classifier in the set of six classifiers* do

  for *SMOTE oversampling rate in range 0.05 to 0.5 with increments of 0.05* do

    Rebalance the training data using SMOTE:

    Retrain the classifier on the oversampled data

    Evaluate the classifier performance

  end

  Determine the optimal SMOTE oversampling rate

end

Tune the SMOTE oversampling rate for each classifier

Evaluate the performance of each combination on the test set

Choose the best combination

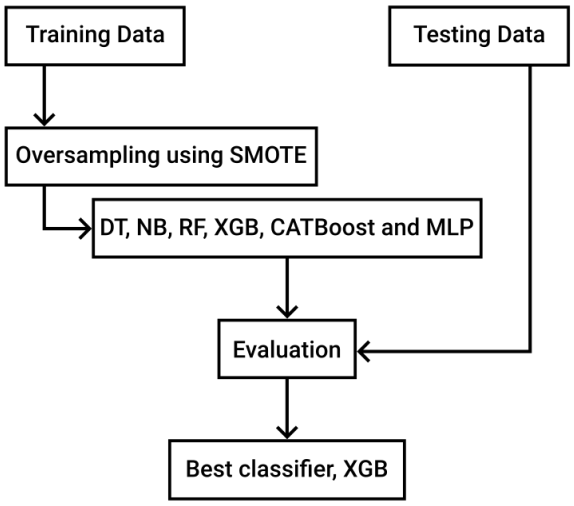

Fig. 3. Comparative Analysis of the Top Six Classifiers Enhanced by Oversampling Techniques

The best-performing classifier was selected and paired with each feature selection method to determine the optimal number of top features in the training set. This generated a set of base learners, each consisting of the top classifier and a feature set optimized for a different selection technique. Evaluating various feature selection methods allowed us to determine which techniques enhance the performance of the top classifier on this dataset. The ensemble of optimized base learners can then be combined to improve the overall predictive accuracy. Algorithm 3 and Figure 4 depict the process of selecting and evaluating various feature selection methods to identify the optimal feature subset for enhancing the performance.

Algorithm 3 Select Best Performing Classifier and Feature Selection Method

Input: Dataset, Best Performing Classifier

Output: Best Feature Selection Method, Set of Base Learners

Select best performing classifier:

for *each feature selection method* do

  Evaluate feature selection technique: Pearson correlation Spearman correlation Information gain Genetic algorithm Particle swarm optimization (PSO) Ant colony optimization (ACO) Aquila

  Select optimal number of top features

  Generate set of base learners

end

Evaluate feature selection methods

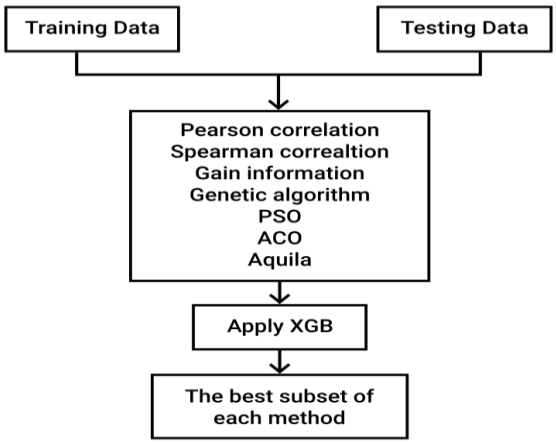

Fig. 4. Selection of Optimal Subsets for Each Method Based on Performance Metrics in the Testing Data

3.6. Ensemble Methods

After evaluating a range of feature selection methods, the top-performing classifier was paired with each optimized feature set to establish a collection of base learners. Specifically, the top classifier was paired with the feature subset selected by each of the following techniques: Pearson correlation, Spearman correlation, information gain, genetic algorithm, particle swarm optimization, ant colony optimization, and Aquila. This resulted in a set of seven base learners, where each base learner consisted of the same top classifier trained on a different optimized feature set. We then combined these base learners using a voting ensemble method to create a strong final classifier. The voting ensemble allows each base learner to independently make a prediction, and then combines the predictions by taking a vote. The final prediction is the class that receives the majority of votes across all base learners. The voting ensemble allows us to leverage the strengths of each feature selection method by training the same top classifier on different views of

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

the data and then combining the predictions. The base learners will make uncorrelated errors, and the ensemble will make a better prediction than any individual base learner. The goal of the voting ensemble is to improve generalization performance and robustness through the diversity and collective intelligence of the base learners. This approach allowed us to construct a strong final classifier that integrates multiple feature selection techniques. Algorithm 4 and Figure 5 demonstrate the process of pairing a top-performing classifier with different feature subsets, derived from a variety of selection methods, to construct a series of unique base learners. These learners are then integrated using a voting ensemble method to create a comprehensive and strong final classifier.

Algorithm 4 Ensemble Learning and Feature Selection

Input: Dataset, Set of Feature Selection Techniques
Output: Final Prediction

Evaluate various feature selection techniques:

Identify best performing classifier:

for *each optimized feature set in Pearson correlation, Spearman correlation, information gain, genetic algorithm, particle swarm optimization, ant colony optimization, Aquila* do

  Create a base learner: Combine the optimized feature set with the best performing classifier to create a base learner

end

Combine all base learners: Combine all base learners using a voting ensemble method

for *each instance to be classified* do

  Let each base learner make a prediction:

end

Combine the predictions: Combine the predictions by taking a vote

Final prediction: The final prediction is the class that receives the majority of votes across all base learners

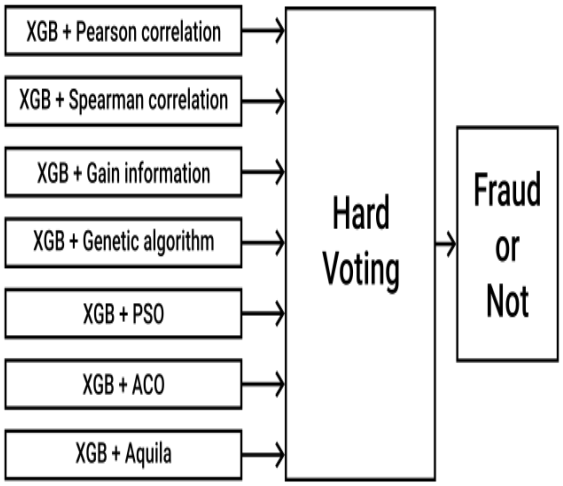

Fig. 5. Ensemble Learning Using Hard Voting Technique

4. Results and Discussion

In this section, we discuss the comprehensive findings of our tests as well as the assessment metrics we used to analyze the performance of our suggested strategy. We will also provide a thorough analysis of the findings and their implications.

4.1. Experimental Setup

The research was conducted on a cloud-based platform called Kaggle, which was outfitted with GPU hardware accelerators and a variety of libraries such as Scikit-Learn, matplotlib, sklearn, and pandas. An Intel Core i-7 3.0 GHz CPU with 8.0 GB of RAM was used in the experiment.

4.2. Performance Metrics

A confusion matrix is often used to evaluate the performance of the machine learning models. It contains the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). These counts allow for the calculation of various evaluation metrics as follows:

– Accuracy measures how often the model makes the correct prediction. It is calculated as the ratio of the correct predictions (TP + TN) to the total predictions:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}; \quad (1)$$

– Recall, also called sensitivity or true positive rate (TPR), measures the proportion of actual positives that are correctly identified. It is calculated as:

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

– The AUC (area under the ROC curve) metric summarizes the model's performance across all classification thresholds. The ROC curve plots TPR vs. FPR (false positive rate). A higher AUC indicates better overall performance.

In summary, the confusion matrix provides the basic counts to compute various metrics that evaluate different aspects of a model's predictive performance. Accuracy measures overall correctness, recall quantifies sensitivity, and AUC summarizes performance across thresholds.

4.3. Comparative Performance Metrics Analysis Across Twelve Classifiers

Table 2 displays the performance classification of twelve different categorization models, which are scored based on their accuracy, recall, and AUC.

The accuracy ratings of the models vary from 0.9662 to 0.9995, showing that all models correctly predict the target variable. However, the accuracy differences among the models were minor, with the best accuracy score being only 0.0333 higher than the lowest accuracy score.

The recall scores, which represent the proportion of true positives accurately detected by the model, show a larger range, with 0.8175 being the greatest and 0.3378 being the lowest. This suggests that certain models excel in terms of accurately recognizing the positive class.

The AUC scores, which reflect the model's ability to differentiate between positive and negative classes, exhibit a similar range of performance to the recall scores, the maximum AUC score is 0.8976, while the minimum is 0.6688.

Overall, the MLPClassifier, XGBClassifier, CatBoostClassifier, and RandomForestClassifier models outperformed the other three models. The GaussianNB model achieved a higher AUC score than some of the top-performing models, but had a lower accuracy and recall score. The AdaBoostClassifier, LGBMClassifier, GradientBoostingClassifier, KNeighborsClassifier, LogisticRegression, and SVM models performed poorly across all three measures. In summary, considering the metrics presented in Table 2, the following six models emerged as the top performers: MLPClassifier, XGBClassifier, CatBoostClassifier, RandomForestClassifier, GaussianNB, and DecisionTreeClassifier.

Table 2

Model performance metrics

Model	Accuracy	Recall	AUC
GaussianNB	0.9773	0.8175	0.8976
MLPClassifier	0.9994	0.777	0.8884
XGBClassifier	0.9995	0.7635	0.8817
DecisionTreeClassifier	0.9992	0.7432	0.8714
CatBoostClassifier	0.9994	0.7364	0.8682
RandomForestClassifier	0.9994	0.7297	0.8648
AdaBoostClassifier	0.9991	0.6689	0.8343
LGBMClassifier	0.9662	0.6486	0.8077
GradientBoostingClassifier	0.9990	0.6148	0.8072
KNeighborsClassifier	0.9992	0.5945	0.7972
LogisticRegression	0.9991	0.5878	0.7938
SVM	0.9986	0.3378	0.6688

4.4. Impact Analysis of Oversampling Rate on Machine Learning Model Performance Metrics

A few key observations from Table 3 are as follows:

– Table 3 compares six different machine learning models (Decision Tree, Naive Bayes, Random Forest, XGBoost, CatBoost, and MLP) across three evaluation metrics, namely accuracy, recall, and AUC;

– XGBoost scored the highest in terms of both accuracy (0.9996) and recall (0.8378), meaning it made the fewest false positive and false negative predictions;

– CatBoost and MLP scored the next highest in accuracy, but slightly lower in recall;

– Random Forest also scored very well in terms of accuracy (0.9995) but scored worse in recall than XGBoost, CatBoost and MLP;

– Naive Bayes had the lowest accuracy score of the models but scored higher in recall than Decision Tree and Random Forest;

– Decision Tree scored very well for accuracy, but it had the worst recall score out of the models. This suggests that it is more prone to false negatives;

– In terms of AUC, XGBoost again performed best (0.9188), followed by Naive Bayes. AUC evaluates the overall discriminative power of the model across different thresholds;

– The performance metrics of several oversampled classification models were also evaluated to determine the optimal oversampling rate for each model;

– The models were analyzed based on their accuracy, recall, and AUC scores, which were then compared to identify the best performing model at each oversampling rate;

– By evaluating the models with different oversampling rates, we can determine the rate that provides the best balance between model performance and the reduction of class imbalance in the dataset.

Overall, XGBoost outperforms the other models, excelling in accuracy, recall, and AUC. CatBoost, MLP, and Random Forest show strong results in specific metrics, while Decision Tree and Naive Bayes fall behind.

Table 3

Model performance with oversampling

Model	Oversampling rate	Accuracy	Recall	AUC
DT	0.15	0.9975	0.7837	0.8908
NB	0.05	0.9762	0.8243	0.9004
RF	0.15	0.9995	0.7972	0.8985
XGB	**0.20**	**0.9996**	**0.8378**	**0.9188**
Catboost	0.25	0.9994	0.8243	0.912
MLP	0.25	0.9988	0.8243	0.9117

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

The analysis of the oversampling rate's impact on the performance of an XGBoost model, as illustrated in Figure 6, reveals that the Area Under the Receiver Operating Characteristic Curve (AUC score) peaks at an oversampling rate of 0.20, achieving a maximum value of 0.912. This optimum point signifies the best balance between enhancing the representation of the minority class and preserving the model's generalization capability.

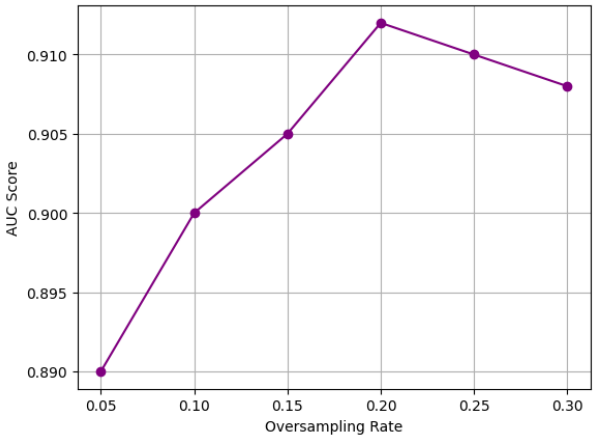

Fig.6. Performance of XGB as a Function of Oversampling Rate Measured by AUC Score

Oversampling, a technique employed to mitigate class imbalance by increasing the instances in the minority class, can enhance model performance by providing additional data for learning; however, excessive oversampling risks overfitting, where the model may learn noise instead of the essential patterns. Beyond the 0.20 rate, the AUC score begins to decline, indicating that further oversampling introduces redundancy and noise, adversely affecting the model performance. Therefore, it is recommended to select the optimum oversampling rate of 0.20 to maximize the model performance while monitoring for overfitting through regular evaluation on a validation set.

4.5. Comparative Efficacy of Feature Selection Techniques in Enhancing Model Performance

Table 4 and Figure 7 detail a comparison of seven feature selection methods in a classification model, focusing on the number and type of selected features, oversampling rate, and performance metrics like accuracy, recall, and AUC. Spearman correlation, Gain information, Genetic algorithm, PSO, and ACO stand out, achieving accuracy above 0.9994 and recall over 0.8412, indicating their effectiveness in selecting predictive features.

In terms of selected features, V14 appears in the feature sets of all methods, indicating its importance as a predictor. V10 and V11 also occurred frequently. The feature sets are mostly small, between 5-7 features, showing that these methods can effectively reduce dimensionality and select compact yet predictive subsets of features.

Spearman correlation selected the fewest features (5) with the highest accuracy, recall, and comparable AUC. It is also worth noting that ACO selected six features with the highest accuracy and recall. Overall, the feature selection methods were found to be quite effective for this classification task.

4.6. Enhancing Classifier Performance through Ensemble Integration of Diverse Feature Selection Methods

The effectiveness of integrating various feature selection methods with the XGB classifier using an ensemble approach was highlighted. This method achieved an accuracy of 0.9998, a recall of 0.8693, and an AUC of 0.9376. A range of feature selection techniques were used, including Pearson and Spearman correlation, information

Table 4

Performance comparison of various feature selection methods

Model	# of selected features	The selected features	Accuracy	Recall	AUC
XGB+Pearson correlation	5	V14', 'V12', 'V11', 'V10', 'V4	0.9995	0.8425	**0.9268**
XGB+Spearman correlation	5	V14', 'V12', 'V4', 'V10', 'V11	0.9997	0.8495	0.9245
XGB+Gain information	6	V10', 'V11', 'V12', 'V14', 'V16', 'V17	0.9997	0.8501	0.9289
XGB+Genetic algorithm	7	V3', 'V4', 'V10', 'V11', 'V14', 'V24', 'V28	0.9994	0.8412	0.9199
XGB+PSO	5	V3', 'V4', 'V10', 'V14', 'V28	0.9996	0.8483	0.925
XGB+ACO	6	V3', 'V4', 'V11', 'V14', 'V16', 'V28	**0.9997**	**0.8505**	0.9255
XGB+Aquila	5	V4', 'V10', 'V13', 'V14', 'V17	0.9996	0.8475	0.9231

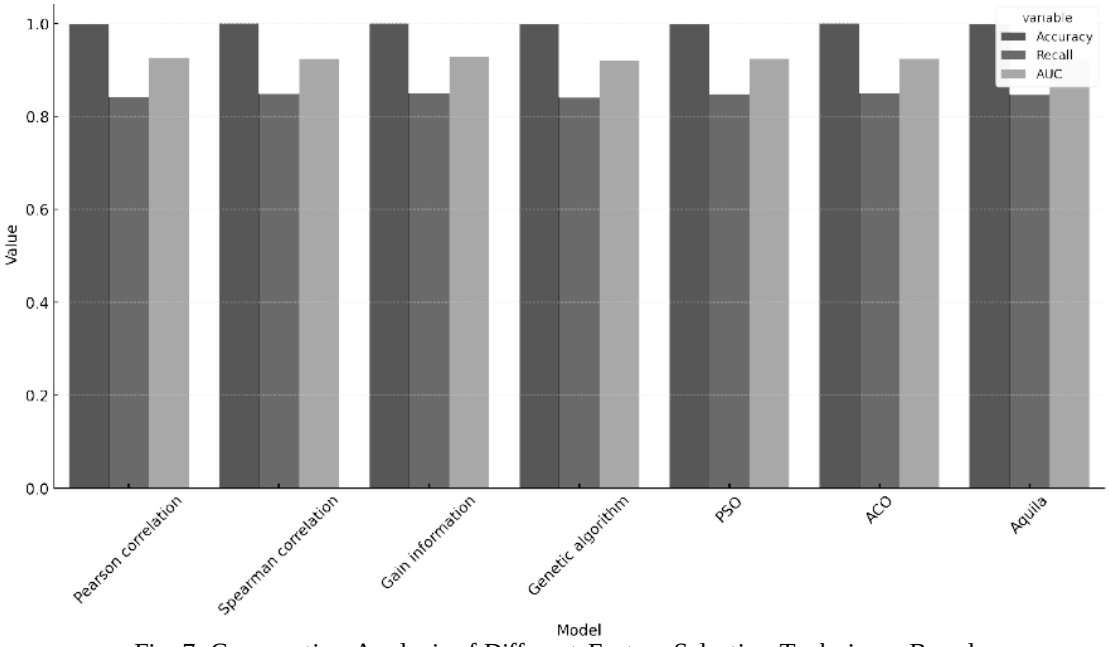

Fig. 7. Comparative Analysis of Different Feature Selection Techniques Based on Accuracy, Recall, and AUC Metrics

gain, and others, to create seven base learners. Each learner was trained on a different feature set and then combined using a voting ensemble method. This approach leveraged the diverse strengths of each method, resulting in a robust final classifier that was more accurate than any individual learner by minimizing uncorrelated errors.

4.7. Outperforming Current State-of-the-Art Techniques: Comparative Analysis of Fraud Detection in Credit Card Transactions

Table 5 displays a comparison between our proposed method and other recent approaches found in the literature, which we selected due to their use of the same dataset. Overall, these existing methods show satisfactory results; however, our approach outperforms most other techniques. A key advantage of our method is the systematic analysis of different machine learning algorithms and feature selection techniques. Additionally, we investigated various oversampling rates to determine the optimal rate for each of the top six models. In contrast to other oversampling-based approaches that oversampled the data before splitting it into training and test sets, our method oversampled only the training data. This prevents synthetic samples from being included in the test set and allows for a more robust evaluation of the model performance. Furthermore, unlike previous studies that used a single oversampling rate (typically 0.5), we optimized this hyperparameter for each model. Evaluating our approach by examining multiple datasets, rather than just one, would strengthen these conclusions and is an area for future work. In summary, our methodological improvements, including tuning oversampling rates and comparing various learning algorithms, are critical factors that allow it to surpass current state-of-the-art techniques.

5. Conclusion and future work

This paper provides a comprehensive analysis of credit card fraud detection, focusing on the impact of varying oversampling rates and the implementation of a novel ensemble method. Specifically, this study investigates the influence of SMOTE oversampling on key performance metrics, revealing an optimal rate of 20% for maximizing accuracy, recall, and AUC scores. This optimal rate balances the benefits of addressing class imbalance with the risks of overfitting and noise amplification. A key innovation of this work lies in the introduction of an ensemble approach that integrates multiple feature selection methods with the XGBoost classifier. This approach significantly outperforms the traditional methods, achieving remarkable results: 99.98% accuracy, 86.93% recall, and 93.76% AUC. These metrics demonstrate the ensemble model's effectiveness in identifying fraudulent transactions while minimizing false positives and negatives. High recall is particularly crucial in fraud detection, as it minimizes the number of fraudulent transactions that go undetected. A critical aspect of the research methodology involved rigorous testing on real-world transaction data, ensuring the model's robustness and practical applicability. This focus on real-world data contrasts with many studies that rely solely on synthetic datasets. By using real transaction data, the research ensures that the model can effectively handle the complexities and the

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Table 5

Performance comparison of different techniques on the credit card transaction dataset

Ref	Dataset	Technique	Evalution Metrics	Results
[7]	Credit card transactions dataset from European cardholders	RF	Accuracy, Precision, Recall and F1-score	Accuracy: 0.9996, Precision: 0.9130, Recall: 0.8571, F1-score: 0.8842
[8]		AdaBoost +XGB	Accuracy, Precision, Recall	Accuracy: 0.9998, Precision: 0.9992, Recall: 0.9997
[9]		JAD	Accuracy, Precision, Recall, F1-score and AUC	Accuracy: 0.9990, Precision: 0.7960, Recall: 0.8480, F1-score: 0.8210, AUC: 0.9810
[11]		Deep AutoEncoder +PCA	Accuracy, Precision, Recall, F1-score and AUC	Accuracy: 0.9990, Precision: 0.9777, Recall: 0.7212, F1-score: 0.8302, AUC: 0.7630
[26]		RF	Accuracy, Precision, andAUC	Accuracy: 0.9800, Precision: 0.8900, AUC: 0.9500
[27]		Bagged tree	Accuracy andAUC	Accuracy: 0.9704, AUC: 0.9489
Our approach		XGB	Accuracy, Recall and AUC	Accuracy: 0.9998, Recall: 0.9358, AUC: 0.9817

nuances of actual credit card transactions. While the results are promising, the study acknowledges the limitation of relying on a single dataset. Future research will address this by validating the methodology across diverse datasets to ensure its generalizability. Further exploration of alternative oversampling and downsampling techniques, along with the integration of diverse data sources, is also planned.

Contributions of authors: Conceptualization, Methodology – **Mohamed AKOUHAR**, **Abdallah ABARDA**, **Mohamed EL FATINI**, **Mohamed OUHSSINI**; Formulation of Tasks, Analysis – **Mohamed AKOUHAR**, **Abdallah ABARDA**, **Mohamed OUHSSINI**; Development of Model, Software, Verification – **Mohamed AKOUHAR**, **Mohamed OUHSSINI**; Analysis of Results, Visualization – **Mohamed AKOUHAR**, **Mohamed OUHSSINI**; Writing – Original Draft Preparation, Writing – Review and Editing – **Mohamed AKOUHAR**, **Abdallah ABARDA**, **Mohamed EL FATINI**.

Conflict of Interest

The authors declare that they have no conflict of interest.

Financing

This study was conducted without financial support.

Data Availability

The data can be accessed through this link https://www.kaggle.com/mlg- ulb/creditcardfraud.

Use of Artificial Intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

All the authors have read and agreed to the published version of this manuscript.

References

1. Rivera, K., Rohn, C., Donker, J., & Butter, C. *Fighting fraud: A never-ending battle*. *PwC's Global Economic Crime and Fraud Survey*, 2020. 14 p. Available at: https://www.pwc.com/gx/en/forensics/gecs-2020/pdf/global-economic-crime-and-fraud-survey-2020.pdf. (accessed 1.08.2024).

2. Zhang, Y.-F., Lu, H.-L., Lin, H.-F., Qiao, X.-C., & Zheng, H. The optimized anomaly detection models based on an approach of dealing with imbalanced dataset for credit card fraud detection. *Mobile Information Systems*, 2022, no. 1, article no. 8027903. DOI: 10.1155/2022/8027903.

3. Li, L., Liu, Z., Chen, C., Zhang, Y.-L., Zhou, J., & Li, X. *A time attention based fraud transaction detection framework*. arXiv preprint. DOI: 10.48550/arXiv.1912.11760.

4. Kültür, Y., & Çağlayan, M. U. Hybrid approaches for detecting credit card fraud. *Expert Systems*, 2017, no. 34, article no. e12191. DOI: 10.1111/exsy.12191.

5. Kurshan, E., & Shen, H. Graph computing for financial crime and fraud detection: Trends, challenges and outlook. *International Journal of Semantic Computing*, 2020, vol. 14, no. 04, pp. 565-589. DOI: 10.1142/S1793351X20300022.

6. Van Vlasselaer, V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. Apate: A novel approach for automated credit card transaction fraud detection using network-based extensions. *Decision Support Systems*, 2015, vol. 75, pp. 38-48. DOI: 10.1016/j.dss.2015.04.013.

7. Hossain, M. A., Khatun, M. S., Bhuiyan, R. A., & Taslim, M. Handling class imbalance in credit card fraud using various sampling techniques. *American Journal of Multidisciplinary Research and Innovation (AJMRI)*, 2022, vol. 1, no. 4, pp. 1-6. DOI: 10.54536/ajmri.v1i4.633.

8. Ileberi, E., Sun, Y., & Wang, Z. Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost. *IEEE Access*, 2021, vol. 9, pp. 165286-165294. DOI: 10.1109/ACCESS.2021.3134330.

9. Plakandaras, V., Gogas, P., Papadimitriou, T., & Tsamardinos, I. Credit card fraud detection with automated machine learning systems. *Applied Artificial Intelligence*, 2022, vol. 36, no. 1, article no. e2086354. DOI: 10.1080/08839514.2022.2086354.

10. Salekshahrezaee, Z., Leevy, J. L., & Khoshgoftaar, T. M. The effect of feature extraction and data sampling on credit card fraud detection. *Journal of Big Data*, 2023, vol. 10, no. 1, article no. 6. DOI: 10.1186/s40537-023-00684-w.

11. Fanai, H., & Abbasimehr, H. A novel combined approach based on deep autoencoder and deep classifiers for credit card fraud detection. *Expert Systems with Applications*, 2023, vol. 217, article no. 119562. DOI: 10.1016/j.eswa.2023.119562.

12. Buonaguidi, B., Mira, A., Bucheli, H., & Vitanis, V. Bayesian quickest detection of credit card fraud. *Bayesian Analysis*, 2022, vol. 17, no. 1, pp. 261-290. DOI: 10.1214/20-BA1254.

13. Ahmad, H., Kasasbeh, B., AL-Dabaybah, B., & Rawashdeh, E. EFN-SMOTE: An effective oversampling technique for credit card fraud detection by utilizing noise filtering and fuzzy C-means clustering. *International Journal of Data and Network Science*, 2023, vol. 7, no. 4, pp. 1025-1032. DOI: 10.5267/j.ijdns.2023.6.003.

14. Habibpour, M., Gharoun, H., Mehdipour, M., Tajally, A., Asgharnezhad, H., Shamsi, A., Khosravi, A., & Nahavandi, S. Uncertainty-aware credit card fraud detection using deep learning. *Engineering Applications of Artificial Intelligence*, 2023, vol. 123, article no. 106248. DOI: 10.1016/j.engappai.2023.106248.

15. Zhao, Z., & Bai, T. Financial fraud detection and prediction in listed companies using SMOTE and machine learning algorithms. *Entropy*, 2022, vol. 24, no. 8, article no. 1157. DOI: 10.3390/e24081157.

16. Mohsen, O. R., Nassreddine, G., & Massoud, M. Credit card fraud detector based on machine learning techniques. *Journal of Computer Science and Technology Studies*, 2023, vol. 5, no. 2, pp. 16-30. DOI: 10.32996/jcsts.2023.5.2.2.

17. Aburbeian, A. M., & Ashqar, H. I. Credit card fraud detection using enhanced random forest classifier for imbalanced data. In *International Conference on Advances in Computing Research*, 2023, Orlando, USA, Springer, pp. 605-616. DOI: 10.1007/978-3-031-33743-7_48.

18. Tripathy, N., Nayak, S. K., Godslove, J. F., Friday, I. K., & Dalai, S. S. Credit card fraud detection using logistic regression and SMOTE approach. *International Journal of Computer and Communication Technology*, 2022, vol. 8, no. 4, pp. 38-47. DOI: 10.47893/IJCCT.2022.1438.

19. Karthikeyan, T., Govindarajan, M., & Vijayakumar, V. An effective fraud detection using competitive swarm optimization based deep neural network. *Measurement: Sensors*, 2023, vol. 27, article no. 100793. DOI: 10.1016/j.measen.2023.100793.

20. Malik, E. F., Khaw, K. W., Belaton, B., Wong, W. P., & Chew, X. Y. Credit card fraud detection using a new hybrid machine learning architecture. *Mathematics*, 2022, vol. 10, no. 9, article no. 1480. DOI: 10.3390/math10091480.

21. Mienye, I. D., & Sun, Y. A deep learning ensemble with data resampling for credit card fraud detection. *IEEE Access*, 2023, vol. 11, pp. 30628-30640. DOI: 10.1109/ACCESS.2023.3262020.

22. Xu, B., Wang, Y., Liao, X., & Wang, K. Efficient fraud detection using deep boosting decision trees. *Decision Support Systems*, 2023, vol. 175, article no. 114037. DOI: 10.1016/j.dss.2023.114037.

23. Roseline, J. F., Naidu, G. B. S. R., Pandi, V. S., Rajasree, S. A. A., & Mageswari, N. Autonomous credit card fraud detection using machine learning approach. *Computers and Electrical Engineering*, 2022, vol. 102, article no. 108132. DOI: 10.1016/j.compeleceng.2022.108132.

24. Ileberi, E., Sun, Y., & Wang, Z. A machine learning based credit card fraud detection using the GA algorithm for feature selection. *Journal of Big Data*, 2022, vol. 9, no. 1, article no. 24. DOI: 10.1186/s40537-022-00573-8.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

25. Esenogho, E., Mienye, I. D., Swart, T. G., Aruleba, K., & Obaido, G. A neural network ensemble with feature engineering for improved credit card fraud detection. *IEEE Access*, 2022, vol. 10, pp. 16400-16414. DOI: 10.1109/ACCESS.2022.3148298.

26. Chang, V., Doan, L. M. T., Di Stefano, A., Sun, Z., & Fortino, G. Digital payment fraud detection methods in digital ages and industry 4.0. *Computers and Electrical Engineering*, 2022, vol. 100, article no. 107734. DOI: 10.1016/j.compeleceng.2022.107734.

27. Shabbir, A., Shabir, M., Javed, A. R., Chakraborty, C., & Rizwan, M. Suspicious transaction detection in banking cyber-physical systems. *Computers and Electrical Engineering*, 2022, vol. 97, article no. 107596. DOI: 10.1016/j.compeleceng.2021.107596.

28. Cherif, A., Ammar, H., Kalkatawi, M., Alshehri, S., & Imine, A. Encoder–decoder graph neural network for credit card fraud detection. *Journal of King Saud University-Computer and Information Sciences*, 2024, vol. 36, no. 3, article no. 102003. Elsevier. DOI: 10.1016/j.jksuci.2024.102003.

29. Leevy, J. L., Hancock, J., & Khoshgoftaar, T. M. Comparative analysis of binary and one-class classification techniques for credit card fraud data. *Journal of Big Data*, 2023, vol. 10, no. 1, article no. 118. DOI: 10.1186/s40537-023-00794-5.

30. Alamri, M., & Ykhlef, M. Hybrid under-sampling and oversampling for handling imbalanced credit card data. *IEEE Access*, 2024, vol. 12, pp. 14050-14060. DOI: 10.1109/ACCESS.2024.3357091.

31. Khalid, A. R., Owoh, N., Uthmani, O., Ashawa, M., Osamor, J., & Adejoh, J. Enhancing credit card fraud detection: an ensemble machine learning approach. *Big Data and Cognitive Computing*, 2024, vol. 8, no. 1, article no. 6. DOI: 10.3390/bdcc8010006.

32. Dhandore, D., Agrawal, C., & Meena, P. Enhancing credit card fraud detection through advanced ensemble learning techniques and deep learning integration. *International Journal of Engineering Applied Science and Management*, 2024, vol. 5, no. 9, pp. 1-9. Paper ID: 2024/IJEASM/5/2024/2185.

33. Zhu, M., Zhang, Y., Gong, Y., Xu, C., & Xiang, Y. *Enhancing credit card fraud detection: A neural network and SMOTE integrated approach*. arXiv preprint/ DOI: 10.48550/arXiv.2405.00026.

34. *Credit card fraud detection*. Available at: https://www.kaggle.com/mlg-ulb/creditcardfraud. (accessed 12 June 2024).

Received 15.06.2024, Accepted 17.02.2025

ПОКРАЩЕННЯ ВИЯВЛЕННЯ ШАХРАЙСТВА З КРЕДИТНИМИ КАРТКАМИ: ВПЛИВ ЧАСТОТИ НАДМІРНОЇ ДИСКРЕТИЗАЦІЇ ТА МЕТОДІВ СУКУПНОСТІ З РІЗНОМАНІТНИМ ВИБОРОМ ФУНКЦІЙ

М. Акухар, А. Абарда,
М. Ель Фатіні, М. Ухсіні

Предметом цієї статті є вдосконалення систем виявлення шахрайства з кредитними картками шляхом вивчення впливу частоти надлишкової дискретизації та методів ансамблю з різноманітними техніками вибору функцій. Шахрайство з кредитними картками стало серйозною проблемою у фінансовому світі, що призвело до значних втрат як для фінансових установ, так і для споживачів. Оскільки обсяг транзакцій з кредитними картками продовжує зростати, точне виявлення шахрайства стає дедалі складнішим. Мета цього дослідження полягає в тому, щоб покращити виявлення шахрайства з кредитними картками шляхом аналізу частот передискретизації, щоб вибрати оптимальну для найефективніших моделей, і використовуючи методи ансамблю, засновані на різноманітних підходах до вибору функцій. Основні завдання, які виконуються в цьому дослідженні, включають оцінку ефективності моделей на основі точності, запам'ятовування та показників AUC, аналіз ефекту передискретизації за допомогою методу синтетичної передискретизації меншості (SMOTE) і пропонування методу ансамблю, який поєднує сильні сторони різні методи вибору ознак і класифікатори. Методи, які використовуються в цьому дослідженні, передбачають застосування низки методів машинного навчання, включаючи логістичну регресію, дерева рішень, випадкові ліси та посилення градієнта, до незбалансованого набору даних, де легітимних транзакцій значно перевищує кількість шахрайських. Щоб усунути дисбаланс даних, дослідники систематично досліджують вплив різних частот передискретизації за допомогою SMOTE. Крім того, вони розробляють модель ансамблю, яка об'єднує сім методів вибору функцій із алгоритмом eXtreme Gradient Boosting (XGB). Результати цього дослідження показують, що застосування SMOTE суттєво покращує продуктивність моделей машинного навчання з оптимальною частотою передискретизації 20%. Модель XGB виділялася своєю винятковою продуктивністю, високою точністю, показниками запам'ятовування та AUC. Крім того, запропонований комплексний підхід, який поєднує в собі сильні сторони різноманітних методів вибору ознак і класифікатора XGB, додатково підвищує точність виявлення та продуктивність системи порівняно з традиційними методами. Висновки, зроблені в результаті цього дослідження, сприяють розвитку галузі виявлення шахрайства з кредитними картками, надаючи розуміння впливу надмірної

вибірки та переваг методів ансамблю з різноманітним вибором функцій. Ці відомості можуть допомогти в розробці більш ефективних і надійних систем виявлення шахрайства, допомагаючи фінансовим установам і споживачам краще захищатися від зростаючої загрози шахрайства з кредитними картками.

Ключові слова: шахрайство з кредитними картками; машинне навчання; виявлення шахрайства; ансамблеві методи; вибір функції; SMOTE.

Акухар Мохамед– аспірант лабораторії аналізу, геометрії та застосувань, Державний університет в Кенітрі, Марокко.

Абарда Абдаллах– доктор статистики та аналізу даних, професор лабораторії математичного моделювання та економічних розрахунків, Перший університет Хасана, Сеттат, Марокко.

Ель Фатіні Мохамед – доктор числового аналізу, професор лабораторії аналізу, геометрії та застосувань, Державний університет в Кенітрі, Марокко.

Ухсіні Мохамед – докторант лабораторії SIV, Університет Зохр, Агадір, Марокко.

Mohamed Akouhar – PhD Student in Lab. of Analysis, Geometry, and Applications, University ibn Tofail, Kenitra, Morocco,
e-mail: m.akouhar@gmail.com, ORCID: 0009-0008-5914-6764, Scopus Author ID: 58881341400.

Abdallah Abarda – Doctor of Statistics and Data Analysis, Professor, Lab of Mathematical Modeling and Economic Calculations, Hassan First University, Settat, Morocco,
e-mail: abardabdallah@gmail.com, ORCID: 0000-0002-9408-4991, Scopus Author ID: 57024196600.

Mohamed El Fatini – Doctor of Numerical Analysis, Professor, Lab. of Analysis, Geometry, and Applications, University ibn Tofail, Kenitra, Morocco,
e-mail: Mohamed.elfatini@uit.ac.ma, ORCID: 0000-0002-1179-4366, Scopus Author ID: 25225378500.

Mohamed Ouhssini – PhD Student in Lab SIV, University ibn Zohr, Agadir, Morocco,
e-mail: mohamed.ouhsini@gmail.com, ORCID: 0000-0002-3851-9962, Scopus Author ID: 57223425494.