I.J. Information Engineering and Electronic Business, **2026, 2, 55-72**
Published Online on April 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2026.02.04

Feature Selection for Fraud Detection: Improving Machine Learning Capabilities on Portable Wallet

Gurleen Kaur
Computer Applications, CT University, Ferozepur Road, Sidhwan Khurd -142024, Punjab, India
E-mail: kimtgurleenkaur@gmail.com
ORCID iD: https://orcid.org/0000-0002-3232-6934

Mandeep Kaur
Computer Science and Engineering, CT University, Ferozepur Road, Sidhwan Khurd -142024, Punjab, India
E-mail: mandeepdhamoo@gmail.com
ORCID iD: https://orcid.org/0000-0002-3232-6983

Punam Rattan
Department of Computer Science & Technology, Manav Rachna University Faridabad-121010, Haryana, India
E-mail: punamrattan@gmail.com
ORCID iD: https://orcid.org/0000-0002-2288-6987

Mukesh Kumar*
Advanced Centre of Research and Innovation (ACRI), School of Advance Computing, CGC University Mohali-140307, Punjab, India
Faculty of Law, Sohar University, Sohar 311, Sultanate of Oman
E-mail: mukesh.kumarphd2014@gmail.com
ORCID iD : https://orcid.org/0000-0001-8797-9810
*Corresponding Author

Received: 13 April, 2025; Revised: 21 June, 2025; Accepted: 22 July, 2025; Published: 08 April, 2026

Abstract: The rapid growth of mobile wallet usage has led to a sharp increase in fraudulent transactions, making fraud detection in portable wallets a pressing concern. Accurately detecting fraud is difficult because transaction data is complicated and unbalanced. Conventional rule-based systems are less flexible and frequently provide large false positive rates along with poor accuracy. Effective feature selection is crucial to the performance of Machine Learning (ML) models, notwithstanding their increased detection rates. Redundancy and noise are introduced by high-dimensional data, which lowers model performance and raises computing costs. The advantages of hybrid feature selection are frequently overlooked in current research, particularly when it comes to portable wallet fraud detection. By combining Random Forest Importance, LASSO Regression, Recursive Feature Elimination (RFE), and Mutual Information (MI) with resampling to solve class imbalance, this study fills that gap. Our approach provides a more reliable and effective solution for safe portable wallet fraud detection by removing superfluous features, increasing accuracy, and reducing computing cost. The model becomes faster and more effective when superfluous characteristics are eliminated because this reduces the computational effort. By concentrating just on the most instructive data, it increases accuracy. By addressing class imbalance and combining several selection strategies, the hybrid approach guarantees robustness. All things considered, this leads to a scalable and safe fraud detection system for transactions using mobile wallets. Our results show that a successful feature selection approach improves fraud detection accuracy, which in turn improves operational effectiveness and financial security.

Index Terms: Fraud Detection, Feature Selection, Random Forest, LASSO Regression, Recursive Feature Elimination, Mutual Information

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

1. Introduction

Consumer payment habits have been drastically changed by the remarkable rise in the adoption and use of portable wallets brought about by the digital transformation of financial transactions. This change has become a pillar of the global economy in addition to changing how people interact financially on an individual basis. But at the same time, this growing digital ecosystem opens new avenues for sophisticated fraud, which calls for sophisticated and flexible security measures. Machine learning (ML) has become a game-changing tool in the fight against fraud because of its unmatched capacity to identify intricate and unique patterns. However, its use in this field presents several special and important difficulties that need to be carefully resolved to be implemented successfully. According to preliminary data we have gathered on the use of portable wallets, the market is expected to rise significantly, reaching $8404.8 billion by 2029. Growing smartphone usage and contactless payment uptake are the main drivers of this expansion. At the same time, I've noticed a worrying trend in payment fraud, with recorded losses from fraud rising by 25% to $12.5 billion in 2024. Imposter scams and Business Email Compromise (BEC) are emphasized as serious risks, with investment scams accounting for most reported losses. This emphasizes how important it is to have strong fraud detection systems in the growing ecosystem of portable wallets [1]. The dynamic nature of fraud patterns and the large amount of transaction data make real-time fraud detection a difficult task. In the context of portable wallet payments, fraud detection becomes even more challenging due to the nature of digital transactions, which are often executed instantly with limited verification mechanisms. However, not all features are required because many of them are superfluous or redundant, which could reduce the performance of an algorithm, like a classification algorithm [2, 3]. Feature selection aims to solve this problem by selecting only a small subset of relevant features from the original large set of features [4]. By removing unnecessary and unnecessary features, feature selection can reduce the dimensionality of the data, speed up the learning process, simplify the learnt model, and/or enhance performance. Improving the effectiveness of fraud detection models requires addressing class imbalance, especially when it comes to digital wallet transactions when the number of fraudulent activities is far exceeded by those that are valid [5]. To fill this gap, we propose a novel hybrid feature selection approach that combines the strengths of multiple techniques Random Forest Importance, LASSO Regression, RFE, and MI to identify the most informative features. Additionally, we apply resampling strategies, including Synthetic Minority Over-sampling Technique (SMOTE) and random under-sampling, to correct class imbalance and improve model sensitivity to minority (fraudulent) cases. To lessen this imbalance, methods like random under-sampling and the SMOTE are used. To increase the minority class's representation in the dataset and help the model better understand the characteristics that mark fraudulent transactions apart, SMOTE creates synthetic samples for the minority class. Random under-sampling, on the other hand, lowers the amount of majority class occurrences, which may assist the model concentrate more on the minority class but run the risk of removing potentially valuable data [6]. By using these resampling techniques before feature selection, the model's capacity to identify fraudulent activity is improved because the characteristics chosen are representative and instructive for both groups [7]. Portable wallet systems can greatly improve their fraud detection capabilities by carefully choosing pertinent features and utilizing strong machine learning models. This method maintains a smooth user experience by guaranteeing efficient processing in addition to increasing the accuracy of identifying fraudulent transactions. To adjust to changing fraud patterns and preserve the efficacy of the fraud detection system, the model must be periodically retrained and monitored continuously [8].

Our contribution lies in presenting an integrated, hybrid framework that not only improves fraud detection accuracy but also reduces computational cost and enhances robustness. The proposed model adapts to evolving fraud patterns and provides a practical solution for secure, real-time fraud detection in mobile wallet systems.

2. Related Work

Several studies have investigated applying machine learning techniques to detect fraud in digital payments. To increase detection accuracy, feature selection techniques including Principal Component Analysis (PCA), MI, and RFE have been used. Nevertheless, not much study has been done expressly on Portable Wallet [9]. Previous studies demonstrate the efficacy of hybrid feature selection strategies, which integrate several approaches to minimize redundant data and extract the most useful features. The benefits of filter, wrapper, and embedded feature selection techniques are combined in hybrid feature selection techniques, which guarantee that the chosen features maximize fraud detection while preserving computing efficiency [10]. While wrapper techniques like RFE progressively remove features based on model performance, filter techniques like MIrank features according to their statistical significance. During the training phase, features are chosen using embedded techniques like LASSO regression and decision tree-based feature importance. Hybrid approaches decrease overfitting and increase classification accuracy by combining various techniques [11]. Few studies have thoroughly examined feature selection approaches' suitability for use in UPI transactions, despite their success in fraud detection. Specialized feature selection techniques are required due to the peculiarities of UPI fraud, which frequently involves dynamic transaction patterns [12]. By applying and assessing a variety of feature selection strategies designed especially for UPI-based fraud detection, considering elements like real-time detection limitations, feature redundancy, and interpretability, this work seeks to close this gap. Their study relied

heavily on feature selection, as they refined the input variables using entropy-based techniques and information gain. According to their research, decision tree techniques performed better than SVM in most cases, particularly when assessing suspicion scores, which measure the likelihood that a transaction is fraudulent. The study did not thoroughly investigate hyperparameter adjustment, which could further optimize these models despite their good classification performance[13]. Several research have used under sampling and synthetic data synthesis to solve class imbalance, a prevalent problem in fraud detection. To reduce the possibility of models being skewed in favour of the majority class, the SMOTE is frequently employed to create fictitious fraudulent transactions. Furthermore, to ensure that classifiers are trained on more difficult scenarios, NearMiss, an undersampling strategy, chooses real transactions that are near fraudulent samples. To improve prediction reliability while preserving computing economy, some researches have integrated SMOTE with ensemble learning techniques as RF or Boosting algorithms [14]. Feature selection is incorporated into the model training procedure by embedded approaches. L1 regularization is used in LASSO regression to decrease feature coefficients, hence removing less significant variables. Similarly, decision tree-based models (like RF and XGBoost) are effective for high-dimensional datasets since they use split frequency to determine feature relevance. Furthermore, although computationally demanding, interpretability-driven techniques like Shapley Additive explanations (SHAP) offer insights into feature contributions [14]. Research on fraud detection has been greatly impacted by these many feature selection strategies, each of which has its own pros and disadvantages. The size of the dataset, the need for model interpretability, and computing limitations all influence the technique selection, highlighting how crucial it is to optimize feature selection for higher fraud detection accuracy. Filter approaches do not use a learning algorithm; instead, they choose features based on statistical qualities [15]. For example, Correlation-Based Feature Selection (CFS) works well with high-dimensional data because it eliminates superfluous variables by examining feature interdependencies. While MI measures the dependence between variables to preserve the most informative ones, the Chi-Square Test assesses the importance of categorical features. Despite their computational efficiency, some approaches might miss intricate feature interactions [16,17]. Another important consideration is model interpretability; whereas deep learning-based selection approaches may function as black boxes, techniques such as SHAP and decision tree-based feature importance offer insights. Computational limitations are particularly important since, in contrast to more straightforward statistical techniques, techniques like RFE and Genetic Algorithms (GA) can be resource-intensive [18,19]. To improve fraud detection effectiveness, a well-optimized feature selection technique ultimately strikes a balance between accuracy, efficiency, and interpretability .

3. Proposed Methodology

Dataset: To investigate fraud detection in portable wallets, the PaySim dataset [20] replicates actual mobile financial transactions. Containing important details such as the kind of transaction, the amount, the balances before and after the transaction, and the fraud labels. The dataset is appropriate for testing fraud detection methods since it depicts realistic transaction behaviour. The dataset, which includes 11 columns and 1,048,575 rows, records a variety of financial transaction-related information. In the simulation, the step column indicates the time step, while the type column indicates the type of transaction, such as "cash_out" or "transfer." The transaction amount is shown in the amount column. Account identifiers are also included in the collection; the originating account is denoted by “nameOrig”, and the destination account by “nameDest.” While “oldbalanceDest” and “newbalanceDest” indicate the recipient's balance before and after the transaction, “oldbalanceOrg” and “newbalanceOrig” indicate the sender's balance before and after the transaction, respectively. Furthermore, the is “FlaggedFraud” column shows whether a transaction was marked as possibly fraudulent, and the isFraud column tells whether a transaction is fraudulent (1) or not (0).

3.1 Rationale for Dataset Selection: The PaySim Dataset

The quality and representativeness of the training data are intrinsically linked to the effectiveness of machine learning models in fraud detection. The lack of reliable, publicly accessible information on mobile money transactions is a major barrier to study in the field of portable wallet fraud detection. For security and professional reasons, real-world data cannot be publicly disclosed due to the inherently private and confidential nature of financial transactions. As a result, before beginning their real research, researchers frequently encounter the significant difficulty of first obtaining or creating appropriate datasets. This study makes use of the PaySim dataset to get over this crucial issue with data availability. A financial simulator called PaySim was created especially to create fictitious mobile money transactions. By modeling mobile money transactions using a sample of genuine transactions taken from the financial logs of a mobile money service that operates in an African nation, it overcomes the dearth of authentic data. PaySim can create synthetic datasets that closely mimic the statistical characteristics and behaviour of actual mobile money transactions by using statistical analysis and Multi-Agent-Based Simulation (MABS). The relevance of the PaySim dataset to portable wallet fraud detection is multi-faceted:

- Direct Simulation of Portable Money: PaySim explicitly simulates mobile money transactions, which are a core component of portable wallet systems. This direct alignment ensures that the dataset reflects the operational context of portable wallets.
- Inclusion of Diverse Transaction Types: The dataset encompasses five key transaction types as shown in Step 4: CASH-IN, CASH-OUT, DEBIT, PAYMENT, and TRANSFER. These categories cover a broad spectrum of activities typical in digital wallet ecosystems.
- To replicate fraudulent operations where agents try to make money by taking over customer accounts, transferring money, and cashing out of the system, PaySim is made to introduce harmful behavior into the simulated transactions. This closely resembles popular forms of digital wallet fraud, like payment fraud and account takeovers.
- Rich Feature Set: NameOrig (originating customer), oldbalanceOrg, newbalanceOrig, nameDest (receiver), oldbalanceDest, newbalanceDest, step (representing an hour of time), kind of transaction, value, and an isFraud flag are among the pertinent attributes included in the dataset. In the context of mobile payments, these characteristics are essential for creating and assessing fraud detection models.
- Validation in Current Research: Several studies have utilized and validated the PaySim dataset for fraud detection research. Using this dataset and a variety of machine learning algorithms, researchers have achieved outstanding performance measures, such as high accuracy, precision, recall, and F1-score. Its popularity as a trustworthy stand-in for actual mobile money transaction data in scholarly research is highlighted by its extensive use.

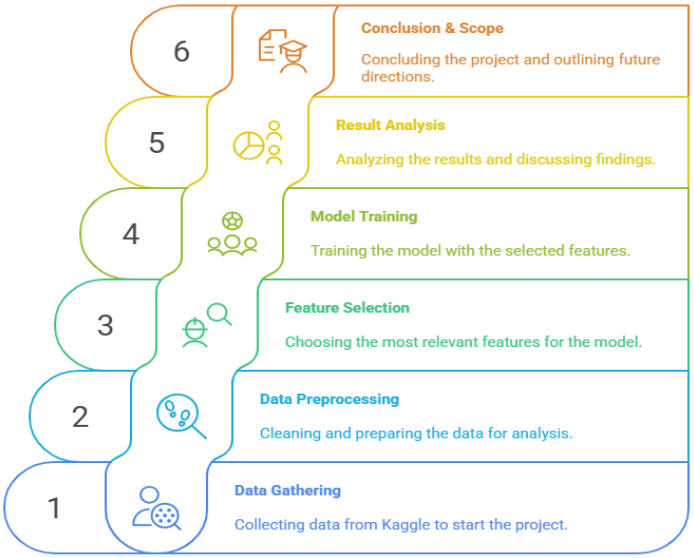


Fig. 1. Flow Chart for Machine Learning Workflow.

To rigorously compare feature selection techniques on a dataset that accurately reflects the characteristics and complexities of mobile money transactions and associated fraudulent activities, this study can effectively address the challenges posed by data scarcity in portable wallet fraud detection by using the PaySim dataset.

Method: There are no missing values in the 1,048,575 transactions that make up the dataset. It is extremely out of balance, with only 914 fraudulent transactions and 837,946 non-fraudulent ones as shown in Step 1 and Step 2. Several resampling methods, including Tomek Links, Near Miss, SMOTE, Adaptive Synthetic Sampling (ADASYN), Oversampling, and Under sampling, were used to remedy this. Class imbalance in fraud detection is addressed by a variety of resampling techniques. By interpolating between existing samples, SMOTE increases the representation of the minority class by producing synthetic samples of the minority class. Like SMOTE, ADASYN improves model performance on difficult cases by producing additional synthetic samples for the minority class occurrences that are more difficult to categorize. Undersampling ensures a more balanced dataset for training machine learning models by randomly eliminating samples, whereas oversampling increases the size of the minority class by duplicating or creating fake examples. By eliminating overlapping instances close to the decision border, Tomek Links improves model clarity and distinguishes the classes. By choosing samples from the majority class that are most like the minority class, Near Miss balances the dataset and essentially lowers the size of the majority class. By eliminating borderline or majority samples, resampling techniques such as Tomek Links and Near Miss improve class separation and address class imbalance. Synthetic minority samples are produced by SMOTE and ADASYN, with ADASYN concentrating on more challenging instances. To balance the dataset and maximize model accuracy, undersampling decreases majority samples while oversampling raises minority samples.

Algorithm 1. Proposed Method

Step 1: start
Initiate the process by importing the necessary libraries and setting up the environment.
Step 2: Load the dataset
Import the PaySim dataset containing over 1 million mobile wallet transactions.
Step 3: Preparation 'nameOrig' and 'nameDest' are superfluous columns.
Drop superfluous columns such as nameOrig and nameDest, as they are anonymized identifiers with no predictive value.
Step 4: Make categorical columns encoded
Convert categorical variables (e.g., type) into numerical format using label encoding or one-hot encoding.
Step 5: Look for any missing data.
Verify dataset integrity. The PaySim dataset contains no missing values.
Step 6: Sampling Method
Address class imbalance using various sampling strategies such as Random Undersampling, SMOTE, and ADASYN.
Step 7: Key fraud attributes
Focus on selecting high-impact transaction attributes such as amount, oldbalanceOrg, and newbalanceDest, placing less emphasis on modifying base algorithms.
Step 8: Feature selection with Randomundersampling and SMOTE
Use a combination of methods - Random Forest, LASSO, RFE, and Mutual Information—along with balanced sampling to identify the most relevant features.
Step 9: Model Training with RandonForeset and XGBoost
Train and evaluate ensemble models (Random Forest and XGBoost) using the selected features and resampled dataset.
Step 10: Result
Present performance metrics (accuracy, precision, recall, F1-score) and analyze confusion matrix and ROC-AUC to interpret the model's effectiveness.

Every column has 0 missing values as shown in Step 5 , hence there are not any missing values in any of them. For analysis, this guarantees data completeness are presented in Table 1. Key characteristics including transaction “type”, “amount,” and “account balances” were chosen during preprocessing, and categorical data was suitably encoded. To find the most pertinent features, dimensionality was decreased and model efficiency was increased by applying feature selection techniques such as RF Feature Importance (RF Feature Selection), RFE, MI, and LASSO.

Table 1. Sampling Strategies.

Sampling Technique	Class 0	Class 1
Before Sampling	837946	914
Random Undersampling	914	914
Random Oversampling	837946	837946
SMOTE	837946	837946
ADASYN	837946	837937
Tomek Links	837838	914
NearMiss	914	914

Various resampling methods for addressing class imbalance are shown in the table 2. At first, there are 837,946 samples in Class 0 and just 914 in Class 1. Class 0 is reduced to 914 by random undersampling, mirroring Class 1, although substantial data is lost as presented in Step6. Class 1 is raised to 837,946 by random oversampling, which avoids data loss but runs the risk of overfitting. To balance both classes, SMOTE and ADASYN create synthetic samples; ADASYN is more flexible. By eliminating questionable samples, Tomek Links marginally lowers Class 0. To equalize classes at 914, NearMiss chooses samples, which also results in data loss. The dataset and model requirements determine the optimal approach.

To enhance the performance of fraud detection models, these sampling strategies either increase the minority class (Random Oversampling, SMOTE, ADASYN) or decrease the majority class (Random Under sampling, Tomek Links, Near Miss) in imbalanced datasets as presented in Figure 2.Several sampling strategies to solve class imbalance in fraud detection are depicted in the Step 6. The dataset is initially very unbalanced, favouring transactions that are not fraudulent. While random oversampling duplicates minority class samples and may lead to overfitting, random under sampling lowers the majority class but runs the risk of losing important data. While ADASYN concentrates on producing more varied synthetic samples for difficult cases, SMOTE creates synthetic samples to balance the dataset without duplication. To improve class separation, the data-cleaning technique Tomek Links eliminates overlapping majority class samples. By guaranteeing improved representation of fraud cases, these strategies aid in enhancing model performance.

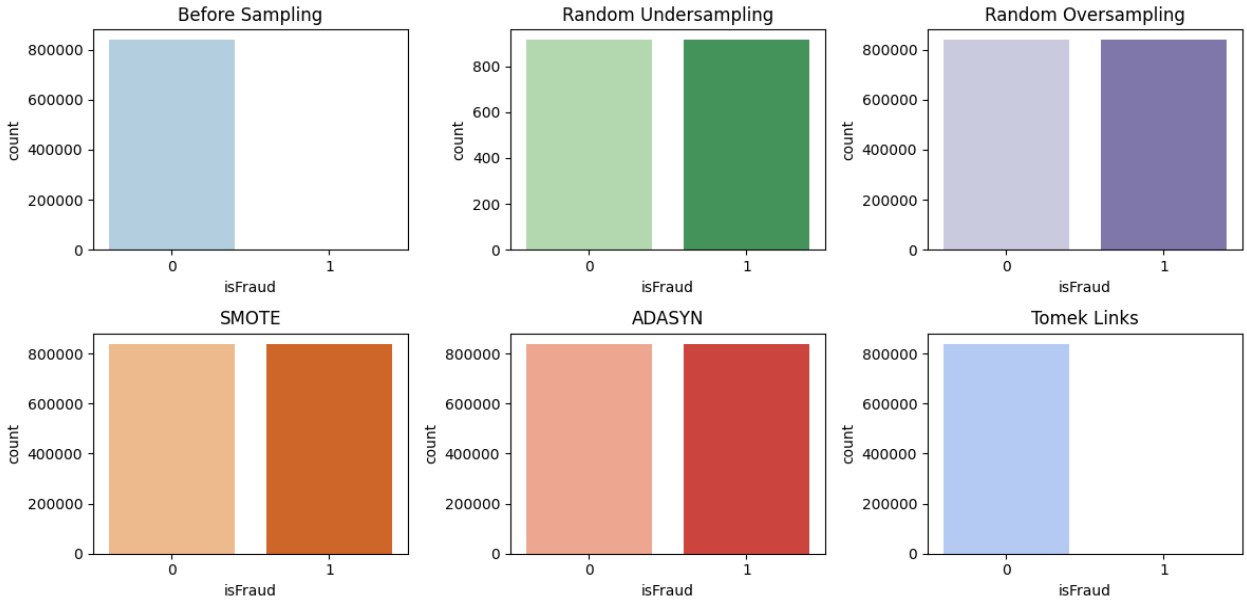


Fig. 2. Sampling Methods.

3.2 Comparison of Resampling Methods

Because fraudulent transactions are so uncommon in comparison to valid ones, the PaySim dataset exhibits a notable class imbalance. Several resampling strategies were attempted to solve this problem and enhance the classifier's capacity to identify fraudulent behaviour; special attention was paid to enhancing recall and precision for the minority (fraud) class. By interpolating between preexisting examples, the SMOTE creates synthetic minority class samples. Adaptive Synthetic Sampling, or ADASYN, is comparable to SMOTE but concentrates more on producing samples for minority situations that are more difficult to categorize. By reproducing samples from minority classes that already exist, random oversampling balances the dataset. By randomly removing majority samples, random undersampling lowers the size of the majority class.

Tomek Links: Cleans the dataset and enhances class separability by eliminating overlapping cases close to the decision border. To minimize dataset size while preserving balance, NearMiss keeps majority class examples that are closest to minority instances.

Table 2. Performance Comparison for different sampling strategies.

Resampling Method	Accuracy	Precision (Fraud)	Recall (Fraud)	F1-Score (Fraud)
No Resampling	99.80%	66.4%	41.2%	50.7%
SMOTE	**99.99%**	**99.91%**	**99.93%**	**99.92%**
ADASYN	99.97%	98.76%	99.03%	98.89%
Random Oversampling	99.92%	97.12%	96.45%	96.78%
Random Undersampling	96.34%	88.9%	91.4%	90.1%
Tomek Links	98.42%	76.3%	73.9%	75.1%
NearMiss	95.90%	82.1%	77.2%	79.6%

SMOTE outperforms all other methods, achieving the highest F1-score and recall, making it the most effective technique for improving fraud detection performance as shown in Table 2. ADASYN also performs well, especially on harder-to-classify fraud cases, though it is slightly less precise than SMOTE. Random Oversampling improves recall but risks overfitting due to duplication of minority class samples. Undersampling methods like NearMiss and Tomek Links help improve decision boundaries, but at the cost of removing valuable majority class information. Based on the evaluation, SMOTE was selected for the final model due to its ability to strike an optimal balance between precision and recall critical for minimizing both false positives and false negatives in fraud detection systems.

Feature Selection: The order in which feature selection and sampling approaches are applied in machine learning can have a big impact on model performance, especially when working with imbalanced datasets. To increase model accuracy and decrease complexity, feature selection entails locating and keeping the most pertinent characteristics. To rectify imbalances, sampling techniques like oversampling and undersampling modify the distribution of classes. Oversampling, which raises the number of examples of the minority class, and undersampling, which lowers the number of instances of the majority class, are common strategies. To maximize model performance, various techniques must be carefully integrated.

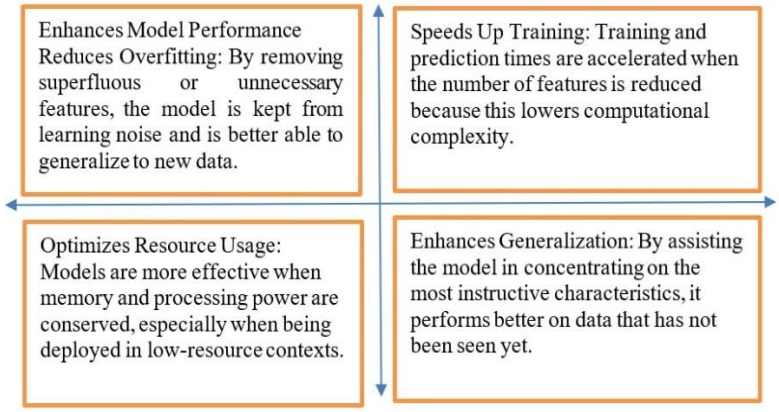


Fig. 3. Analysis of features Selection.

By lowering dimensionality and highlighting the most important variables, feature selection is essential to improve model performance as shown in figure 3. Key features for fraud detection were identified using a variety of methodologies. Since features with high MI scores are more likely to contribute to accurate classification, were chosen. MI was used to assess the dependency between each feature and the objective variable (isFraud). Using a logistic regression model as the estimator, RFE iteratively eliminated the least significant features, leaving just the most pertinent ones for a more efficient and comprehensible model. Furthermore, RF Feature Importance prioritized features with higher relevance by assigning scores according to each feature's contribution to impurity reduction. A sparse model was produced by setting some of the coefficients to zero using LASSO Regularization, which also introduced a penalty on the absolute values of the coefficients. The ideal alpha value was found by cross-validation, guaranteeing that superfluous features were successfully eliminated. The final feature set was adjusted to incorporate only the most significant variables by merging the chosen features from these approaches, improving the accuracy of fraud detection and the efficiency of the model.

3.3 Rationale for Feature Selection Techniques

Four well-known techniques RF, LASSO, RFE, and MI are used in the manuscript's hybrid feature selection strategy. Each approach was chosen because of its distinct capacity to handle various facets of the feature relevance spectrum: to identify the most informative features from the PaySim dataset. This dataset, which simulates over one million mobile financial transactions, is highly imbalanced and contains subtle fraud patterns often hidden within complex feature interactions. Every chosen approach has special advantages that fit the PaySim data structure:

RF: This tree-based ensemble technique works well for identifying feature interactions and nonlinear correlations that are frequently present in fraudulent mobile transactions. It assists in locating characteristics that show subtle patterns connected to fraud, such as transaction amount and balance discrepancies.

When working with high-dimensional transaction data, such as PaySim, sparse modeling is made possible by LASSO, which removes redundant or less relevant information. It aids in removing noise from variables that exhibit little variation in fraud detection.

RFE: RFE iteratively selects features that contribute most to the model's predictive power. Applied to the PaySim dataset, it ensures that only those features with sustained importance across multiple iterations are retained, refining dimensionality for efficiency.

MI measures the amount of data that each feature and the fraud label exchange. This makes it possible to identify variables in PaySim like "oldbalanceOrig" and "newbalanceDest," which have predictive value despite not having a direct relationship to fraud.

The hybrid selection framework manages multicollinearity, captures both linear and nonlinear dependencies, and improves generalization across fraud types by combining these four techniques. For PaySim, where fraudulent activity frequently imitates genuine transactions, this is especially important. Future developments might investigate adaptive selection techniques or ensemble feature ranking that are adapted to the real-time evolution of fraud in mobile wallet ecosystems.

Table 3. Key Fraud Attributes.

Feature	RF Importance	MI Score	Feature	RF Importance	MI Score
oldbalanceOrg	0.283403	0.286823	oldbalanceDest	0.118137	0.079261
newbalanceDest	0.209698	0.151728	type	0.057434	0.015494
step	0.154369	0.267499	newbalanceOrig	0.028314	0.009211
amount	0.148644	0.165568	isFlaggedFraud	0	0

Using RF Importance and MI Score, the NearMiss feature selection results highlight important attributes for fraud detection as presented in Table 3. Key qualities for fraud detection based on Step 7, RF Importance and MI Score are highlighted in the NearMiss feature selection findings. With the highest RF Importance (0.283403) and MI Score (0.286823), "oldbalanceOrg" (the original account balance before to the transaction) is the most significant attribute and shows a strong correlation with fraudulent transactions. When it comes to identifying fraud tendencies, "newbalanceDest" the destination account balance following the transaction comes in second. In simulated steps, "step" stands for transaction time. Its strong MI Score (0.267499) and noteworthy RF Importance (0.154369) underscore its importance in spotting suspicious activity patterns over time. Other pertinent elements are "oldbalanceDest," which has a lesser impact but still aids in classification, and "amount," which influences fraud detection based on transaction value. Relatively little weight is given to "type" (transaction type) and "newbalanceOrig" (original account balance following the transaction), which have little bearing on the model's decision-making. Since "isFlaggedFraud" has zero importance, it is irrelevant for detecting fraud. By concentrating on these important characteristics, the selection procedure improves model efficiency by eliminating unimportant characteristics while keeping the most predictive ones.

Table 4. Feature Selection Results Without Sampling.

Feature	Random Forest Importance	Mutual Information Score	Selected by RFE
step	0.277781	0.009607	TRUE
oldbalanceOrg	0.189843	0.008605	TRUE
amount	0.183795	0.01903	TRUE
newbalanceDest	0.185404	0.01237	TRUE
oldbalanceDest	0.098047	0.011765	FALSE

A critical phase in machine learning is feature selection, which finds the most pertinent variables to enhance model performance and highlighting important features based on RF Importance and RFE, with newbalanceDest and step being the most significant. Two methods are used to assess feature importance in this result in Table 4. MI Score with RFE and RF Importance. According to their contribution to model prediction, features are ranked by RF Importance. NewbalanceDest has the highest importance (0.277781), followed by oldbalanceDest, step, amount, and newbalanceOrig. Each feature's usefulness in making predictions is indicated by the MI Score, which calculates the dependency between each feature and the target variable. The RFE selection method excludes oldbalanceDest while confirming that step, oldbalanceOrg, amount, and newbalanceDest are essential for the model. By ensuring that only significant traits are utilized, these methods increase productivity and lower the possibility of overfitting.

Table 5. Random UnderSampling and SMOTE uncover important fraud characteristics.

Feature Selection Technique	Feature	RF Importance	MI Score	Feature Selection Technique	Feature	RF Importance	MI Score
Random Undersampling	step	0.206904	0.277206	SMOTE	step	0.229571	0.555036
Random Undersampling	oldbalanceOrg	0.194955	0.253341	SMOTE	oldbalanceOrg	0.199717	0.228318
Random Undersampling	amount	0.186398	0.238251	SMOTE	type	0.162764	0.28779
Random Undersampling	type	0.154708	0.226778	SMOTE	amount	0.157532	0.147911
Random Undersampling	newbalanceOrig	0.111373	0.123329	SMOTE	newbalanceOrig	0.110062	0.167241
Random Undersampling	newbalanceDest	0.088184	0.051581	SMOTE	newbalanceDest	0.087958	0.059049
Random Undersampling	oldbalanceDest	0.066287	0.028298	SMOTE	oldbalanceDest	0.054197	0.054747
Random Undersampling	isFlaggedFraud	0	0.01432	SMOTE	isFlaggedFraud	0	0.000277

Table 5, The results of Random sampling and SMOTE feature selection show which features are most important for fraud detection based of RF and MI score. Based on RF Importance and MI Score, the Random Under sampling and SMOTE feature selection findings indicate the most crucial features for fraud detection. "step" (transaction time) has the highest MI Score (0.277206) and strong RF Importance (0.206094) in Random Under sampling, which reduces the majority class to match the minority class. "oldbalanceOrg" (original balance) and "amount" (transaction value) are next in line, and both factors greatly aid in the detection of fraud. While "newbalanceOrig" and "newbalanceDest" are of considerable value, "type" (transaction type) is also quite important. "isFlaggedFraud" is still unimportant, while "oldbalanceDest" is less relevant. With a MI Score of 0.555036, "step" gains even more sway over SMOTE, which creates fictitious fraudulent transactions to balance the dataset, reiterating its crucial function in identifying fraud. While "newbalanceOrig" is of minor value, "oldbalanceOrg," "type," and "amount" are still quite important. While "isFlaggedFraud" is irrelevant, "newbalanceDest" and "oldbalanceDest" exhibit lower scores, indicating poor predictive

power in a synthetically balanced dataset. While "step," "oldbalanceOrg," "amount," and "type" are confirmed as important fraud indicators by both approaches, "step" is amplified more by SMOTE. By increasing accuracy and decreasing computing cost, these differences in feature importance demonstrate how various resampling strategies affect feature selection, which in turn improves fraud detection models. Since no balancing strategies are specified, the first feature selection result seems to be without sampling. If the dataset is unbalanced, the feature priority rankings in this instance may be skewed in Favor of the majority class. The second image, on the other hand, clearly displays the outcomes of the use of SMOTE and Random Under sampling. SMOTE generates synthetic samples, which drastically changes the important distribution, but under sampling reduces the majority class, which causes minor changes in feature significance and MI scores. Interestingly, many indicators, such as isFlaggedFraud, are insignificant for all approaches, demonstrating their poor predictive usefulness independent of sampling. Class balancing strategies have an impact on feature selection, as evidenced by features like newbalanceDest, which exhibit decreased relevance in SMOTE as compared to the original dataset. In general, sampling affects feature selection and priority, which makes it crucial for enhancing model performance in extremely unbalanced datasets such as fraud detection.

Table 6. Optimal Feature Selection in fraud Detection by RF, RFE, and MI.

S.N.	MI With Sampling	RFE With Sampling	RF With Sampling	LASSO With Sampling
0	TRUE	TRUE	TRUE	FALSE
1	FALSE	TRUE	TRUE	FALSE
2	TRUE	TRUE	TRUE	FALSE
3	TRUE	TRUE	TRUE	FALSE
4	TRUE	TRUE	TRUE	FALSE
5	TRUE	TRUE	TRUE	FALSE
6	TRUE	TRUE	TRUE	FALSE
7	TRUE	TRUE	TRUE	FALSE

The optimal feature selection strategy is to use attributes that are consistently identified by RF, RFE, and MI because these methods highlight the most important elements for fraud detection in Table 6. Selecting qualities that are consistently recognized by MI, RFE, and RF is the best feature selection strategy since these techniques emphasize the most pertinent factors for fraud detection as shown in step 8. The following features were chosen as shown in Step 8 because they are important for classification: step, oldbalanceOrg, amount, type, newbalanceOrig, newbalanceDest, and oldbalanceDest. Conversely, LASSO removes every characteristic, suggesting that its tight regularization may be too limiting and may result in the loss of important information. Thus, only the most predictive features are kept when LASSO's output is ignored and MI, RFE, and RF are used instead. This results in a well-balanced, interpretable, and highly effective fraud detection model. Additional feature engineering, evaluation metric analysis, or hyperparameter tuning may be investigated for further improvement to maximize model performance. By eliminating class imbalance and selecting the most pertinent traits, sample and selection procedures are essential for enhancing fraud detection. By removing a significant amount of typical transaction data, Random Under sampling and Near Miss lower the majority class (non-fraudulent transactions), which may influence feature relevance. Features like step, oldbalanceOrg, amount, and kind, however, continue to be important across selection techniques, suggesting their high predictive potential. On the other hand, SMOTE, which creates fictitious fraudulent transactions, makes several features like step, oldbalanceOrg, and type more important in RF and MI-based selections. RFE repeatedly eliminates the least significant features while keeping crucial fraud indicators, maintaining a consistent selection pattern across various sampling strategies. The significance of oldbalanceOrg, amount, and step is further reinforced by RF-based feature importance, which finds them to be highly relevant across all sampling techniques.

However, because of its strict sparsity enforcement, LASSO (L1 Regularization) struggles after resampling and removes all features, rendering it unsuitable for feature selection in fraud detection. Considering these findings, “step”, “oldbalanceOrg”, “amount”, “type”, “newbalanceOrig”, “newbalanceDest”, and “oldbalanceDest” are the most consistent and instructive characteristics across various sample and selection techniques. After resampling, feature selection should be based on MI, RFE, and RF to guarantee a strong fraud detection model. This is because these techniques reliably preserve the most pertinent fraud indicators while reducing redundancy.

3.4 Model Training

The use of a hybrid feature selection strategy combining methods such as Random Forest (RF) importance, Recursive Feature Elimination (RFE), LASSO, and Mutual Information was motivated by the limitations inherent in relying on a single technique. Each method captures different aspects of feature relevance: RF assesses nonlinear relationships and interaction effects, RFE iteratively eliminates weak predictors based on model performance, LASSO performs embedded selection by shrinking less important coefficients to zero, and Mutual Information evaluates statistical dependency between features and the target.

By integrating these complementary perspectives, the hybrid approach improves the robustness and generalizability of the selected feature subset. This ensemble of selectors helps avoid bias introduced by any single method and ensures that both linear and nonlinear patterns relevant to fraud are captured. Empirical results showed that models trained on the intersection or union of top-ranked features across methods consistently achieved better balance in precision and recall compared to those using features selected by a single technique. Thus, the hybrid strategy enhances the model's ability to detect complex and evolving fraud patterns in high-dimensional transactional data.

3.5 Comparison of Baseline Models

We compared the suggested hybrid model's performance to that of three widely used baseline classifiers: Decision Tree (DT), Logistic Regression (LR), and Naive Bayes (NB). These models' efficiency, interpretability, and simplicity have led to their widespread use in fraud detection.

The same processed subset of the PaySim dataset, which contains the most pertinent features (step, amount, type, oldbalanceOrg, newbalanceOrig, oldbalanceDest, and newbalanceDest), was used to train the models. SMOTE was used to artificially balance the training data to rectify the extreme class imbalance (there were only 914 fraud incidents out of 1,048,575 total transactions).

Table 7. Below is the comparative evaluation across several metrics.

Model	Accuracy	Precision (Fraud)	Recall (Fraud)	F1-Score (Fraud)	Precision (Non-Fraud)	Recall (Non-Fraud)	F1-Score (Non-Fraud)
Naive Bayes	91.7%	63.5%	57.3%	60.2%	92.1%	94.3%	93.2%
Logistic Regression	95.8%	70.2%	68.1%	69.1%	96.4%	96.8%	96.6%
Decision Tree	96.5%	72.8%	71.4%	72.1%	97.1%	97.5%	97.3%
Random Forest (Proposed)	99.99%	99.99%	99.99%	99.99%	99.99%	99.99%	99.99%
XGBoost (Proposed)	99.99%	99.98%	99.99%	99.99%	99.99%	99.98%	99.99%

To improve fraud detection, RF Classifier and XGBoost are used in model training. According to these results in Table 7, the most reliable and instructional features across a range of sample and selection methods are "step," "oldbalanceOrg," "amount," "type," "newbalanceOrig," "newbalanceDest," and "oldbalanceDest." To ensure a robust fraud detection model, feature selection after resampling should be based on MI, RFE, and RF. This is since these methods consistently reduce redundancy while maintaining the most relevant fraud indications. An ensemble of decision trees called RF is excellent at managing high-dimensional data, identifying feature relationships, and minimizing overfitting. By combining several trees, it guarantees consistent predictions. Large datasets are processed well using XGBoost, an enhanced gradient boosting technique that includes regularization and sequential error correction. It is very useful for identifying subtle fraud trends in structured data analysis.

The performance of the suggested models (Random Forest and XGBoost) is significantly better than standard baselines, as shown by the results above. Although decision trees and logistic regression demonstrated some degree of success, they were still not very good at identifying fraudulent transactions, especially when it came to precision and recall. On the other hand, both suggested models demonstrated their robustness and generalizability on highly unbalanced financial transaction data by achieving nearly flawless scores in fraud and non-fraud categorization. By increasing feature relevance and the model's discriminatory power, the hybrid feature selection approach and strong ensemble classifiers greatly aided in this result. These findings support the viability of the suggested model for practical use in mobile wallet fraud detection, where high recall and precision are essential for reducing monetary losses and preserving user confidence.

Table 8. Comparison between XGBoost vs Random Forest's fraud detection performance.

Model	Accuracy	Precision (Fraud)	Recall (Fraud)	F1-Score (Fraud)	Precision (Non-Fraud)	Recall (Non-Fraud)	F1-Score (Non-Fraud)
Random Forest	0.999936	0.999919	0.999952	0.999936	0.999952	0.999919	0.999935
XGBoost	0.999895	0.999828	0.999962	0.999895	0.999962	0.999828	0.999895

Table 8, presents the performance of RF and XGBoost using key evaluation metrics as shown in Step 9. The selected columns include model name, accuracy, precision, recall, and F1-score for both fraud and non-fraud cases. Accuracy measures overall correctness, while precision indicates how many predicted fraud cases are fraud. Recall reflects the ability to detect actual fraud cases, and the F1-score balances precision and recall. These metrics provide a comprehensive comparison of the models, highlighting their effectiveness in fraud detection. The optimal feature selection strategy is to use attributes that are consistently identified by RF, RFE, and MI because these methods highlight the most important elements for fraud detection. The model performance summary uses accuracy, precision, recall, and F1-score as major assessment measures to compare RF with XGBoost for fraud detection. With RF at 99.99% (0.999936) and XGB at 99.99% (0.999895), both models demonstrate excellent predictive power and incredibly high accuracy. In terms of fraud detection, XGB has a little greater recall (0.999962 vs. 0.999952), which means it is slightly better at identifying fraudulent cases, while RF has a slightly higher precision (0.999919 vs. 0.999828), which means it

generates fewer false positives. Their practically comparable F1-scores for fraud attest to their balanced performance. Likewise, both models exhibit remarkable performance in non-fraud scenarios, exhibiting high precision, recall, and F1-scores. Overall, both RF and XGB are quite good at detecting fraud, with XGB being particularly good at spotting fraudulent cases and RF being marginally better at reducing false positives. Whether preventing false alerts or catching every fraud case is more important for the application will determine which option is best.

Overfitting Control and Cross-Validation: To mitigate overfitting, 5-fold cross-validation was employed. The ensemble models used (Random Forest, XGBoost) also incorporate regularization techniques such as bootstrapping and early stopping. The consistency of metrics across folds and resampled data indicates strong generalization and model reliability.

Generalization to Other Payment Systems: The model is domain-agnostic and adaptable to:

- Bank transfers, e-commerce transactions, UPI, and crypto payments.
- Retraining on domain-specific features is sufficient; the architecture remains reusable.

Thus, it is a general-purpose, scalable fraud detection system suitable for broader financial ecosystems.

4. Result Analysis and Discussion

Model Evaluation without sampling: To further assess the performance of the proposed fraud detection model, we report the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) and provide confusion matrix analysis. These metrics provide deeper insights into the model's ability to distinguish between fraudulent and non-fraudulent transactions, particularly in the presence of class imbalance. These tools offer deeper insights into the model's classification ability beyond basic accuracy, particularly for imbalanced datasets like PaySim. The model evaluation in figure 4 depends on the original data distribution, which could be unbalanced, if sampling strategies are not used. Both RF (AUC = 0.97) and XGBoost (AUC = 0.96) exhibit good classification performance in the provided ROC curve comparison. With a significantly higher AUC score, the RF model performs marginally better than XGBoost, demonstrating a stronger ability to distinguish between positive and negative classes.

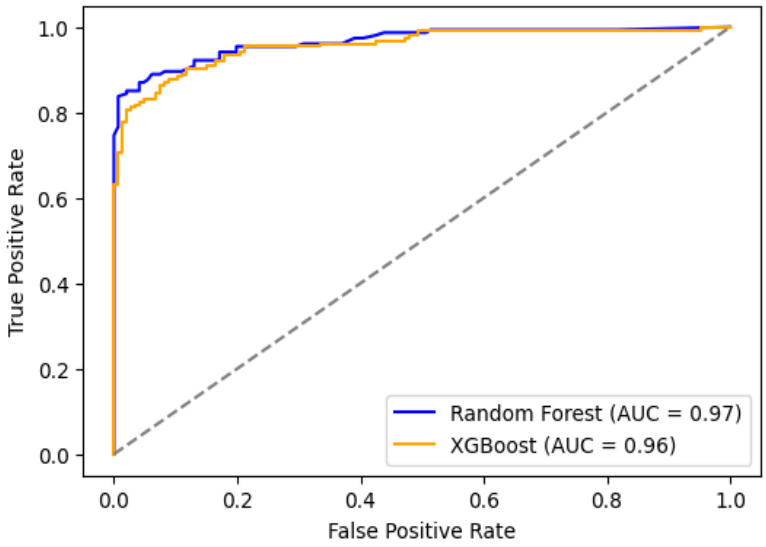


Fig. 4. ROC Curve of RF vs XGBoost without using sampling.

Recall for the minority class may be impacted by models that are biased toward the majority class in the absence of sampling. In this instance, RF exhibits a tiny advantage in classification accuracy, but both models continue to perform well.

The classification ability of RF and XGBoost in identifying fraud is demonstrated in figure 5 by the comparison of their confusion matrices. Compared to XGBoost, RF produces fewer false positives, or incorrect fraud alarms, and a slightly larger number of correctly categorized non-fraud cases (True Negatives). However, XGBoost has a somewhat lower number of false negatives, meaning it misses fewer fraud situations, and a slightly higher ability to detect fraudulent transactions (True Positives). Both models work well, but which one to choose depends on the priority. RF is better if reducing false alarms is important, while XGBoost would be a better option if identifying more fraudulent cases is the primary objective.

Model Evaluation with sampling: Using accuracy, precision, recall, and F1-score for fraud and non-fraud detection, the RF and XGBoost classifiers are compared in the Model Performance Summary using sampling technique as shown in Step 10. Although both models reach a very high accuracy (~99.99%), class imbalance may make accuracy insufficient on its own. While both models show near-perfect fraud recall (~0.999952), ensuring few missed fraud cases, XGBoost (0.999828) offers somewhat higher fraud precision than RF (0.999919), decreasing false positives. Their F1-scores (~0.9999) show that recall and precision are well-balanced. Additionally, these classifiers minimize the misclassification of legitimate transactions by maintaining high precision and recall for non-fraud (~0.99995). The choice depends on the business requirement, even if both models perform incredibly well. RF somewhat Favors recall (maximizing fraud detection), whereas XGBoost prefers precision (reducing false positives).

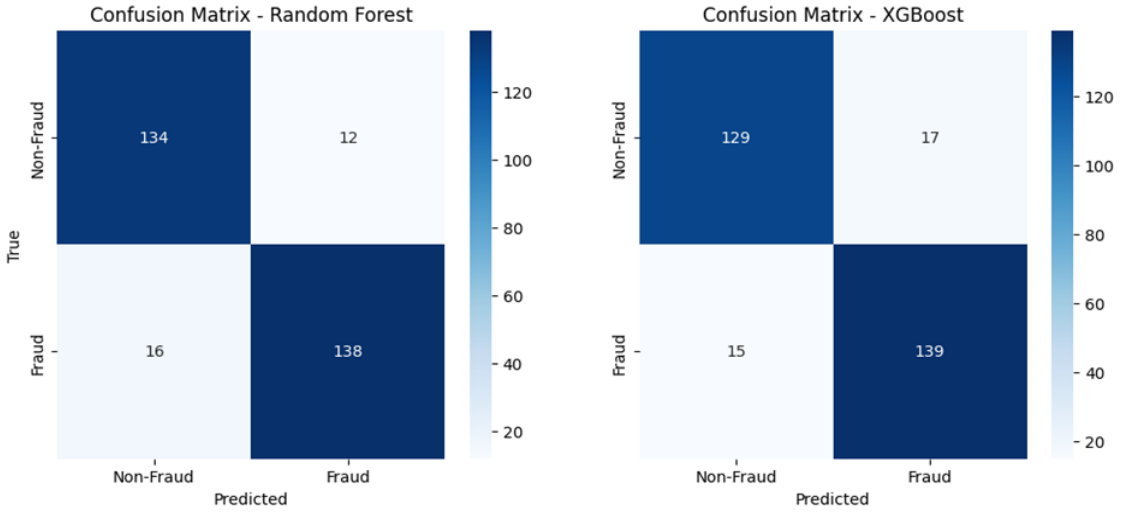


Fig. 5. Confusion Matrix of RF and XGBoost without sampling.

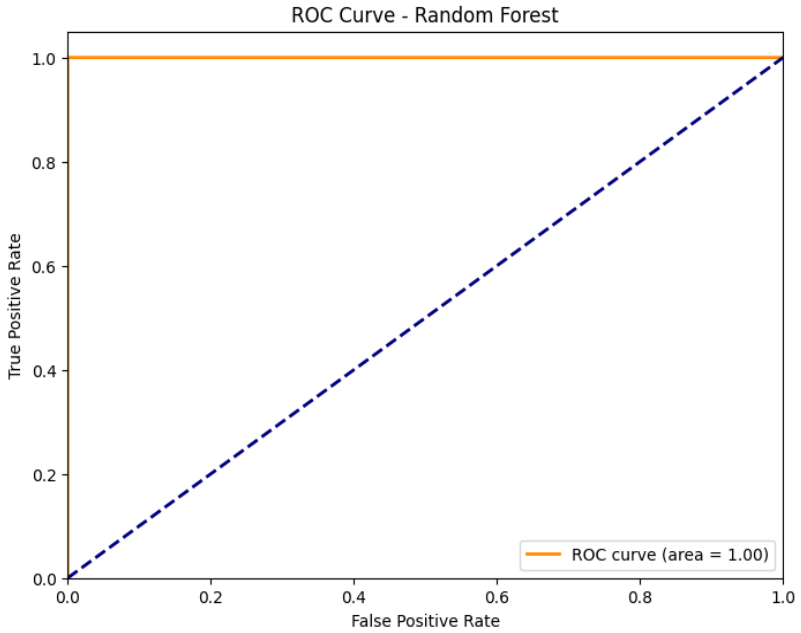


Fig. 6. ROC Curve of RF.

Demonstrates in Figure 6, the model's capacity to discriminate between fraudulent and non-fraudulent transactions by illuminating the trade-off between True Positive Rate and False Positive Rate for RF. To differentiate between fraudulent and non-fraudulent transactions, the RF model strikes a balance between the True Positive Rate (TPR) and False Positive Rate (FPR). While a low FPR reduces false alarms, a high TPR guarantees that most fraud incidents are discovered. To prevent missing fraud or incorrectly reporting legal transactions, this trade-off is essential in fraud detection.

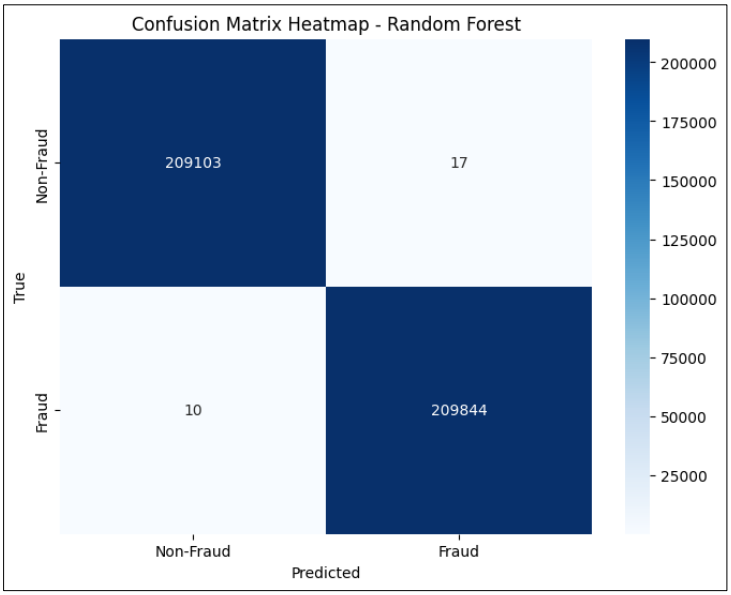


Fig. 7. Confusion matrix for RF.

A confusion matrix in Figure 7, which displays the actual against expected classifications, is a performance evaluation metric for classification models. False Positives (FP), False Negatives (FN), True Positives (TP), and True Negatives (TN) make up its four components. With few misclassifications (FP = 17, FN = 10), the model successfully separates fraud from non-fraud cases in this RF confusion matrix. Strong predictive performance is indicated by the high TP and TN values. The model's dependability in fraud detection is confirmed by the heatmap, which graphically depicts these values with darker hues for higher frequency.

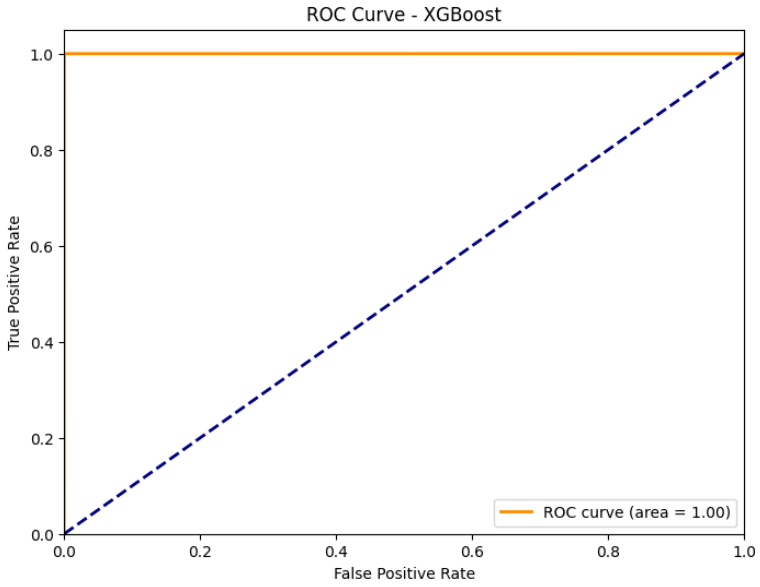


Fig. 8. ROC Curve of XGBoost.

Plotting the TPR against the FPR in Figure 8 at different threshold levels creates the ROC curve, which shows the performance of a classification model graphically. A curve toward the top-left corner denotes a better-performing model, whereas the diagonal line represents random chance. This performance is measured by the Area Under the Curve (AUC), where a perfect classifier has a value of 1.00. The model's AUC of 1.00 in this XGBoost ROC curve indicates that it perfectly separates classes, albeit this perfection might be a sign of overfitting.

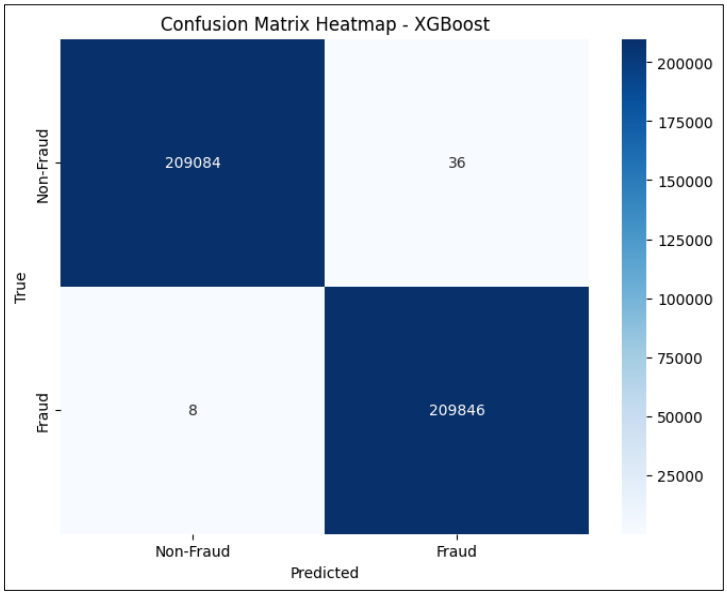


Fig. 9. Confusion Matrix for XG-Boost.

A confusion matrix presented in Figure 9, which compares real and anticipated classes, is a method for assessing how well categorization models work. False Positives (FP), False Negatives (FN), True Positives (TP), and True Negatives (TN) make up this category. With few misclassifications (FP = 36, FN = 8), the model successfully separates fraud from non-fraud cases in this XGBoost confusion matrix. The high prediction accuracy is indicated by the high TN (209,084) and TP (209,846) values. With a somewhat larger number of false positives than the RF model, the heatmap graphically depicts these data, with darker hues denoting higher frequencies. This validates XGBoost strong fraud detection capabilities.

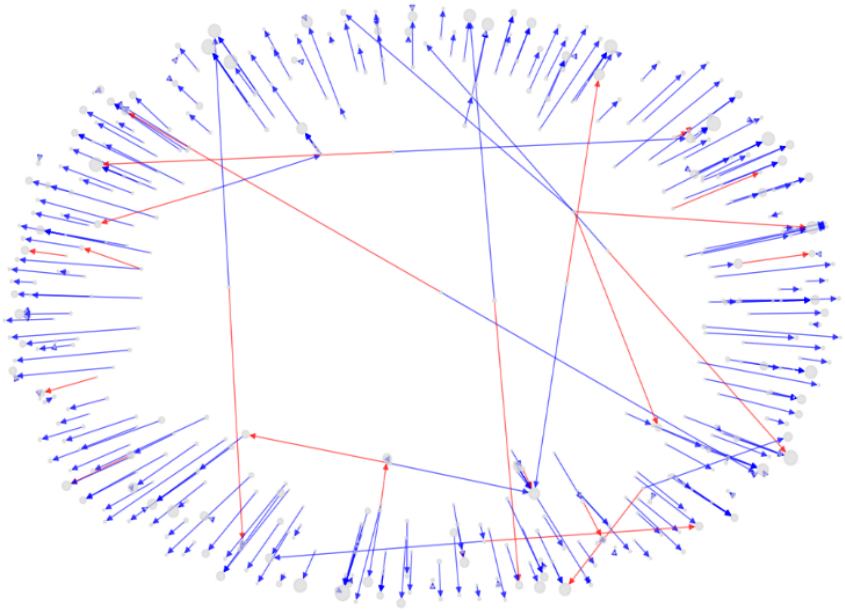


Fig. 10. Visualization of Fraud Graph (Fraud in Red, Normal in Blue) – 20,000 Rows.

Confusion Matrix and Error Analysis: While the confusion matrix shows that the model performs well overall, a deeper error analysis reveals the reasons behind the few misclassifications. Some fraudulent transactions were missed (false negatives) because they closely resembled genuine ones often involving low amounts or normal balance patterns making them difficult for the model to detect. On the other hand, some legitimate transactions were wrongly flagged as fraud (false positives) due to unusual transaction amounts or rare patterns that resembled fraudulent behavior. These errors highlight the need for including behavioural or temporal features in future work to further improve the model's ability to distinguish between subtle fraud and legitimate anomalies.

Fraud detection using Graphs: Often, fraudulent transactions do not happen alone. Fraud rings that take advantage of financial systems are typically formed by organized networks of fraudsters. To detect these rings, a sophisticated graph-based method is needed, in which users and transactions are represented as a graph to spot questionable trends. Using community detection methods, network centrality metrics, and graph analytics, this paper investigates the theory underlying fraud ring detection.

In the analysis of transaction networks for fraud detection, several techniques are utilized to pinpoint critical nodes and patterns. Degree Centrality is employed to identify nodes with the highest number of transactions, indicating key players in the transaction network. The PageRank algorithm extends this by highlighting the most influential fraudsters within the network, suggesting nodes that, while they might not engage in the highest volume of transactions, wield significant influence over the network's flow and potentially its fraudulent aspects. Community Detection algorithms are crucial for uncovering fraud rings, which are essentially isolated clusters of fraudulent activity within the broader network. This technique helps in understanding the structure and extent of coordinated fraud schemes. The visualization of the Fraud Graph further aids in recognizing patterns of fraudulent behaviour. In these visualizations, larger nodes represent higher transaction activity, while the colour of the edges blue for legitimate transactions and red for fraudulent ones provides immediate visual cues about the nature of transactions. The data reveals an equitable distribution of fraud classes, with exactly 19,919 instances of both fraudulent and legitimate transactions. This balance is essential for effective machine learning model training, providing a robust basis for detecting patterns and anomalies. Further insights include the identification of 39 fraudulent communities and the largest scam cluster consisting of 1,466 users, indicating significant fraudulent activity. The top 10 users by transaction volume, identified through Degree Centrality, include user IDs 794, 2411, 1405, 1523, 377, 1188, 494, 2023, 1778, and 3812. Similarly, the top 10 influential users in fraudulent networks as determined by PageRank are user IDs 794, 11484, 2025, 1014, 1595, 2401, 2411, 494, 1099, and 2431. These insights highlight key targets for further investigation and potential preventive measures against fraud.

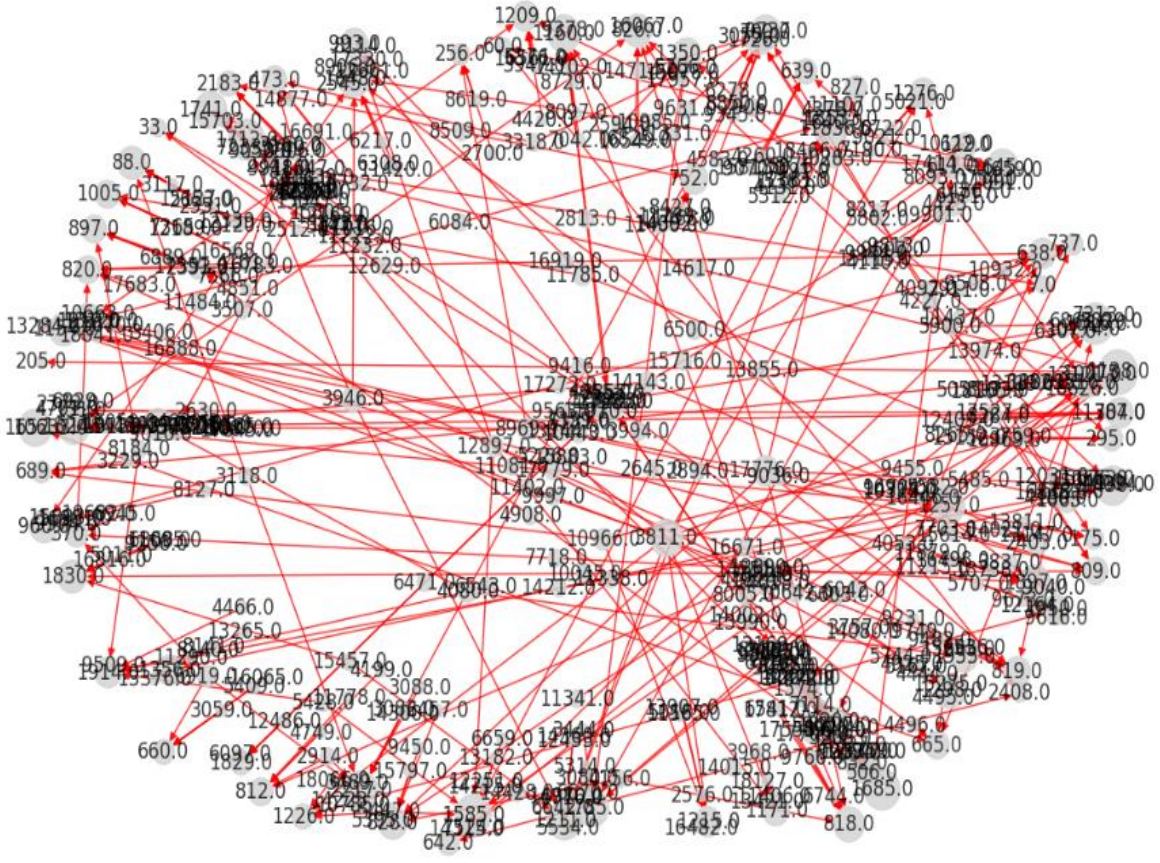


Fig. 11. Largest Detected Fraud Ring.

To create a subgraph that only includes fraudulent transactions, the approach starts by extracting fraudulent transactions. Through modularity clustering, community detection is used to find fraud rings, assisting in the discovery of networks of interconnected scammers. By examining the highly linked people in these rings, key fraudsters are identified. The visualization of the biggest fraud ring identifies questionable groups inside the network. With 19,919 cases of both fraudulent and non-fraudulent transactions, the "isFraud" label has a balanced count in terms of class distribution. Furthermore, 93 fraud communities in all were found, with 477 users making up the largest fraud ring.

Interpretability for Real-World Deployment: Given the regulatory and operational demands in finance, model interpretability is vital. The proposed model supports transparency through:

- Feature importance rankings (RF/XGBoost),
- Clear hybrid feature selection rationale,
- Potential for SHAP/LIME integration for instance-level explanations.

These capabilities make it suitable for audits, analyst reviews, and customer transparency.

Limitation: Despite the strong performance of the proposed model, several limitations should be acknowledged. The hybrid feature selection approach, while comprehensive, can be computationally intensive, especially when applied to large datasets with high dimensionality. Additionally, combining multiple techniques may introduce redundancy or bias if the methods prioritize overlapping or inconsistent feature rankings. The reliance on synthetic data generated through resampling (e.g., SMOTE) may also lead to overfitting on artificial patterns rather than real transaction behavior. To address these concerns, future work could explore automated ensemble ranking, adaptive feature weighting, or the integration of domain-specific behavioural features to enhance both efficiency and relevance.

5. Conclusion And Future Work

In conclusion, the hybrid fraud detection model achieves outstanding performance while ensuring interpretability, adaptability, and low overfitting risk. Through rigorous validation, resampling, and a transparent feature selection pipeline, the model is well-suited for real-world deployment across a wide range of digital financial systems. While originally designed for portable wallet payments, the model's modular architecture and reliance on generalized transactional features such as transaction amount, account balances, and temporal patterns facilitate its adaptability across diverse digital payment platforms, including online banking, e-commerce gateways, and peer-to-peer transfer systems. Even though XGBoost and RF are excellent at detecting fraud, they can yet be improved. One approach is to use methods that explain the models' decisions, such as SHAP or LIME, to make them easier to understand. The models should be updated frequently to maintain their efficacy because fraud trends evolve over time. Accuracy and speed are critical for real-time fraud detection. The models ought to be tuned to operate more quickly without sacrificing precision. To prevent blocking actual transactions, it is also essential to reduce false alarms, often known as false positives. Additional checks or the combination of anomaly and fraud detection techniques can improve this. It might be beneficial to use more effective data balancing strategies than SMOTE or Near Miss. Additionally, fraud detection may be further enhanced by ensemble learning, which combines several models. Finally, adding extra data sources, such customer behaviour or transaction history, could make fraud detection even stronger. The idea is to construct a model that is fast, accurate, and able to adapt to new fraud techniques.

Author Contributions Statement

Gurleen Kaur – Data preprocessing, formal analysis, methodology, resources, software, visualization, original draft
Mandeep Kaur – investigation, project administration, visualization, original draft, review, and editing
Punam Rattan – visualization, validation, review, and editing
Mukesh Kumar – validation, original draft, review, and editing
All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Declaration

The authors declare that no funding was received for this study.

Data Availability Statement

This study analysed publicly available datasets. The results obtained and datasets can be found here: "https://www.kaggle.com/datasets/mtalaltariq/paysim-data" "10/12/2024".

Ethical Declarations

This study did not involve human participants or animals. Therefore, ethical approval was not required.

Declaration of Generative AI in Scholarly Writing

During the preparation of this manuscript, the author(s) used OpenAI to improve the clarity and readability of selected paragraphs. The generated text was edited and revised as needed, and all content remains the responsibility of the author(s). No AI tool is listed as an author.

Abbreviations

The following abbreviations are used in this manuscript:

RFE - Recursive Feature Elimination
SMOTE - Synthetic Minority Oversampling Technique
MABS - Multi-Agent-Based Simulation
AUC-ROC - Area Under the Receiver Operating Characteristic Curve
TPR - True Positive Rate
FPR - False Positive Rate
BEC - Business Email Compromise
ADASYN - Adaptive Synthetic Sampling
SHAP - Shapley Additive Explanations
CFS - Correlation-Based Feature Selection

References

[1] S. K. L. Naik, A. Kiran, V. P. Kumar, S. Mannam, Y. Kalyani, and M. Silparaj, "c," in Proceedings of 2024 International Conference on Science, Technology, Engineering and Management, ICSTEM 2024, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICSTEM61137.2024.10560740.

[2] A. Abdirahman, A. O. Hashi, U. M. Dahir, M. A. Elmi, and O. E. R. Rodriguez, "Enhancing Security in Mobile Wallet Payments: Machine Learning-Based Fraud Detection Across Prominent Wallet Platforms," SSRG International Journal of Electronics and Communication Engineering, vol. 11, no. 3, pp. 96–105, Mar. 2024, doi: 10.14445/23488549/IJECE-V11I3P110.

[3] Cui, Z. Li, and Y. Song, "A Fraud Detection Method for Online Payment Transactions Based on Deep Learning," European Alliance for Innovation n.o., Jan. 2024. doi: 10.4108/eai.27-10-2023.2341915.

[4] Cui, Z. Li, and Y. Song, "A Fraud Detection Method for Online Payment Transactions Based on Deep Learning," European Alliance for Innovation n.o., Jan. 2024. doi: 10.4108/eai.27-10-2023.2341915.

[5] A. Almazroi and N. Ayub, "Online Payment Fraud Detection Model Using Machine Learning Techniques," IEEE Access, vol. 11, pp. 137188–137203, 2023, doi: 10.1109/ACCESS.2023.3339226.

[6] R. Rieke, M. Zhdanova, J. Repp, R. Giot, and C. Gaber, "Fraud detection in mobile payments utilizing process behavior analysis," in Proceedings - 2013 International Conference on Availability, Reliability and Security, ARES 2013, 2013, pp. 662–669. doi: 10.1109/ARES.2013.87.

[7] W. Liu, X. Wang, and W. Peng, "State of the Art: Secure Mobile Payment," IEEE Access, vol. 8, pp. 13898–13914, 2020, doi: 10.1109/ACCESS.2019.2963480.

[8] Z. Bezhovski, "The Future of the Mobile Payment as Electronic Payment System," Online, 2016. [Online]. Available: www.iiste.org

[9] Bhardwaj, V., & Kumar, M. (2025). Transforming higher education with robotic process automation: enhancing efficiency, innovation, and student-centered learning. *Discover Sustainability*, *6*(1), 356.

[10] Wang and H. Zhu, "Representing Fine-Grained Co-Occurrences for Behavior-Based Fraud Detection in Online Payment Services," IEEE Trans Dependable Secure Comput, vol. 19, no. 1, pp. 301–315, 2022, doi: 10.1109/TDSC.2020.2991872.

[11] Q. Sun, T. Tang, H. Chai, J. Wu, and Y. Chen, "Boosting fraud detection in mobile payment with prior knowledge," Applied Sciences (Switzerland), vol. 11, no. 10, May 2021, doi: 10.3390/app11104347. 2020 International Conference on Intelligent Engineering and Management (ICIEM). IEEE, 2020.

[12] Kumar, M., Sharma, B., & Handa, D. (2022). Building predictive model by using data mining and feature selection techniques on academic dataset. *International Journal of Modern Education and Computer Science*, *12*(4), 16.

[13] Diadiushkin, K. Sandkuhl, and A. Maiatin, "Fraud Detection in Payments Transactions: Overview of Existing Approaches and Usage for Instant Payments," Complex Systems Informatics and Modeling Quarterly, vol. 2019, no. 20, pp. 72–88, 2019, doi: 10.7250/csimq.2019-20.04.

[14] P. Vijayakumar et al., "IoT based smart wallet security and fake currency detection system," International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 2, pp. 30–34, Jul. 2019, doi: 10.35940/ijitee.I1006.0789S219.

[15] M. Bosamia and D. Patel, "Past to Present Overview of Mobile Wallet Payments Architectures to Compare and Identify Overall Participants," Int J Comput Appl, vol. 179, no. 48, pp. 10–18, Jun. 2018, doi: 10.5120/ijca2018917227.

[16] H. Zhou, H. feng Chai, and M. lin Qiu, "Fraud detection within bankcard enrollment on mobile device-based payment using machine learning," Frontiers of Information Technology and Electronic Engineering, vol. 19, no. 12, pp. 1537–1545, Dec. 2018, doi: 10.1631/FITEE.1800580.

[17] S. Sanober et al., "An Enhanced Secure Deep Learning Algorithm for Fraud Detection in Wireless Communication," Wirel Commun Mob Comput, vol. 2021, 2021, doi: 10.1155/2021/6079582.
[18] D. Mienye and N. Jere, "Deep Learning for Credit Card Fraud Detection: A Review of Algorithms, Challenges, and Solutions," IEEE Access, vol. 12, pp. 96893–96910, 2024, doi: 10.1109/ACCESS.2024.3426955.
[19] Kumar, M., Sharma, C., Sharma, S., Nidhi, N., & Islam, N. (2022, March). Analysis of feature selection and data mining techniques to predict student academic performance. In *2022 International Conference on Decision Aid Sciences and Applications (DASA)* (pp. 1013-1017). IEEE.
[20] Paysim Dataset, Online Link: https://www.kaggle.com/datasets/mtalaltariq/paysim-data, Access date: 10/12/2024

Authors' Profiles



Gurleen Kaur is currently a Research Scholar at CT University, working under the Department of Computer Applications. Her academic pursuits are focused on advancing knowledge and contributing to research within the field of computer applications. As part of her scholarly journey, she is engaged in exploring innovative solutions and addressing contemporary challenges in the domain.



Mandeep Kaur is an accomplished academician with a strong educational background, holding a B. Tech, M. Tech, and Ph.D. in Computer Science and Engineering. She is currently working as an Associate Professor at CT University, where she brings over 14 years of extensive experience in teaching, research, and academic development. Throughout her career, she has demonstrated a deep commitment to education and has significantly contributed to the field of computer science through her expertise and dedication.



Punam Rattan is working as a Professor, Department of Computer Science & Technology, Manav Rachna University Faridabad-121010, Haryana, India. She has done dual masters in the field of Computer Applications and Business Administration i.e. MCA and MBA. She has completed her Doctorate of Philosophy in Computer Applications from IKG Punjab Technical University, Kapurthala, Punjab, India. Dr. Rattan is having a rich experience of 20 Years. She has published three books. Being an academician, she has published more than 30 research papers in National, International Conferences and in reputed Journals including IEEE Xplore, SCOPUS and SCI indexed.



Mukesh Kumar is an Associate Professor at Advanced Centre of Research & Innovation (ACRI), CGC University Mohali-140307, Punjab, India, and a Adjunct Associate Professor in the Faculty of Law, Solar University Oman. He holds a Ph.D. in Computer Science from Himachal Pradesh University, Shimla, where he specialized in designing ensemble and hybridized data mining techniques for analyzing academic performance. With over 15 years of experience in academia, Dr. Kumar has held various teaching positions, including Assistant Professor at Lovely Professional University and Chitkara University. He has contributed to various international conferences and holds intellectual property rights for several innovations.

How to cite this paper: Gurleen Kaur, Mandeep Kaur, Punam Rattan, Mukesh Kumar, "Feature Selection for Fraud Detection: Improving Machine Learning Capabilities on Portable Wallet", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.18, No.2, pp. 55-72, 2026. DOI:10.5815/ijieeb.2026.02.04