Digital payment fraud: Detection and Prediction

Geeta Aggarwal[1] and Seema Aggarwal[2*]

ABSTRACT

The widespread implementation of digital payment methods has greatly simplified financial transactions but has also increased the potential for online payment fraud. Creating reliable prediction models for fraud detection is crucial for protecting against fraudulent actions. This research employs machine learning techniques, notably Random Forest or Logistic Regression, to differentiate between genuine and fraudulent financial dealings reliably. Researchers assess these models' efficacy in real-time fraud detection using a comprehensive dataset, including transaction information and labelled fraud incidents. The findings of this study will help strengthen the safety and trustworthiness of online payment systems, reducing the risk of fraud and other security breaches for consumers.

Keywords : *Cybercrime, Fraud, Digital Payment.*

1. Introduction

Digital payment has become increasingly popular in recent times because of the convenience it offers. This has been made possible due to the advancement in technology. Technology is a boon as well as a bane and so is digital payment. Digital payment offers various benefits like convenience, speed, accessibility, contactless payment and many cash back offers to lure. On the other hand, one must be cautious about security and privacy. There is a tradeoff between benefits and risks of digital payments which consumer must weigh before proceeding further.

This paper aims at creating prediction models to help detect fraud in digital payments. Machine learning techniques have been used for this purpose. Random forest and linear regression are two techniques used in this paper to detect fraud in digital payments. The models developed have been assessed in real time using a comprehensive dataset.

The paper is organized as follows: the problem is defined in section 2. Section 3 discusses the related work. The methodology used is discussed in detail in section 4. The paper is concluded with recommendations in section 5.

2. Problem Definition

Online payment fraud detection is a significant challenge in the banking sector. The potential for fraudulent operations targeted at exploiting weaknesses in online payment systems rises in tandem with the rising use of digital transactions (Nicholls, Kuppa, & Le-Khac, 2019). Financial damages from fraudulent transactions may be substantial for people and businesses, making detecting and preventing such acts crucial.

The project's overarching objective is to construct reliable prediction models that can promptly and reliably spot illicit financial dealings as they occur. Each transaction's attributes will be used to predict whether it is fraudulent, and the dataset will be used to train and evaluate these models. Random Forest, Logistic Regression, and other machine-learning methods may solve this issue. The models will be trained on the transaction history to recognize abnormalities and trends that indicate fraudulent activity (Punithalakshmi & Rajakumar, 2021). Financial institutions and payment service providers may include these models in their systems to increase security measures and protect against possible fraud concerns, all while attaining high accuracy in fraud detection.

Several criteria, such as the characteristics used, the algorithm's efficiency, and the dataset's completeness and accuracy, will determine the effectiveness of the prediction models. Users' funds, digital payment system confidence, and overall, online financial security might benefit from better fraud detection algorithms.

3. Related Work

Detecting and avoiding cyber fraud are only two of the many data analysis and processing issues that may be tackled with the use of machine learning (ML). ML may be categorized into three distinct subfields: supervised, unsupervised, and semi-supervised. Training ML systems with tagged and adjustable data and known variable targets is what supervised learning is all about. Training

[1] *Associate Professor, Department of Computer Science, PGDAV College, University of Delhi,*

[2] *Associate Professor, Department of Computer Science, Miranda House, University of Delhi,*

* *Corresponding Author ✉ seema.aggarwal@mirandahouse.ac.in*

Received: 14 June, 2024
Available online: 31 July, 2024

models for fraud detection using carefully labelled transactions is widespread practice in many industries (Marazqah Btoush, et al., 2023). In contrast, unsupervised learning seeks for meaningful data patterns via dimension reduction and cluster segmentation while training ML on unknown target variables. To train models that can detect novel fraud techniques, semi-supervised learning combines labelled and unlabeled data.

Several methods have been explored in ML research that aim at detecting fraud. Support Vector Machines (SVM) and Decision Trees (DT) are two examples of popular supervised classifiers used for credit card fraud detection. To detect fraud, SVM employs datasets annotated with categories and a variety of kernel functions, including linear, radial, polynomial, and sigmoid. Financial fraud may also be detected using DT classifiers like C4.5 and LMT, both of which have shown impressive performance (Marazqah Btoush, et al., 2023). Credit card theft was correctly recognized by the C4.5 decision tree classifier with bagging ensemble 82.08% of the time. Another effective algorithm often used in fraud detection is called Random Forest (RF). It employs many decision trees to accurately identify credit card fraud (Marazqah Btoush, et al., 2023). In addition to the above-mentioned methods, Logistic Regression (LR), K-nearest neighbour (KNN), and Bayesian classification are also often used to detect fraud. Credit card fraud was identified using Random Forest(RF) with 95.19% accuracy, despite skewed nature of the data.

According to the (Marazqah Btoush, et al., 2023), Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) are two deep learning algorithms that have showed promise in identifying cyber fraud. These programs analyze massive databases using artificial neural networks to spot anomalies that can indicate fraud. Catboost-RF successfully identified fraudulent credit card transactions 99.5% of the time. Fraud detection also uses ensemble learning, which combines numerous classifiers to increase accuracy. To improve upon the functionality of preexisting classifiers, several boosting algorithms have been used. In sum, ML has shown to be a useful tool in identifying cyber fraud, and it may be used in tandem with more conventional approaches to further improve fraud detection. Researchers and practitioners may increase the safety of financial transactions and respond to unique fraud detection challenges using a wide range of algorithms.

4. Methodology

4.1 Dataset Description

The dataset used in the study consists of data on 362,620 unique online payment transactions. The eight columns in the dataset are : 'oldbalanceOrg,' 'newbalanceOrig,' 'oldbalanceDest,' 'newbalanceDest,' 'step,' 'type,' 'amount,' and 'fraud'. The 'step' column indicates the unit of time, with one step equaling one hour, and each row represents a separate transaction. The 'type' column indicates the nature of the corresponding electronic exchange ('TRANSFER,' 'PAYMENT,' 'CASH_OUT,' and 'DEBIT,' for example). The 'amount' column in the table indicates the monetary value of the transaction, whereas the 'oldbalanceOrg' and 'newbalanceOrig' columns show the originating customer's account balance before and after the transaction. Like the column, 'oldbalanceDest', column 'newbalanceDest' displays the previous and current balances for the designated recipient (Haoxiang & Smys, 2021). To achieve these objectives, 'isFraud' is the dataset's target variable and hence the most critical column. In this indication, a value of 1 indicates that a transaction is fraudulent, whereas a value of 0 indicates that it is valid. Predictive algorithms to identify fraudulent bank card and credit card purchases made online may benefit from this dataset.

Dataset Splitting : Dataset was divided into training and testing set so that online payment fraud may be detected. The Random Forest and Logistic Regression models were trained on the training set and tested on the testing set to see how well they performed. The standard split ratio 70:30(train: test) provides enough data for both purposes.

Feature Selection : Building reliable fraud detection algorithms relies heavily on feature selection. Features like 'OldbalanceOrg', 'NewbalanceOrig', 'Step', 'Type', 'Amount', 'OldbalanceDest', and 'NewbalanceDest' were utilized as inputs. The output label will be the value of the 'IsFraud' target variable.

Data Preprocessing : The early data exploration and cleaning phases have already taken care of data pretreatment tasks, including addressing missing values, encoding categorical variables, and scaling numeric features ensuring that the data is in a trainable format.

Model Selection : Due to their proficiency in dealing with complicated and unbalanced datasets, the Random Forest and Logistic Regression models have been selected for fraud detection (Dalal, Seth, Radulescu, Secara, & Tolea, 2022). Logistic Regression provides a binary classification approach well-suited to this issue, while Random Forest is a collaborative approach that constructs many decision trees and aggregates their results.

Model Training and Evaluation : Both the Random Forest and the Logistic Regression models were used to fit the training data. After training, models were scored according to assessment measures, including accuracy, precision, recall, and F1-score. Metrics like these indicate how well the algorithms can distinguish between fraudulent and legitimate deals.

Model Performance Comparison : The two models' efficacy was compared regarding accuracy, precision, recall, and F1 score. The objective is to choose the best model that can detect fraudulent activity (Deng, Ruan, Zhang, & Zhang, 2020).

Model Deployment and Monitoring : The top-performing model was used in factory environments to prevent fraudulent charges on online payments. Maintaining ever-evolving fraud tendencies and maintaining the model's efficacy, continuous monitoring, and regular model upgrades will be required.

4.2 Analysis and Discussion

A python script has been used that implements machine-learning methods to identify fraudulent online payments using numerous libraries and functions. Pandas has been used to read the dataset and preprocess the operations, including missing value management and categorical variable encoding. This process guarantees that the data is in a trainable model-friendly format (Sadineni, 2020). Random Forest and Logistic Regression are two machine-learning techniques used in the script for fraud detection. Models are trained and refined with the help of sci-kit-learn Random Forest and Logistic Regression classifier.

The algorithm excels at detecting patterns and outliers in data and is particularly useful for binary classification applications. Following model training, the code evaluates based on several criteria, including accuracy, confusion matrix, and classification report. The algorithms' capacity to identify fraudulent transactions from authentic transactions may be gauged using these criteria (Alabi & David) . The code reduces the steps of data preparation, model creation, and assessment typically associated with detecting fraudulent online payments by using these libraries and functions. It helps to construct reliable and accurate prediction models, letting banks and payment processors better safeguard their customers from fraudulent activities.

We start with preparing a .csv file containing information about electronic payment transactions. The code restricts the data frame to the first 362,620 rows to control the data size. Despite the computational complexity of processing the complete dataset, the code effectively loads and preprocesses the digital payment dataset, allowing for targeted analysis and building predictive algorithms for online payment fraud detection.

The code is a primary data exploration method for locating blanks in a dataset named "digital payment." The sum of missing values in a dataset's columns is found (Ilyas & Chu, 2019). The code then displays the results, making it easy to see empty columns. Also, the data types of each column and the total number of non-null items are summarized in one line employing digital payment.info().

Seaborn module in python is used to produce a scatter plot of the various digital payment transaction types. A count plot in which the x-axis reflects each transaction type, and the y-axis displays the total number of occurrences or frequency is shown in Figure 1. It is shown that the Cash-Out transaction type is higher than the other transaction details (Purohit & Vishwakarma, 2021). This visual aid clarifies the frequency distribution of the various transaction types in the data set. It is possible to get insight into the dataset's properties and may be used to draw conclusions about online payment fraud detection by analyzing the distribution of transaction types.

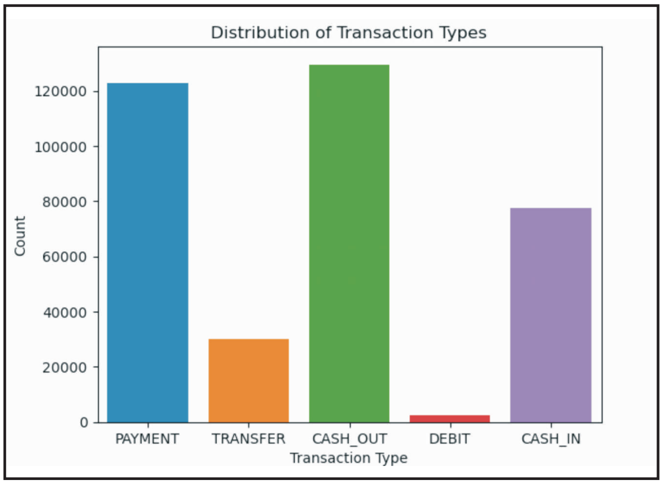

***Figure 1 :** Distribution of transaction types in the dataset*

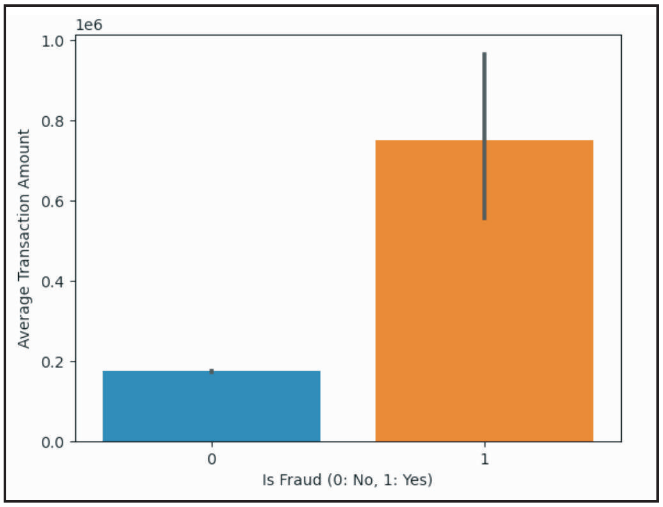

***Figure 2 :** Average transaction amount for fraudulent and non-fraudulent transactions*

We also examine the differential between the average transaction amounts of fraudulent and non-fraudulent transactions. Data visualization is accomplished via the bar plot function in the Seaborn library as shown in Figure 2. Utilizing the estimator option set to 'math_calc.mean,' the code determines the median transaction amount for fraudulent and genuine transactions. The data is then shown as a bar chart due to the Seaborn Library (Odegua & Ikpotokin, 2019). The generated graph gives users an easy way to compare the typical fraud and legitimate transaction values. Using this graph, users can find discrepancies in transaction amounts between the two groups. This data helps figure out how to spot fraudulent activity in online payment systems and prevent it in the future.

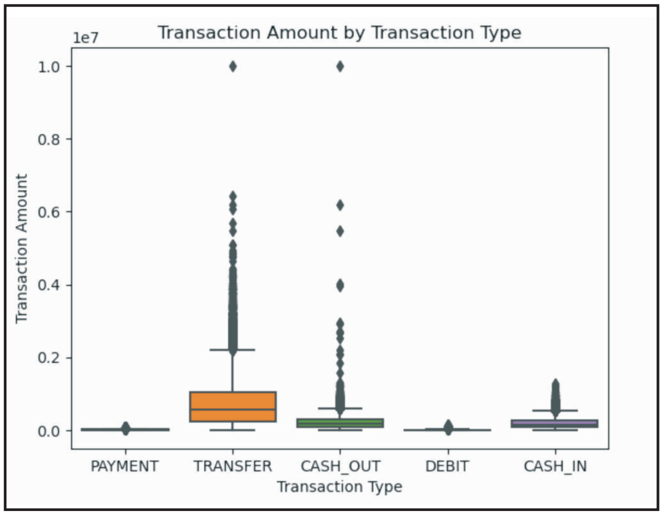

Figure 3 : *Relationship between the transaction amount and the type of transaction*

Figure 3 shows a boxplot, a visual representation of the dataset's connection between transaction amount and type, by using the Seaborn library. The boxplot provides an easy-to-understand visual representation of the median, range, and probable outliers of transaction amounts for each transaction type by graphing the 'Transaction Amount' on the y-axis alongside the 'Transaction Type' on the x-axis (Sahoo, Samal, Pramanik, & Pani, 2019). This visual aid makes it easy to see differences in the typical amount of money for various transaction types. It allows consumers to see whether there is a pattern of more significant or lower amounts in various transaction categories. Anomalies or patterns in fraudulent transactions may be easier to spot employing this data, which might help in the fight against fraud.

A scatter plot demonstrating how account balances for the source accounts have changed due to the transactions is shown in Figure 4. Each data point's 'oldbalanceOrg' and 'newbalanceOrig' values are shown in the scatter plot along the x- and y-axes, respectively. The 'isFraud' column assigns a corresponding colour to each data point; fraudulent transactions are shown in one colour, while legitimate ones are shown in another. The 'alpha=0.5' option makes the points half-transparent, highlighting overlaps between them. The impact of transactions on the balances of the source accounts is made clear by this visualization (Sorkun, Mullaj, Koelman, & Er, 2022). It helps spot out-of-the-ordinary behaviour that might indicate fraudulent activity. Scatter plots are helpful for immediately spotting cases when the 'oldbalanceOrg' considerably deviates from the 'newbalanceOrig,' suggesting the presence of questionable behaviour. During the first investigation phase in online payment fraud detection, this visualization helps comprehend the dynamics of balance changes in origin accounts.

Using a bar plot shown in Figure 5, we intend to determine the most common transaction types connected to fraudulent activity in the dataset of digital payments. With this paper, stakeholders, including financial institutions and payment service providers, can concentrate their efforts and resources on reducing risks and improving fraud detection strategies for transactions most frequently associated with fraud (Tayebi & El Kafhali, 2022).

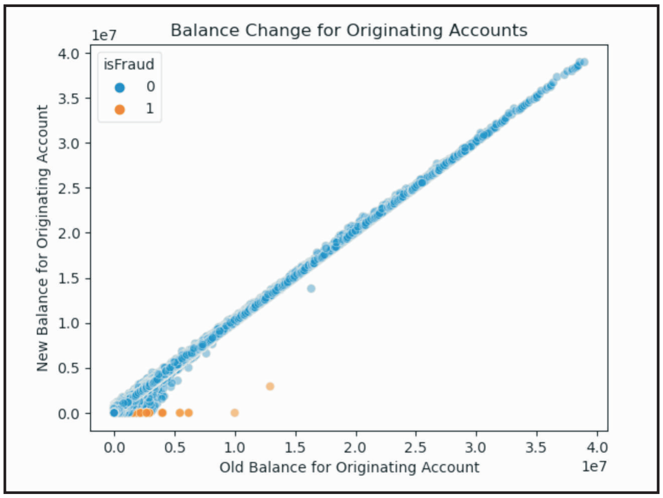

Figure 4 : *Balances change (old vs. new) for originating accounts after transactions*

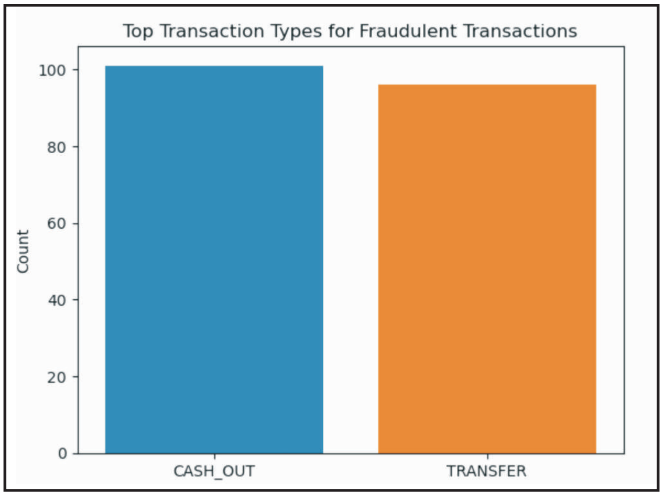

Figure 5 : *The top transaction types associated with fraudulent transactions*

Decision-makers may take steps to safeguard customers from possible financial losses and enhance the overall security of digital payment systems by visualizing the most common transaction types for fraudulent transactions.

The code has been developed to determine and display the average transaction amount for any transaction in the dataset for digital payments. The graph shown in Figure 6 makes it simple to compare the typical transaction amounts for various transaction kinds. This allows

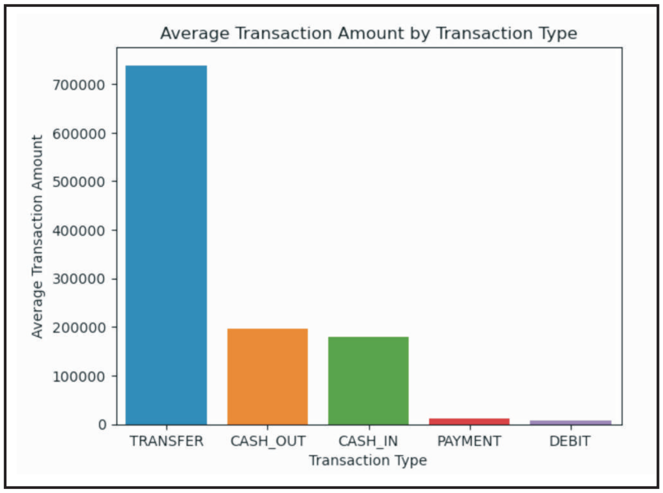

Figure 6 : *The average transaction amount for different type of transactions*

analysts and stakeholders to understand the average transaction amounts related to various categories (Nadim, Sayem, Mutsuddy, & Chowdhury, 2019). The identification of transaction trends, the detection of possible outliers, and a greater comprehension of the behaviour of the digital payment system all benefit from such insights. This data is crucial for detecting online payment fraud since it makes it easier to spot unusual activities or behaviour that could need additional examination.

With the help of the Seaborn library, we study how the average transaction amount changes over various time steps in the context of digital payments. The change in average transaction amount over time steps is shown in Figure 7. Users may see whether there are any trends or changes in the average transaction amount across various time steps by visually visualizing this data (Baratzadeh & Hasheminejad, 2022). A better understanding of transaction behaviour and the ability to spot any unusual activity that might be an indication of online payment fraud may be achieved by financial institutions as well as payment service providers with the aid of this visualization. Overall, this paper helps analyze the temporal distribution of transaction amounts, improves the effectiveness of fraud detection systems, and ensures the safety of digital financial transactions.

The paper helps in developing predictive models for detecting fraudulent online payments. The category 'type' variable is encoded as a number using Label Encoder in the code. Categorical data must be encoded since machine learning algorithms only accept numerical inputs. The 'type' feature is converted to numeric values so that the model may learn from the data and generate accurate predictions based on transaction kinds. StandardScaler has been used to uniformly scale the numeric features ('amount,' 'oldbalanceOrg,' 'newbalanceOrig,' 'oldbalanceDest,' as well as 'newbalanceDest') and to avoid any feature from overwhelming the learning process (Lewinson, 2020). Distance-based methods like kNN and gradient-based algorithms like Logistic Regression are susceptible to feature magnitudes; therefore, they, along with other methods, require careful consideration of scaling. Using sci-kit-learn' train_test_split function, we divide the data into training and testing sets. The predictive models will be trained using the training set (consisting of 70% of the data) and evaluated using the testing set (consisting of 30%). By separating it, researchers can evaluate the models' ability to extrapolate to new data.

Implementing the Random Forest model one can detect online payment fraud. To generate more accurate forecasts, the Random Forest algorithm averages the results from many different decision trees. The Random Forest model is built utilizing the Random Forest Classifier class from the scikit-learn package in this code (Devi, Biswas, & Purkayastha, 2019). The initial number of decision trees used in the model (n_estimators) is set to 100 and may be modified depending on experimental results and the model's overall effectiveness. Setting the random seed for random number production using the random_state option assures repeatability. After the model has been developed, it must be put through its paces by being fed data from the X_train and y_train datasets. The training data's features (input variables) are stored in X_train, while their labels (fraudulent or legal) are stored in y_train as the target variable. Once the model has been trained, the predict() method has been employed to produce predictions on the testing dataset (X_test). The outcomes are kept in the random forest_predictions variable. The trained model may predict the fraudulent status of fresh, unseen transactions for real-time fraud detection and better online payment system security.

We apply a Logistic Regression model to identify fraudulent transactions while making an online payment. In the

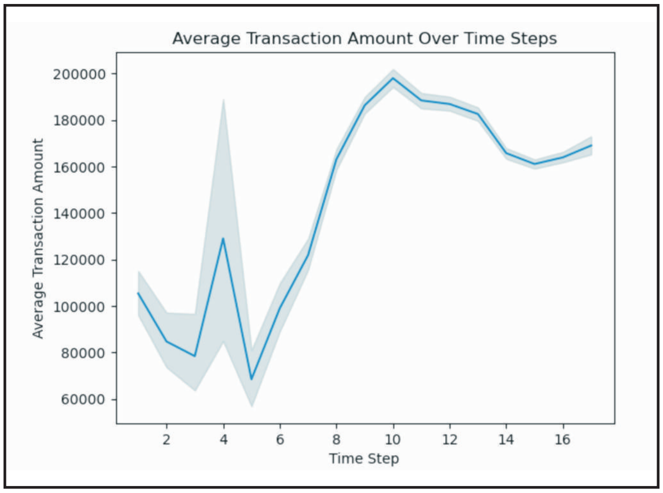

Figure 7 : *The average transaction amount variation over different time steps*

field of binary classification, Logistic Regression is a well-liked approach. We begin by generating a Logistic Regression model and then teaching it to make predictions based on the training data (X_train and y_train) (Hussein, Khairy, Najeeb, & Alrikabi, 2021). Predictions are made using the training model on the testing data (X_test), and the resulting file, lreg_predict, provides the predicted labels (fraudulent or non-fraudulent) for the test set. This method helps researchers determine how well the Logistic Regression algorithm detects fraudulent transactions and improves our perception of how well the model performs.

The effectiveness of the Random Forest model in detecting fraudulent online payments is discussed. A model's accuracy score indicates the rate at which its predictions are accurate. As the model is relatively successful at identifying fraudulent transactions from actual transactions, the printed accuracy of 0.9998 (or 99.98%) is very encouraging. Precision, recall, and F1-score are only a few of the assessment metrics included in the classification report for both the fraudulent and non-fraudulent classes. Moreover, it features support, representing each category's total number of samples (Trivedi, Simaiya, Lilhore, & Sharma, 2020). According to the classification report, the Random Forest model has excellent recall and accuracy for the non-fraudulent class (class 0), suggesting that it accurately detects the most valid transactions. But the model might not record certain fraudulent transactions because of the in comparison poor recall for the fraudulent class (class 1).

The measurement of the Logistic Regression model's effectiveness in detecting fraudulent online payments is done. The model has proven reliable in distinguishing fraudulent from valid transactions, as seen by the reported accuracy of 0.9996 (or 99.96%) (Muharemi, Logofătu, & Leon, 2019). According to the classification results, the Logistic Regression model is very accurate in identifying valid transactions (class 0). Class 1 fraud, however, has a poor recall, suggesting that the model may fail to catch all instances of fraud.

4.3 Model Comparison

Using the assessment findings, a model comparison between Random Forest and Logistic Regression for detecting fraud in online payments may be made. When comparing fraudulent and legitimate purchases, both models show a high degree of accuracy. However, each has its advantages and disadvantages.

The Random Forest model has achieved an accuracy of 99.98%, giving it an excellent choice for detecting fraudulent actions. Also, it correctly labels most valid transactions, demonstrating high accuracy and recall for this class. Although it has a high recall overall, its comparatively poor recall for the fraudulent class suggests it may miss certain fraudulent transactions.

On the other hand, the Logistic Regression model only managed 99.96% precision. Recall for the fraudulent class is much lower than the Random Forest model, despite its great accuracy for the non-fraudulent class. Therefore, the logistic regression approach may have a more significant false-negative rate and miss more fraudulent transactions.

The Random Forest model is the most effective alternative for identifying fraudulent online payments because of its higher recall rate. Both models have shown promise in detecting fraudulent activity. Still, with more tuning and feature engineering, they could do so much more, strengthening online payment systems' safety and confidence.

5. Conclusion and Recommendation

This study offers an in-depth investigation using machine learning algorithms, especially the Random Forest and Logistic Regression models, to identify fraudulent online financial transactions. The dataset is the basis for constructing and assessing the models, and it includes details on 362,620 distinct online payment transactions. High accuracy in identifying fraudulent transactions from legitimate transactions has been demonstrated by both the Random Forest and Logistic Regression models. The Random Forest model is entirely trustworthy in detecting fraud, reaching an astounding accuracy of 99.98%. For genuine purchases, it demonstrated similarly high levels of accuracy and memory. On the other hand, its recall for fraudulent transactions appeared low, which might cause fraudulent activity to be ignored. In contrast, the Logistic Regression model achieved an accuracy of 99.96%, significantly behind the other models.

This model has predicted a more significant false-negative rate than the Random Forest model, despite its excellent accuracy for legitimate transactions. The study emphasizes the significance of employing suitable machine-learning algorithms for fraud detection. Given its superior memory, the Random Forest model is the most reliable strategy for spotting illicit financial dealings. Both models have room for improvement in accuracy and efficiency in detecting fraudulent activity, but this might be achieved via more fine-tuning and feature engineering.

The research highlights the need to establish reliable fraud detection algorithms for online payment systems to prevent customers from suffering financial losses and maintain faith in digital payment platforms. The models must be constantly monitored and updated to maintain changing fraud patterns and ensure optimal performance in real-world circumstances. Increasing consumer trust and promoting a more secure and trustworthy financial system may be attained by improving online payment system security.

6. References

1. Alabi, B., & David, A. (n.d.). Framework for Detection of Fraud at Point of Sale on Electronic Commerce sites using Logistic Regression. *EAI Endorsed Transactions on Scalable Information Systems, 10*(2).

2. Baratzadeh, F., & Hasheminejad, S. M. (2022). Customer Behavior Analysis to Improve Detection of Fraudulent Transactions Using Deep Learning. *Journal of AI and Data Mining, 10*(1), 87-101.
3. Dalal, S., Seth, B., Radulescu, M., Secara, C., & Tolea, C. (2022). Predicting fraud in financial payment services through optimized hyper-parameter-tuned XGBoost model. *Mathematics, 10*(24).
4. Deng, R., Ruan, N., Zhang, G., & Zhang, X. (2020). FraudJudger: Fraud detection on digital payment platforms with fewer labels. *21st International Conference in Information and Communications Security, ICICS 2019.* Beijing, China.
5. Devi, D., Biswas, S., & Purkayastha, B. (2019). A cost-sensitive weighted random forest technique for credit card fraud detection. *2019 10th international conference on Computing, communication and networking technologies (ICT).*
6. Haoxiang, W., & Smys, S. (2021). A survey on digital fraud risk control management by automatic case management system. *Journal of Electrical Engineering and Automation, 3*(1), 1-14.
7. Hussein, A., Khairy, R. S., Najeeb, S. M., & Alrikabi, H. T. (2021). Credit Card Fraud Detection Using Fuzzy Rough Nearest Neighbor and Sequential Minimal Optimization with Logistic Regression. *International Journal of Interactive Mobile Technologies, 15*(5).
8. Ilyas, I. F., & Chu, X. (2019). *Data Cleaning .* ACM New Yorks, US.
9. Lewinson, E. (2020). *Python for Finance Cookbook: Over 50 recipes for applying modern Python libraries to financial data analysis.* Packt Publishing Ltd.
10. Marazqah Btoush, E. A., Zhou, X., Gururajan, R., Chan, K. C., Genrich, R., & Sankaran, P. (2023). A systematic review of literature on credit card cyber fraud detection using machine and deep learning. *PeerJ Comput Sci.* Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10280638/
11. Muharemi, F., Logofătu, D., & Leon, F. (2019). Machine learning approaches for anomaly detection of water quality on a real-world data set. *Journal of Information and Telecommunication, 3*(3), 294-307.
12. Nadim, A., Sayem, I., Mutsuddy, A., & Chowdhury, M. (2019). Analysis of machine learning techniques for credit card fraud detection. *2019 International Conference on Machine Learning and Data Engineering (iCMLDE) .*
13. Nicholls, J., Kuppa, A., & Le-Khac, N. A. (2019). Financial cybercrime: A comprehensive survey of deep learning approaches to tackle the evolving financial crime landscape. *IEEE Access, 9*, 163965-163986.
14. Odegua, R., & Ikpotokin, F. (2019). DataSist: A Python-based library for easy data analysis, visualization and modelling. arXiv preprint.
15. Punithalakshmi, M. P., & Rajakumar, M. (2021). An Analysis of Cyber Crime Prediction Model in Financial Sector Using Big Data Analytics. *webology, 18*(5).
16. Purohit, N., & Vishwakarma, R. (2021). Credit Card Fraud Detection Using Machine Learning Algorithms Using Python Technology. *webology, 18*(6).
17. Sadineni, P. (2020). Detection of fraudulent transactions in credit cards using machine learning algorithms. *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) .*
18. Sahoo, K., Samal, A., Pramanik, J., & Pani, S. (2019). Exploratory data analysis using Python. *International Journal of Innovative Technology and Exploring Engineering, 8*(12), 4727-4735.
19. Sorkun, M., Mullaj, D., Koelman, J., & Er, S. (2022). ChemPlot, a Python library for chemical space visualization.
20. Tayebi, M., & El Kafhali, S. (2022). Deep neural networks hyperparameter optimization using particle swarm optimization for detecting fraudulent transactions. *In Advances on Smart and Soft Computing: Proceedings of ICAC In 2021.* Singapore.
21. Trivedi, N., Simaiya, S., Lilhore, U., & Sharma, S. (2020). An efficient credit card fraud detection model based on machine learning methods. *International Journal of Advanced Science and Technology, 29*(5), 3414-3424.

❖ ❖ ❖