International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Online Payment Fraud Detection using Random Forest Algorithm

Vulugundam Anitha[1*], Chamakura Siri [2], Gandepelly Akanksha [3], Mathe Joshna [4], Kandula Sai Meghana [5]

[1]Assistant Professor, Dept. of Electronics and Telematics Engineering, G. Narayanamma Institute of Technology & Science (For Women), Hyderabad, Telangana, India.
[2,3,4,5] B. Tech students, Dept. of Electronics and Telematics Engineering, G. Narayanamma Institute of Technology & Science (For Women), Hyderabad, Telangana, India.

Abstract: In today's digital world, online transactions have become a part of everyday life, offering convenience, speed, and ease of use. However, they also come with risks like fraud, phishing, and data breaches. To tackle these challenges, we propose a machine learning-based fraud detection model that leverages feature engineering. By analyzing large volumes of data, the model learns, adapts, and improves over time, enhancing bothstability and accuracy in identifying fraudulent activities. These techniques play a crucial role in detecting online transaction fraud. By analyzing a dataset of online transactions, machine learning algorithms can spot unusual patterns that indicate fraudulent activity. Among these, the Random Forest Classifier has proven to be the most effective, achieving an impressive accuracy of 94.94%, outperforming other models in identifying suspicious transactions.

Keywords: Fraud Detection, Machine learning, Fraud Transactions, Random Forest, Classifiers, Accuracy.

1. Introduction

The use of artificial intelligence (AI) and machine learning in finance offers significant benefits, including improved customer satisfaction, reduced operational costs, and greater efficiency. In particular, machine learning techniques have been developed to detect credit card fraud by monitoring user activity to identify and prevent suspicious transactions.

Unfortunately, many fraud victims remain unaware of the scam until it's too late. While fraud detection systems are highly effective, implementing them in real-world scenarios comes with challenges. These systems must rapidly process large volumes of payment requests, determining which transactions to approve. Machine learning algorithms analyze approved transactions to detect unusual patterns, flagging potential fraud. Before investigators step in, cardholders are often asked to verify whether a transaction was legitimate or fraudulent.

2. Random Forest Detection Technique

Random Forest builds accuracy through numerous decision trees which function as ensemble learning for classification and regression operations. Random Forest achieves classification and regression outcomes through combining subsets of training data using features that create decision trees for group

voting decisions in classification and regression averaging predictions. Random Forest models present two main benefits because they decrease data variance while achieving higher performance from complex patterns and outlier conditions.

Three essential parameters of Random Forest need proper definition for proper implementation.

1. Each decision tree in the model generates the number of predictions in line with the parameters established ahead of time.

2. Decision tree structures receive their maximum depth specification through the max depth control.

3. Every tree within the forest contains a definite maximum number of features permitted for use.

4. Each node splitting requires a minimum number of samples which users must establish.

5. The decision tree structure contains leaf nodes that need minimum sample numbers defined as Min samples leaf.

Robust model prediction results are possible through hyperparameter optimization because it generates accurate prediction results efficiently.

The Random Forest algorithm works for both classification and regression tasks so it becomes effective in managing complex data types. The random forest can develop bias when individual decision trees are too basic yet it produces minimal improvements in cases of well-separated datasets. Its broad applications span bioinformatics sector and finance and marketing departments.

A procedure for Random Forest fraud detection involves multiple steps:

1. Dataset - The fraud detection dataset includes financial transaction records which have been named either fraudulent or legitimate. A transaction contains three main elements which are transaction type (Payment, Cash Out, Debit, Transfer, Cash In) and transaction amount in addition to old and new account balances of originator and beneficiary. The model trains its ability to detect fraudulent activities with the help of these features. [6]

2. Data Preprocessing – The dataset must go through cleaning first to generate correct model predictions. Our procedures find and remove unacceptable input data values plus prevent duplicate transactions from entering the system to maintain reporting accuracy. The team transforms categorical transaction types into numbers while making all numerical values equal each other. By creating new features the team uncovers significant information regarding account balance adjustments and transaction frequency. The data needs balancing because valid and fraudulent transactions differ in amount but we solve this issue by using either oversampling or under-sampling methods. [7]

3. Feature Extraction- Using feature extraction methods helps us find fraudulent actions by noting how transactions usually occur. Diverse components like transaction amount indicate fraud when combined with changes in balance and payment methods plus how often customers use them. The research method performs better when we use calculated features such as the balance change percentage and rapid payment indicators. Chosen features help accuracy and system efficiency improvement to make fraud detection systems more dependable. [8]

4.Machine learning model:The fraud detection model works with Random Forest which blends advanced analysis from many decision trees for better results. The system determines fraud based on what the combined decision of several trees indicates about transactions. This method decreases overfitting problems and enhances our fraud detection system's effectiveness. [9]

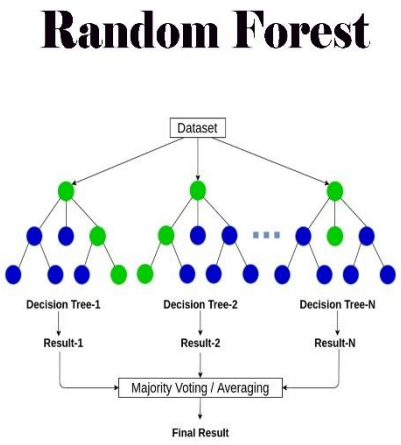

Fig 1: Description of how random forest performs.

Here's how the Random Forest model works:

1. The dataset gets divided into multiple random subsets so each resulting subset trains an individual decision tree.

2. The autonomous nature of each tree allows it to examine individual transactions until it identifies recognized patterns that define behaviour as legitimate or fraudulent.

3. The final outcome emerges through majority voting because it safeguards against any single tree controlling the result.

The Random Forest method demonstrates superiority in extensive data applications where it eliminates overfitting issues and achieves top detection results when identifying fraudulent activities.

4. Model Training and Evaluation:The dataset is divided in two based on model needs where 70% goes to training while testing uses the remaining 30%. Both building the Random Forest classifier and its accuracy evaluation utilize the training data while the testing data evaluates its effectiveness.

5. User Interface and System Implementation - A real-time fraud detection system implements a web interface as its primary mechanism. The system incorporates an easy-to-use interface which enables users to provide transaction information until it generates instant results at their fingertips.

Here's how the system works.

1. User Input: User Input includes transaction details that incorporate the type along with amount and sender's balance information.

2. Process Management: The system turns incoming data into an analytical format while performing normalization functions on it.

3. The Random Forest model used for fraud prediction evaluates transactions to produce a classification result between Fraud and No Fraud.

4. The system shows output information to the user which includes warning alerts when a transaction seems doubtful.

The system operates in real-time to defend against fraud during transactions and it functions as a component of banking apps and payment gateways.

3. Results and Discussion

A Random Forest model reaches 94.94% accuracy in identifying fraudulent activities with superior performance than other detection systems. The system proves effective for transaction fraud exposure through better security performance and false alarm management. The algorithm processes extensive data to evolve its fraud detection abilities which gives fraud prevention greater effectiveness.

	step	type	amount	oldbalanceOrg	newbalanceOrig	oldbalanceDest	newbalanceDest	isFraud
0	1	2	9839.64	170136.0	160296.36	0.0	0.0	No Fraud
1	1	2	1864.28	21249.0	19384.72	0.0	0.0	No Fraud
2	1	4	181.00	181.0	0.00	0.0	0.0	Fraud
3	1	1	181.00	181.0	0.00	21182.0	0.0	Fraud
4	1	2	11668.14	41554.0	29885.86	0.0	0.0	No Fraud

Fig2: Results of Transaction

Fig2 is described as:

- step: Represents the transaction's time step, likely in sequence.
- type: Specifies the transaction type (e.g., Payment, Cash Out, Transfer).
- amount: The total amount involved in the transaction.
- oldbalanceOrg: The sender's account balance before the transaction.

- newbalanceOrig: The sender's account balance after the transaction.
- oldbalanceDest: The receiver's account balance before the transaction.
- newbalanceDest: The receiver's account balance after the transaction.
- isFraud: Indicates if the transaction is fraudulent ("Fraud") or legitimate ("No Fraud").

First Transaction (Row 1):

- Amount: 9,839.64
- Sender's balance dropped from 170,136.00 to 160,296.36.
- Receiver's balance remained unchanged.
- Not fraudulent (**No Fraud**).

Second Transaction (Row 2):

- Amount: 1,864.28
- Sender's balance decreased from 21,249.00 to 19,384.72.
- Receiver's balance remained unchanged.
- Not fraudulent (**No Fraud**).

Third Transaction (Row 3):

- Amount: 181.00
- Sender's balance dropped from 181.00 to 0.00.
- Receiver's balance remained unchanged.
- Marked as fraudulent (**Fraud**).

Fourth Transaction (Row 4):

- Amount: 181.00
- Sender's balance decreased from 181.00 to 0.00.
- Receiver's balance increased from 0.00 to 21,182.00.
- Marked as fraudulent (**Fraud**).

Fifth Transaction (Row 5):

- Amount: 11,668.14
- Sender's balance dropped from 41,554.00 to 29,885.86.
- Receiver's balance remained unchanged.
- Not fraudulent (**No Fraud**).

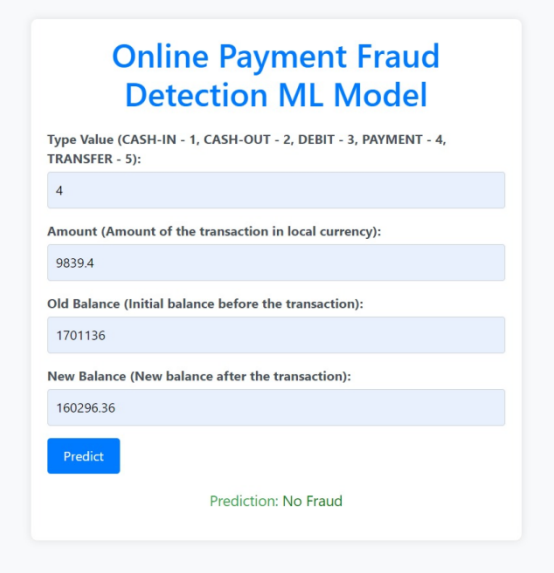

Fig3: Online payment fraud detection ML Model - I

Type: PAYMENT-4
Amount: 9,839.4
Old Balance: 1,701,136
New Balance: 160,296.36

- **Payment Type:** A standard payment transaction, common in financial systems.
- **Transaction Amount:** The payment of 9,839.4 is moderate compared to the total balance.
- **Balance Change:** The account started with 1,701,136 and now holds 160,296.36, leaving a significant remaining balance.
- **Fraud Prediction:** The model classifies this as **No Fraud**, as the transaction, though large, does not show suspicious patterns like a sudden withdrawal or irregular behaviour.

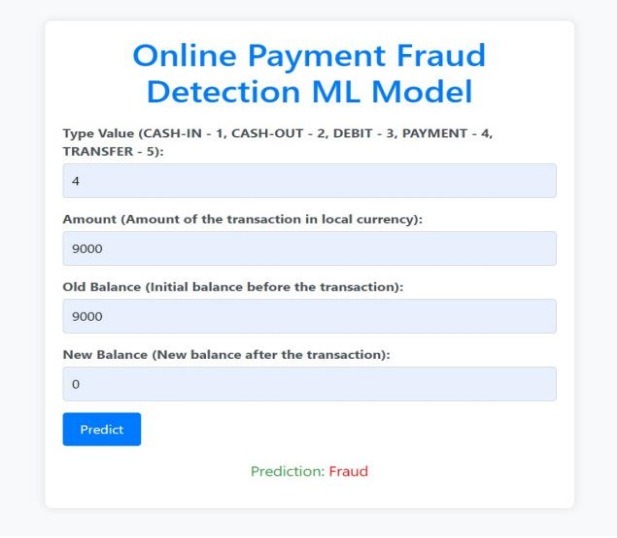

Fig4: Online payment fraud detection ML Model - II

Type: PAYMENT - 4
Amount: 9,000 (local currency)
Old Balance: 9,000
New Balance: 0

- **Full Balance Depletion:** The entire account balance is spent in a single transaction. Sudden depletion, especially in smaller accounts, is often flagged as suspicious.
- **Unusual Behavior:** If the account usually holds more than 9,000, spending everything at once could be a red flag.
- **Fraud Risk:** Using all available funds in one go is uncommon and might suggest fraudulent activity, particularly if unexpected or unauthorized.

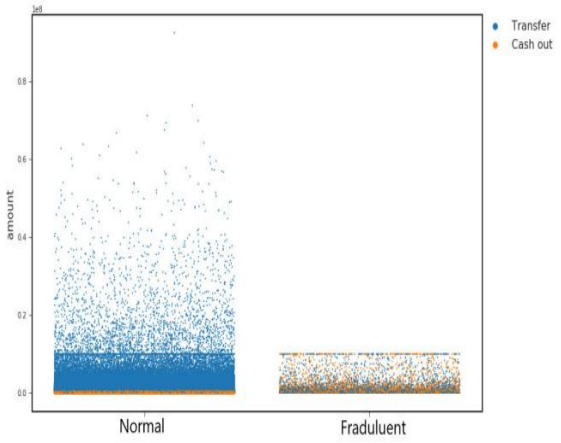

Fig5: Dispersion of fraudulent and normal transactions over time.

Figure 5 shows how fraudulent and normal transactions are distributed over time, highlighting clear patterns.

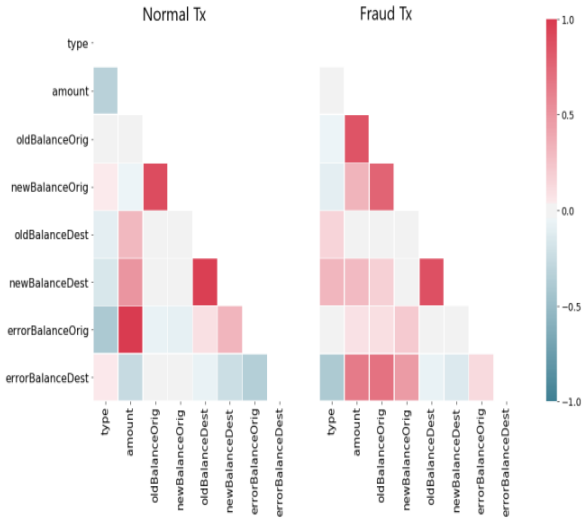

Fig6: Fraudulent and genuine transactions by analysing their correlations using Heatmap.

Fraudulent transactions are spread more evenly over time, while genuine ones mostly involve CASH-OUTs. In contrast, fraud cases show a balanced mix of CASH-OUTs and TRANSFERs. The 'jitter' parameter in the plotStrip function helps separate overlapping transactions for better visualization.
Figure 6 uses heatmaps to compare fraudulent and genuine transactions, highlighting key differences. Among supervised machine learning models, Random Forest performs best, achieving 99.994% accuracy, a precision of 0.9548, recall of 0.5075, Log-loss of 0.01888, and an F1 score of 0.66274 (TABLE 1). With the lowest Log-loss and strong overall performance, Random Forest outperforms other models.[10]

Parameters	Accuracy	Precision score	Recall score	Log loss	F1-score
SVM	55.41 %	0.00049	0.21021	15.3986	0.00099
LR	95.73 %	0.00722	0.28828	1.47404	0.01410
Naïve Bayes	98.61 %	0.02203	0.27678	0.47992	0.04081
KNN	97.10 %	0.00170	0.04524	0.99856	0.00328
Decision tree	99.94 %	0.01470	0.61538	0.02190	0.02873
Random Forest	99.994%	0.95480	0.50750	0.01888	0.66274

Table 1: Accuracy, Precision Score, Recall Score and F1 Score when Different Algorithms are used

Table 2 presents the classification performance without feature selection, evaluated using various metrics. Logistic Regression, Naïve Bayes, Decision Tree, and Random Forest are compared with SVM. The results show that the proposed SVM method achieves 95.35% accuracy, 94.20% sensitivity, 93.72% specificity, and an AUC of 0.939, outperforming other classifiers.[11]

Methods	Accuracy (%)	Specificity (%)	Sensitivity (%)	AUC
Logistic Regression	90.78	88.64	87.40	0.875
Navie Bayes	92.65	91.45	90.59	0.923
Decision Tree	89.65	90.89	92.23	0.896
Random Forest	93.49	86.57	89.99	0.911
SVM	95.35	93.72	94.20	0.939

Table 2: Classification Without Feature Selection Performance

CONCLUSION

Random Forest is the most effective algorithm for detecting online transaction fraud, achieving 99.994% accuracy. Its strength lies in handling large datasets and recognizing key patterns, improving fraud detection. Accurate results depend on feature engineering, data preprocessing, and managing class imbalances. While realtime fraud detection helps reduce banking losses, it cannot fully prevent fraud before processing. Combining Gradient Boosting with data selection and balancing techniques enhances detection further. AIpowered fraud detection strengthens financial security, safeguarding users and fostering trust.

REFERENCES

1. Vedant Mayekar, Siddharth Mattha, Sohan Choudhary, Prof Amruta Sankhe´ "Online Fraud Transaction Detection Using Machine Learning", Vol 08, May 2021.
2. Darshan Aladakatti, Gagana P, Ashwini Kodipalli, Shoaib Kamal, "Fraud detection in Online Payment Transaction using Machine Learning Algorithms", 2022, IEEE.
3. Ashwini Gajakosh, R Archana Reddy, Myasar Mundher Adnan, G Rajalaxmi, PramodhiniR "Fraud Detection In Credit Card Using Competitive SwarmOptimizationWithSupportVectorMachine." 2024International Conference On Distributed Computing And Optimization Techniques (ICDCOT).
4. U. Siddaiah, P. Anjaneyulu, Y. Haritha, M. Ramesh, "Fraud Detection in Online Payments using Machine Learning Techniques", 2023, IEEE.
5. Lochan S, Sumanth H V, Ashwini Kodipalli, Rohini B.R., Trupthi Rao, Pushpalatha V., "Online Payment Fraud Detection Using Machine Learning", 2023, IEEE.
6. Abdulwahab Ali Almazroi, Nasir Ayub, "Online Payment Fraud Detection Model Using Machine Learning Techniques", 2023, IEEE Access.
7. Lochan S, Sumanth H V, Ashwini Kodipalli, Rohini B.R., Trupthi Rao, Pushpalatha V., "Online Payment Fraud Detection Using Machine Learning", 2023, IEEE.
8. Lochan S, Sumanth H V, Ashwini Kodipalli, Rohini B.R., Trupthi Rao, Pushpalatha V., "Online Payment Fraud Detection Using Machine Learning", 2023, IEEE.
9. Vandavasi Baba Mahesh, Koneru Venkata Sai Chandra, Lammata Shiva Prasad Babu, VelagalAarthiSowjanya, Dr Moulana Mohammed, "Clicking Fraud Detection for Online Advertising using Machine Learning", KL University, 2024, IEEE.
10. Darshan Aladakatti, Gagana P, Ashwini Kodipalli, Shoaib Kamal, "Fraud detection in Online Payment Transaction using Machine Learning Algorithms", 2022, IEEE.
11. Ashwini Gajakosh, R Archana Reddy, Myasar Mundher Adnan, G Rajalaxmi, PramodhiniR "Fraud Detection in Credit Card Using Competitive Swarm Optimization with Support Vector Machine." 2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT).