



Archives available at [journals.mriindia.com](http://journals.mriindia.com)

International Journal on Advanced Electrical and Computer Engineering

ISSN: 2349-9338

Volume 14 Issue 01, 2025

## Online Payment Fraud Detection using Machine Learning

Pritam Eknath Bhakve<sup>1</sup>, Aniket Nana Jambhale<sup>2</sup>, Vijaykumar. S. Kumbhar<sup>3</sup>

Department of Computer Science, Shivaji University Kolhapur<sup>1,2,3</sup>

[pritambhakve@gmail.com](mailto:pritambhakve@gmail.com)<sup>1</sup>, [jambhaleaniket97@gmail.com](mailto:jambhaleaniket97@gmail.com)<sup>2</sup>, [vsk\\_csd@unishivaji.ac.in](mailto:vsk_csd@unishivaji.ac.in)<sup>3</sup>

Peer Review Information	Abstract
<p><i>Submission: 08 Jan 2025</i>  <i>Revision: 11 Feb 2025</i>  <i>Acceptance: 04 March 2025</i></p> <p><b>Keywords</b></p> <p><i>Machine Learning</i>  <i>Fraud Detection</i>  <i>Decision Tree</i>  <i>Random Forest</i>  <i>Transaction Data</i></p>	<p>The rapid growth of e-commerce has greatly boosted the frequency of online payment fraud, and poses a great challenge to both financial institutions and consumers. This paper proposes a machine learning-based fraud detection system utilizing random forest and decision structure algorithms, which are capable of detecting suspicious transactions efficiently. This model is trained according to historical transaction data and has characteristics such as steps, transaction type, Amount, oldbalance(origin), newbalance(origin), oldbalance(dest), newbalance(dest) and fraud indicators. Comprehensive testing shows strong accuracy, accuracy, recall and F1 scores for the model, which reduces both false positives and false negatives. This work also examines how selecting and adjusting characteristics via hyperparameters affects and reports that fine-tuning of such factors improves the effectiveness of identification. To compensate for the lessons that characterize fraud perception, this study uses class weighting techniques to improve model skills and recognize fraudulent transactions. The results examine the potential for machine learning to improve fraud prevention systems and provide a scalable, customizable solution to improve the security of digital financial transactions.</p>

### Introduction

The threat of payment fraud has grown exponentially with the accelerated expansion of online financial transactions, and this has posed severe issues for both customers and financial institutions. Deceptive actions will result in heavy economic losses, impact the business operations and erode confidence in electronic payment systems. Traditional dominant fundamental scale detection systems often stay behind the ever-changing strategies used by fraudsters. Machine

learning (ML) has proven to be an effective tool for these problems thanks to its ability to address large data records, capture undiscovered patterns, and create accurate projections.

Numerous transaction attributes like procedure (timestep), type (transaction type: Payment, Cash-in, Debit, Transfer), amount (transaction amount), target credits and isFraud (target label: fraud 1, legal 0) are part of the Kaggle dataset used for training the model. Data records are handled to maintain consistency by addressing missing

values, categorical variables, and normalization. Accuracy is the primary evaluation measure, and training and testing are done in a split of 80-20. Jolib will be used in the future to save it to maintain uniformity in applying and testing trained models of new data. Assess the performance of your model and present a detailed performance check with metrics that incorporate accuracy, accuracy, recall and F1 scores. In addition, we use feature analysis to find out which features have the largest predictions of fraudulent transactions and offer valuable information regarding causes of payment fraud.

By using strategies such as a class weighting to prevent the model from being biased in majority classes (legal submission), this study also addresses class imbalance, common issues of fraud detection. This improves fraud detection accuracy.

By using random forest and decision-making models, this study aims to develop a robust and scalable fraud making systems that can classify transaction a fraudulent or legal.

## LITERATURE REVIEW

Online payment fraud detection is heavily based on ML methods such as Decision Tree (DT), Random Forest (RF), Logistics Regression (LR), and SVM, providing greater accuracy. Functional engineering (amount, location, device) provides improved performance, and real-time detection via streaming data increases fraud prevention. Hybrid models using two or more algorithms show higher accuracy and reliability, but have difficulty with class failures and false positive results [1].

This work identifies flaws in traditional fraud detection systems and introduces ML-based solutions with improved accuracy random forests. Analyze transaction data (payment methods, accounting, etc.) to mark transactions as fraudulent or actual. The system implements a flask to recognize real-time and is evaluated using accuracy, accuracy, recall, and F1 scores. This means that increasing online payment security demonstrates benefits [2].

The literature highlights the shift to machine learning models (ML) for conventionally based

systems to identify online payment fraud. Decision Tree (DT) and Random Forest (RF) are common algorithms, with RF improving accuracy and having fewer numbers. The big challenges are class, real-time detection, and data protection protection. The model is tested using accuracy, accuracy, recall, and F1 scores. Future trends demonstrate hybrid approaches and deep learning to improve fraud detection [3].

With the rapid growth of e-commerce, there is a necessity to detect fraud in online payment. Machine learning models such as Decision Trees and Random Forest efficiently identify suspicious transactions based on factors like amount, location, and user behavior. Overcoming class imbalance and using real-time detection can improve the accuracy. Latest researches consider hybrid models to enhance the efficiency of fraud detection [4].

The literature highlights the use of algorithms in machine learning, such as decision-making, to identify online payment fraud. Key transaction attributes, including amounts and locations, help identify suspicious transactions. The assignment is a real time processing to the class hungry. The new development emphasizes ensemble models and deep learning to improve accuracy and efficiency [5].

With the rise in online transactions, fraud detection has become essential. The literature states that machine learning models such as decision-making-random forests, deep learning, etc. are recorded for fraudulent payment detection. Some of the challenges are class disorders and actual time detection. While Small technology can help improve accuracy, research focuses on future hybrid models to improve fraud detection [6].

If online scams are growing, accurate recognition is essential. Machine learning models such as decision making, random forests and neural networks are precision options highlighted in the literature. Although RF is stable, deep learning manages complex patterns. Data - Changes in echo weights and fraud strategies are challenges. Methods such as Small improve the output of the model [7].

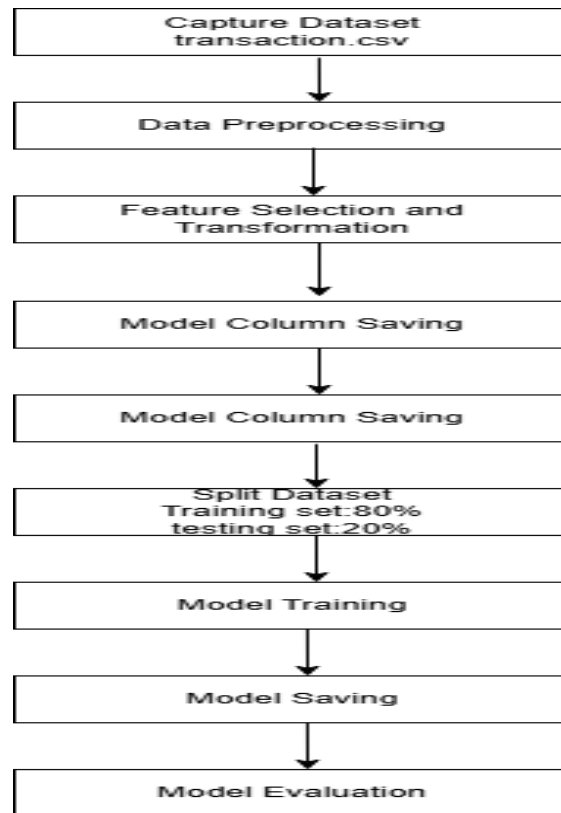
**Flowchart:**

Fig. 1 : Flowchart of the working model

transactions-checkpoint.csv X

C:\Users\admin\Desktop> dataset > transactions-checkpoint.csv > data

	step	type	amount	nameOrig	oldbalanceOrig	newbalanceOrig	nameDest	oldbalanceDest	
1	1	PAYMENT	7107.77	C154988899	183195	176087.23	M408069119	0	0
8	1	PAYMENT	7861.64	C1912850431	176087.23	168225.59	M633326333	0	0
9	1	PAYMENT	4024.36	C1265012928	2671	0	M1176932104	0	0
10	1	DEBIT	5337.77	C712410124	41720	36382.23	C195600860	41898	40348.79
11	1	DEBIT	9644.94	C1900366749	4465	0	C997608398	10845	157982.12
12	1	PAYMENT	3099.97	C249177573	20771	17671.03	M2096539129	0	0
13	1	PAYMENT	2560.74	C1648232591	5070	2509.26	M972865270	0	0
14	1	PAYMENT	11633.76	C1716932897	10127	0	M801569151	0	0
15	1	PAYMENT	11633.76	C1716932897	10127	0	M801569151	0	0

Fig. 2: Sample transaction collected and stored in csv format

**RESEARCH METHODOLOGY**

**Capture Dataset:** Pandas is used to import transaction-checkpoint.csv dataset and is stored in dataframe. This includes recordings of money transaction and detail about the type, amount and effectiveness of the transaction. These raw data are extremely important for training and testing the model.

**Data Preprocessing:** Problem prevention, the isFraud field is quantified in numeric format to prevent data records with missing values. Data cleaning improve performance by receiving accurate and constant information of the model.

**Feature Selection and Trans formation:** Key attributes such as step, type, amount and balance

etc selected for the training the model. Type column are formatted and converted to numeric format with machine learning algorithms.

**Model Column Saving:** The function column names used for training are stored in the model columns. pkl. if you provide a model for new data, you will receive a consistency format.

**Split Dataset:** Using train\_test\_split, the training data is split 80% and 20% for testing. The test set determines the accuracy of the model for unknown data, while the training set trains for model for the pattern.

**Model Training:** Model trained into,

**Random Forest:** An ensemble model that reduces overfitting by combining multiple

decision trees.

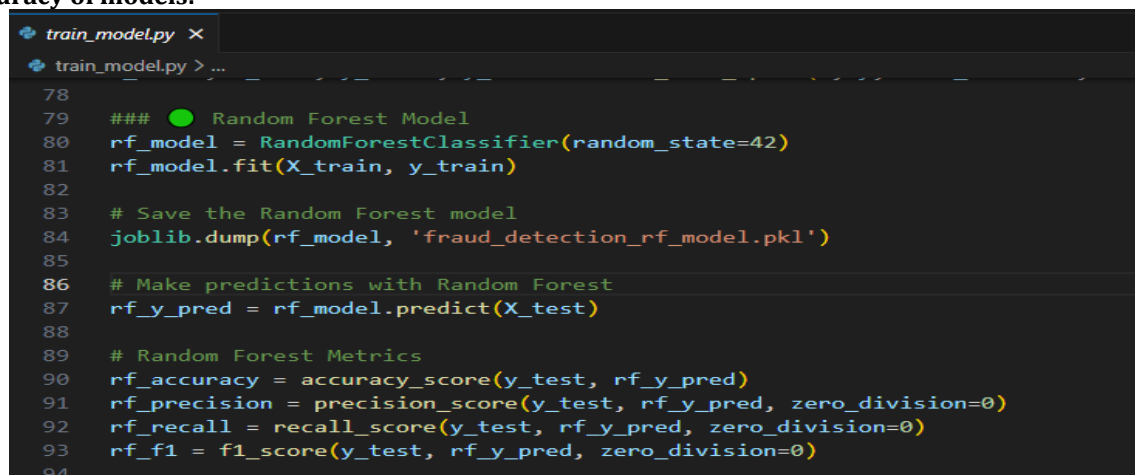
**Decision Tree:** A less complex model that employs a tree-like model to categorize transactions.

**Model Saving:**

Both models decision tree and random forest are stored in .pkl file, so you can remain for future use without retraining. This way you can reuse the solution.

**Model Evaluation:** The model can be evaluated with accuracy\_score() to see how it differentiates between invalid or regular transactions. The evaluation includes accuracy, precision, recall and F1 scores to measure overall performance.

### Accuracy of models:

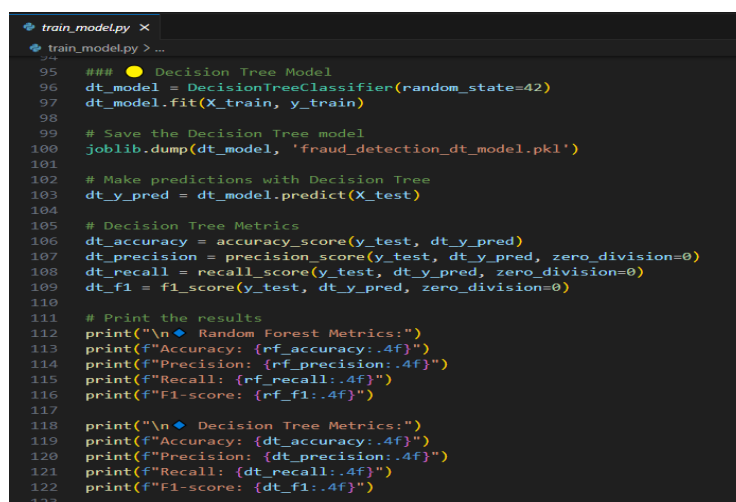


```

train_model.py x
train_model.py > ...
78
79 ### Random Forest Model
80 rf_model = RandomForestClassifier(random_state=42)
81 rf_model.fit(X_train, y_train)
82
83 # Save the Random Forest model
84 joblib.dump(rf_model, 'fraud_detection_rf_model.pkl')
85
86 # Make predictions with Random Forest
87 rf_y_pred = rf_model.predict(X_test)
88
89 # Random Forest Metrics
90 rf_accuracy = accuracy_score(y_test, rf_y_pred)
91 rf_precision = precision_score(y_test, rf_y_pred, zero_division=0)
92 rf_recall = recall_score(y_test, rf_y_pred, zero_division=0)
93 rf_f1 = f1_score(y_test, rf_y_pred, zero_division=0)
94

```

Fig. 3: Sample code used during calculation of accuracy of model



```

train_model.py x
train_model.py > ...
95 ### Decision Tree Model
96 dt_model = DecisionTreeClassifier(random_state=42)
97 dt_model.fit(X_train, y_train)
98
99 # Save the Decision Tree model
100 joblib.dump(dt_model, 'fraud_detection_dt_model.pkl')
101
102 # Make predictions with Decision Tree
103 dt_y_pred = dt_model.predict(X_test)
104
105 # Decision Tree Metrics
106 dt_accuracy = accuracy_score(y_test, dt_y_pred)
107 dt_precision = precision_score(y_test, dt_y_pred, zero_division=0)
108 dt_recall = recall_score(y_test, dt_y_pred, zero_division=0)
109 dt_f1 = f1_score(y_test, dt_y_pred, zero_division=0)
110
111 # Print the results
112 print("\n Random Forest Metrics:")
113 print(f"Accuracy: {rf_accuracy:.4f}")
114 print(f"Precision: {rf_precision:.4f}")
115 print(f"Recall: {rf_recall:.4f}")
116 print(f"F1-score: {rf_f1:.4f}")
117
118 print("\n Decision Tree Metrics:")
119 print(f"Accuracy: {dt_accuracy:.4f}")
120 print(f"Precision: {dt_precision:.4f}")
121 print(f"Recall: {dt_recall:.4f}")
122 print(f"F1-score: {dt_f1:.4f}")
123

```

```
PS E:\OnlinePaymentFroudDetectoion> python train_model.py

◆ Random Forest Metrics:
Accuracy: 0.9960
Precision: 1.0000
Recall: 0.7143
F1-score: 0.8333

◆ Decision Tree Metrics:
Accuracy: 0.9920
Precision: 0.7500
Recall: 0.6429
F1-score: 0.6923
PS E:\OnlinePaymentFroudDetectoion> |
```

## RESULT AND DISCUSSION

### Random Forest Metrics:

Accuracy: 0.9960  
Precision: 1.0000  
Recall: 0.7143  
F1-score: 0.8333

### Decision Tree Metrics:

Accuracy: 0.9920  
Precision: 0.7500  
Recall: 0.6429  
F1-score: 0.6923

## CONCLUSION

The Online Payment Fraud Detection solution accurately detects the fraud transactions utilizing Random Forest and Decision Tree algorithms. Through the data preprocessing and conversion, the solution provides appropriate and consistent inputs to train models. The models prove to have high accuracy with consistent performance as assured by testing metrics. Trained and saved in.pkl format, the models easily allow reuse when predicting in future instances, resulting in the scalability and effectiveness of the solution to use in actual fraud detection.

## References

- R. A. Priya Gupta Shubhi Jain Prof, "Online payment fraud detection using ml," *JETIR*, p. 7, may 2024.
- T. B. Saikiran Namani Harsh Mordharia Nayan Gajare, "Online payment fraud detection using machine learning," *IRJETS*, p. 7, 2024.
- A. R. R. M, "FRAUD DETECTION IN ONLINE PAYMENT," *IJARIE*, p. 7, 2024.
2. N. 1M.N. Naga Keerthi, "ONLINE PAYMENT FRAUD DETECTION," *ijcrt*, p. 11, 2024.
- S. N. H. M. N. G. T. Bemila, "Online payment fraud detection using machine learning," *IRJETS*, 202.
- P. P. G. M. 2Mallipudi Devi Siva Sai, "Online payment fraud detection using ml," *international Journal of Innovative Science and Research Technology*, vol. 8, no. 10, p. 5, october 2023.
- M. C. M. B. B. S. P. PRATHYUSHA, "ONLINE PAYMENT FRAUD DETECTION," (*INT-JECSE*, p. 8, 2023.
- C. M. D. M. Dr. M. Venkatesh Bhukya Keerthi Budati Bhargavi, "Online payment fraud detection using machine learning," *IJARIE*, p. 5, 2024.
- R. A. priya gupta Shubhi JainProf, "Online payment fraud detection using machine learning," *JETIR*, p. 7, May 2024.
- T. S. Chawla, "Online Payment Fraud Detection using ml," *National College of Ireland*, p. 19, 2023.
- P. P. G. M. N. V. S. G. ., I. G. 2Mallipudi Devi Siva Sai, "Online payment fraud detection using machine learning," *International Journal of Innovative Science and Research Technology*, vol. 8, no. 10, p. 5, oct 2023.
- A. A. A. 1. A. N. A. 2, Online payment fraud detection model using machine learning, *IEEE*, 2023.
- L. X. B. M. F. G.-P. 2. J. J. M. H. Ludivia Hernandez Aros 1✉, Financial fraud detection through the application of machine learning, *Humanaties and social science communication*, 2023.
- M. A. Y. M. A. R. M. H. O. R. B. R. R. Ananya Sarker, credit card fraud detection using machine learning, *scientfic research publishing*, 2024.
- Kharade, S. K., Kharade, K. G., Kamat, R. K., & Kumbhar, V. S. (2020). Setting Barrier to Removable Drive through Password Protection for Data Security. *Our Heritage*, 68(27), 19-23.