iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

ONLINE PAYMENT FRAUD DETECTION USING MACHINE LEARNING

Mr. N. NAVVEN KUMAR
Assistant Professor, Department of Information Technology,
Jawaharlal Nehru Technological University Hyderabad,
naveen.cse.mtech@gmail.com

JANGALA NAGALAXMI
Post Graduate Student, M. TECH (SE) Department of Information Technology,
Jawaharlal Nehru Technological University Hyderabad,
nagalaxmijangala7@gmail.com

ABSTRACT
Online payment fraud has become a critical challenge in the digital economy, leading to substantial financial losses and eroding consumer trust. The rise of web surfing and online shopping, so came the use of credit cards for online transactions, as did the prevalence of online financial fraud. This study focuses on developing a machine learning-based system to detect and prevent fraudulent transactions in online payment platforms. The proposed solution involves data preprocessing, feature engineering, and the selection of appropriate machine learning models such as Logistic Regression, XG Boost Classifier, Random Forests, and SVC. Given the imbalanced nature of the dataset, where fraudulent transactions are rare, advanced techniques are employed to enhance model accuracy. The evaluation metrics include accuracy, confusion matrix. The system is designed for real-time deployment, offering a robust mechanism to reduce fraudulent activities and improve the security and reliability of online payment systems.

Keywords:
Fraud detection, Machine learning, Random Forest, SVM, Gradient Boosting algorithms, classification, Data preprocessing.

INTRODUCTION
The rapid growth of online financial transactions has made digital payments an integral part of modern commerce. However, this convenience has also led to an increase in fraudulent activities, resulting in substantial financial losses for businesses and consumers. Online payment fraud involves unauthorized transactions where fraudsters exploit vulnerabilities in payment systems, causing significant damage to financial institutions. In the digital age, the proliferation of online payment systems has revolutionized commerce, offering unprecedented convenience and accessibility. However, this growth has also been accompanied by a rise in fraudulent activities, posing significant risks to both consumers and businesses. Online payment fraud encompasses a range of illicit activities, from unauthorized transactions to identity theft, and it can lead to substantial financial losses and a loss of trust in digital platforms. As fraudsters continuously develop new tactics, traditional rule-based detection methods struggle to keep pace with evolving threats.Traditional rule-based fraud detection systems rely on predefined rules and thresholds to identify suspicious activities. While effective in some cases, these systems struggle to adapt to evolving fraud patterns, leading to a high number of false positives and missed fraudulent transactions.

OBJECTIVES
The Main objective of the study is to identify the challenges in the implementation of the newly adopted depth Building an ML model for detecting fraudulent transactions in real-time. Exploring various ML techniques, including Supervised and Unsupervised Learning approaches. Handling data imbalance using techniques like SMOTE (Synthetic Minority Over-sampling Technique) or cost-sensitive learning. Reducing false positives to ensure a seamless user experience .Evaluating model performance using key metrics like Precision, Recall, F1-Score. Deploying the trained model for real-time fraud detection in a financial system.

Volume-09 Issue 06, June-2025
ISSN: 2456-9348
Impact Factor: 8.232

IJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

SYSTEM ARCHITECTURE

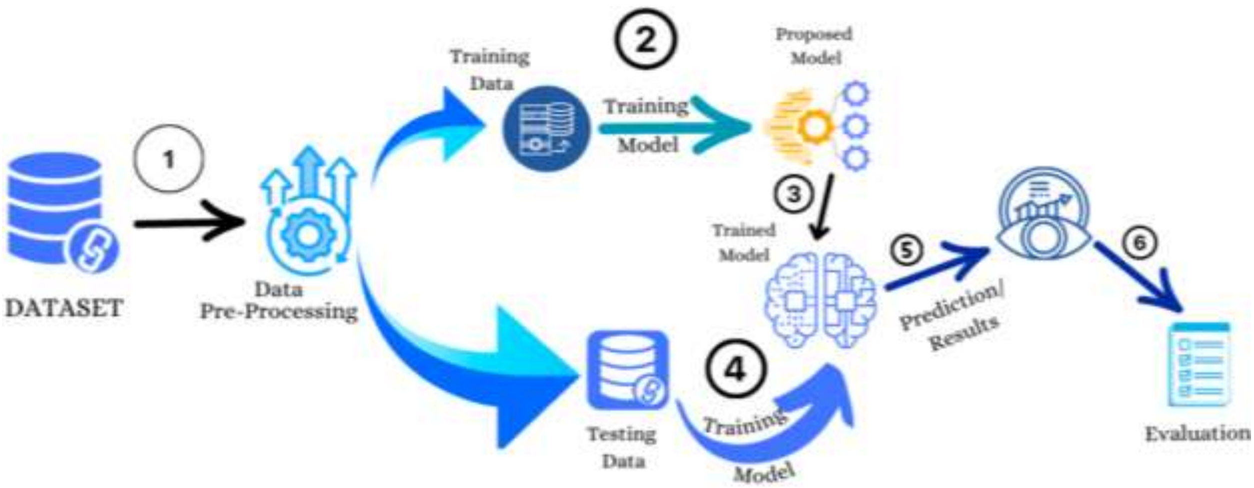


Fig 1: System architecture

METHODOLOGY

1.Data Collection
Collect transaction data including features such as transaction amount, timestamp, user details, payment method, and geographical location. Ensure that data includes both legitimate and fraudulent transactions
The dataset used for training and testing the model contains online transaction data.
It includes the following columns:
type: Type of online transaction.
amount: The amount of the transaction.
NameOrig: Customer starting the transaction.
OldbalanceOrg: Balance before the transaction.
newbalanceOrig: Balance after the transaction.
nameDest: Recipient of the transaction.
oldbalanceDest: Initial balance of recipient before the transaction.
newbalanceDest: The new balance of recipient after the transaction.

.

ISSN: 2456-9348
Impact Factor: 8.232

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

A	B	C	D	E	F	G	H	I	J	K
step	type	amount	nameOrig	oldbalance(	newbalance	nameDest	oldbalance(	newbalance	isFraud	isFlaggedFraud
1	PAYMENT	9839.64	C12310068	170136	160296.36	M19797871	0	0	0	0
1	PAYMENT	1864.28	C16665442	21249	19384.72	M20442822	0	0	0	0
1	TRANSFER	181	C13054861	181	0	C55326406	0	0	1	0
1	CASH_OUT	181	C84008367	181	0	C38997010	21182	0	1	0
1	PAYMENT	11668.14	C20485377	41554	29885.86	M12307017	0	0	0	0
1	PAYMENT	7817.71	C90045638	53860	46042.29	M57348727	0	0	0	0
1	PAYMENT	7107.77	C15498889	183195	176087.23	M40806911	0	0	0	0
1	PAYMENT	7861.64	C19128504	176087.23	168225.59	M63332633	0	0	0	0
1	PAYMENT	4024.36	C12650129	2671	0	M11769321	0	0	0	0
1	DEBIT	5337.77	C71241012	41720	36382.23	C19560086	41898	40348.79	0	0
1	DEBIT	9644.94	C19003667	4465	0	C99760839	10845	157982.12	0	0
1	PAYMENT	3099.97	C24917757	20771	17671.03	M20965391	0	0	0	0
1	PAYMENT	2560.74	C16482325	5070	2509.26	M97286527	0	0	0	0
1	PAYMENT	11633.76	C17169328	10127	0	M80156915	0	0	0	0
1	PAYMENT	4098.78	C10264838	503264	499165.22	M16353782	0	0	0	0
1	CASH_OUT	229133.94	C90508043	15325	0	C47640220	5083	51513.44	0	0
1	PAYMENT	1563.82	C76175070	450	0	M17312179	0	0	0	0
1	PAYMENT	1157.86	C12377626	21156	19998.14	M18770629	0	0	0	0
1	PAYMENT	671.64	C20335245	15123	14451.36	M47305329	0	0	0	0
1	TRANSFER	215310.3	C16709931	705	0	C11004390	22425	0	0	0
1	PAYMENT	1373.43	C20804602	13854	12480.57	M13445190	0	0	0	0
1	DEBIT	9302.79	C15665112	11299	1996.21	C19735381	29832	16896.7	0	0

2.Data Preprocessing

Data Cleaning: Handle missing values, remove duplicates, and correct any inconsistencies.

Feature Engineering: Create new features that could help in distinguishing fraudulent transactions (e.g., transaction frequency, average transaction amount, user behaviour patterns).

Normalization/Standardization: Normalize or standardize numerical features to ensure they are on a similar scale.

Categorical Encoding: Convert categorical variables into numerical formats using techniques like one-hot encoding or label encoding.

3.Data Splitting

Divide the data into training, validation, and test sets (e.g., 70% training, 15% validation, 15% test).

4.Model Selection and Training

Logistic Regression: A baseline method for binary classification.

Decision Trees: For their interpretability and ability to handle non- linear relationships.

Random Forests: An ensemble method that improves classification accuracy and reduces overfitting.

Gradient Boosting Machines(GBM): Includes XG Boost , which are powerful for handling complex patterns.

Neural Networks: For deep learning approaches, if the dataset is large and complex.

Train Models : Fit each selected model on the training dataset.

5. Model Evaluation

Evaluate Performance:

-Accuracy: The proportion of correctly classified transactions.

-Precision: The proportion of true positives among the predicted positives.

-Recall: The proportion of true positives among the actual positives.

F1 Score: The harmonic mean of precision and recall.

6. Model Selection

select the best model for detect the fraud transactions

7.Testing and Final Evaluation

Test Final Model: Evaluate the chosen model on the test set to assess its performance on unseen data.

Analyze Results: Review the results and determine if further tuning or adjustments are needed.

8. Deployment

Deploy Model: Integrate the model into the production environment where it can analyze real-time transactions.

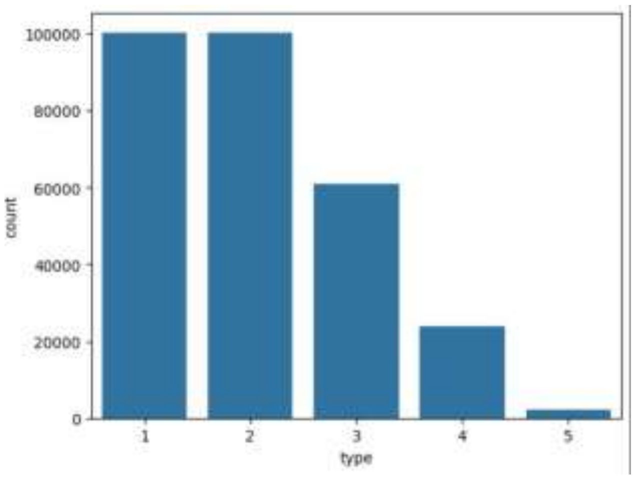
ISSN: 2456-9348
Impact Factor: 8.232

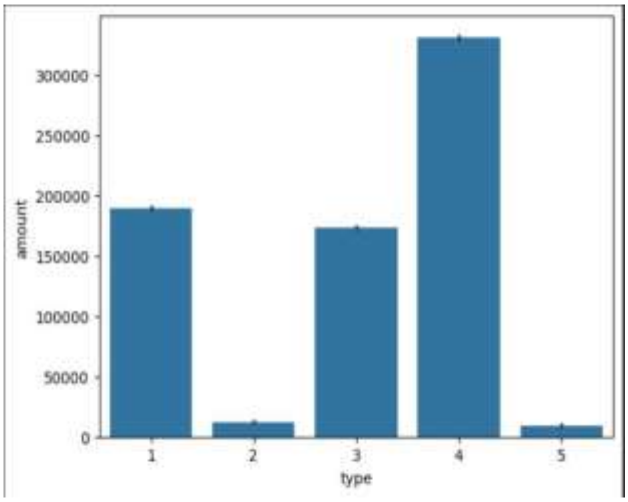
iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

RESULTS AND DISCUSSION

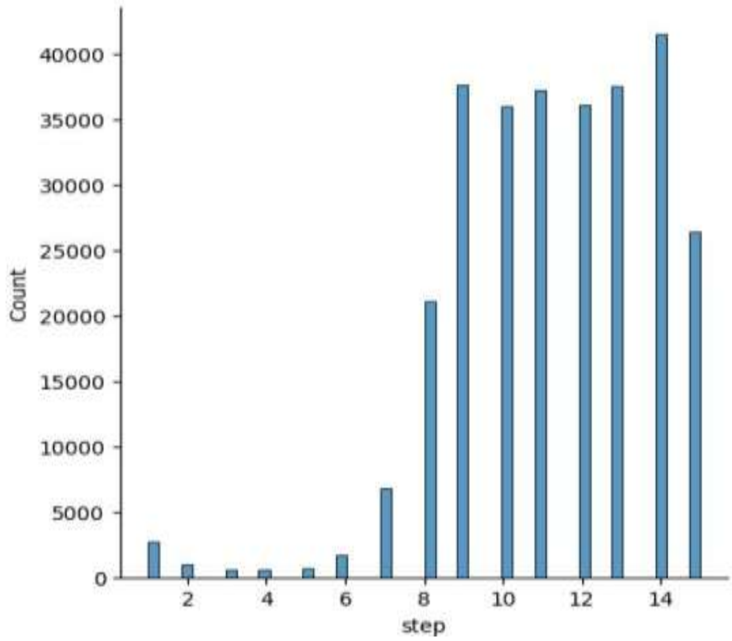
Considering the results of all the above-mentioned supervised machine learning algorithms we came to know random forest algorithm is the best suited algorithm for the detection of online transaction fraud with accuracy of 99.94%. with respect to accuracy by comparing logistic regression is 96.1% and XGB classifier is 99.04% and SVM is 95.7% . However, looking to performance we conclude random forest is the best accuracy results.

a.Count plot of payment type using Seaborn Library: **b .Bar plot for analyzing Type & Amount:**





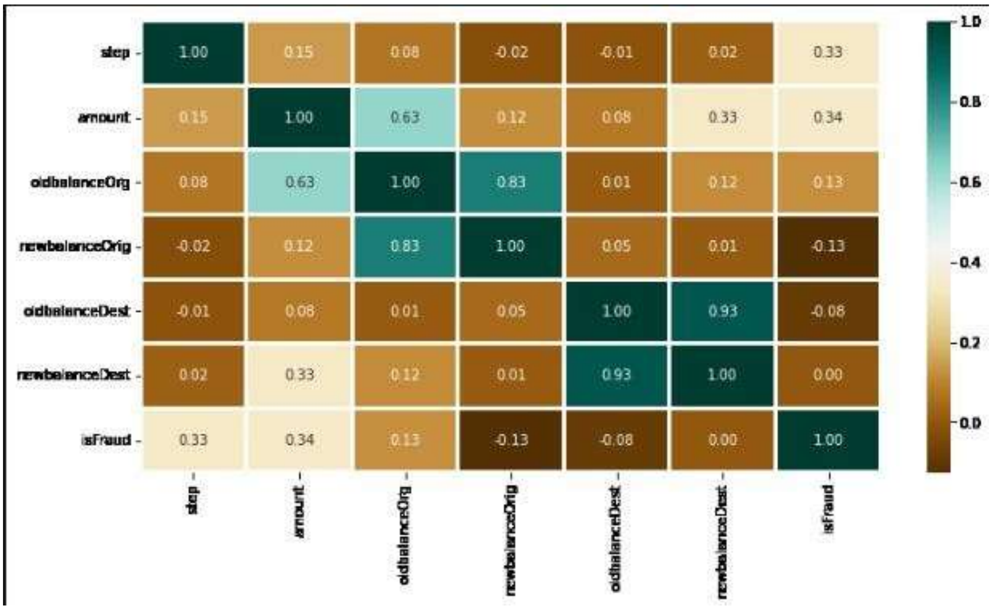
c.Distribution of Step column:



ISSN: 2456-9348
Impact Factor: 8.232

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

d.Correlation among different features using Heatmap



e. Model Training and Model Evaluation

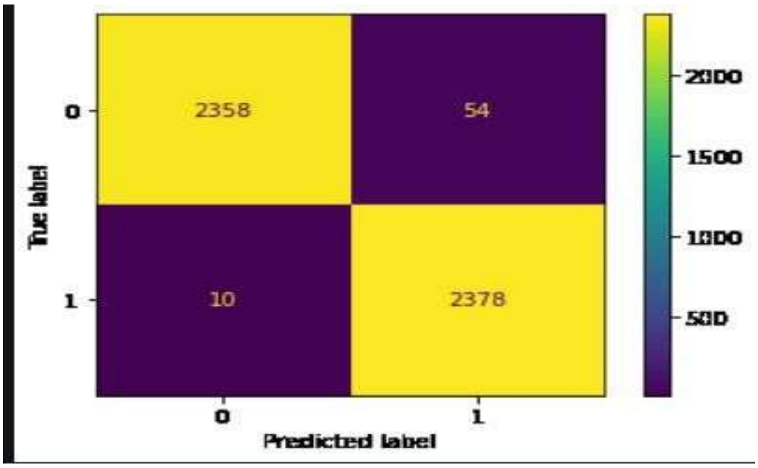
```
LogisticRegression() :
Training Accuracy :  0.9610946236487818
Validation Accuracy :  0.9650647516187905

XGBClassifier() :
Training Accuracy :  0.9990647916240432
Validation Accuracy :  0.9988292242028274

SVC(probability=True) :
Training Accuracy :  0.9577130392435476
Validation Accuracy :  0.9610511096110737

RandomForestClassifier(criterion='entropy', n_estimators=7, random_state=7) :
Training Accuracy :  0.9999942442337746
Validation Accuracy :  0.9966858546463663
```

f.Confusion Matrix



iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

OUTPUTSCREENS

International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

CONCLUSION

The project on online payment fraud detection using machine learning algorithms has demonstrated the potential and effectiveness of leveraging advanced computational techniques to combat financial fraud. By employing a combination of supervised learning models, such as Random Forest, XGBoost, and Support Vector Classifier, along with feature engineering and data preprocessing, the system has achieved a significant level of accuracy in identifying fraudulent transactions. The results underscore the importance of utilizing diverse algorithms and ensemble methods to enhance detection capabilities and reduce false positives. The implementation of these models in real-time scenarios can substantially mitigate the risk of financial loss for businesses and consumers alike, enhancing the overall security of online payment systems..

REFERENCES

[1] Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy" by Dal Pozzolo, Andrea, et al. (2014).
[2] "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection" by Khan, Z. U., & Fiaidhi, J. (2020). Journal of King Saud University - Computer and Information Sciences.
[3] "Pattern Recognition and Machine Learning" by Christopher M. Bishop (2006). Springer.
[4] “A Fraud Detection System Using Machine Learning” by Dhananjay Kalbande, Pulin Prabhu, Anisha Garat, Tanja Rajab Pumsirirat, Apapan, and Yan Liu.
[5] " Moreno, Felipe Marino, et al. "Transfer Learning." Solanki, Jatin et al.“A STEP FORWARD IN FRAUD DETECTION SYSTEM USING MACHINE LEARNING.” Journal of critical reviews (2020): n. pag.
[6] Devika, S. P., Nisarga, K. S., Rao, G. P., Chandini, S. B., & Rajkumar,N. (2019). A research on credit card fraudulent detection system.
[7] International Journal of Recent Technology and Engineering, 8(2),5029– 503 - https://doi.org/10.35940/ijrte.B1083.078219
[8] International Conference on Advanced Computational Intelligence,ICACI 2012, 188–190. https://doi.org/10.1109/ICACI.2012.6463148
[9] Raghavan, P., & Gayar, N. E. (2020). Fraud Detection using Machine Learning and Deep Learning. In 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) (pp.334-339).IEEE. https://doi.org/10.1109/ICCIKE47802.2019.90042 31
[10] Sweers, Tom, Tom Heskes, and Jesse Krijthe. "Autoencoding Credit Card Fraud." Bachelor Thesis (2018).
[11] Maes, Sam, et al. "Credit card fraud detection using Bayesian and neural networks." Proceedings of the 1st international naiso congress on neuro fuzzy technologies. Vol. 261. 2002.