Credit Card Froud Detection Using Machine Learning and Logistic Regration

Sonali Ananda Avatade[1], Dr. Bere Sachin. S[2]
[1,2]*Department of Computer Engineering, Dattakala Group of Institutions*

***Abstract*— Credit card frauds are easy and friendly targets. E-commerce and many other online sites have increased the online payment modes, increasing the risk for online frauds. Increase in fraud rates, researchers started using different machine learning methods to detect and analyse frauds in online transactionsThe main aim of the paper is to design and develop a novel fraud detection method for Streaming Transaction Data, with an objective, to analyse the past transaction details of the customers and extract the behavioural patterns. Where cardholders are clustered into different groups based on their transaction amount. Then using sliding window strategy [1], to aggregate the transaction made by the cardholders from different groups so that the behavioural pattern of the groups can be extracted respectively. Later different classifiers [3],[5],[6],[8] are trained over the groups separately. And then the classifier with better rating score can be chosen to be one of the best methods to predict frauds. Thus, followed by a feedback mechanism to solve the problem of concept drift [1]. In this paper, we worked with European credit card fraud dataset**

I. INTRODUCTION

Credit card generally refers to a card that is assigned to the customer (cardholder), usually allowing them to purchase goods and services within credit limit or withdraw cash in advance. Credit card provides the cardholder an advantage of the time, i.e., it provides time for their customers to repay later in a prescribed time, by carrying it to the next billing cycle.

Credit card frauds are easy targets. Without any risks, a significant amount can be withdrawn without the owner's knowledge, in a short period. Fraudsters always try to make every fraudulent transaction legitimate, which makes fraud detection very challenging and difficult task to detect.In 2017, there were 1,579 data breaches and nearly 179 million records among which Credit card frauds were the most common form with 133,015 reports, then employment or tax-related frauds with 82,051 reports, phone frauds with 55,045 reports followed by bank frauds with 50,517 reports from the statics released by FTC [10].

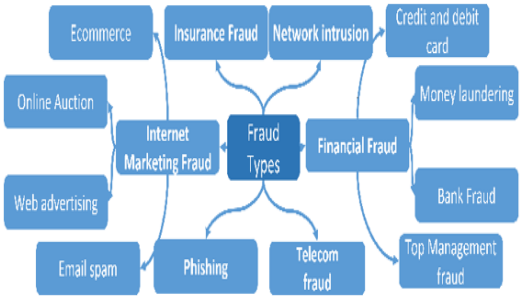


Fig. 1: Taxonomy for Frauds

With different frauds mostly credit card frauds, often in the news for the past few years, frauds are in the top of mind for most the world's population. Credit card dataset is highly imbalanced because there will be more legitimate transaction when compared with a fraudulent one. As advancement, banks are moving to EMV cards, which are smart cards that store their data on integrated circuits rather than on magnetic stripes, have made some on-card payments safer, but still leaving card-not-present frauds on higher rates. According to 2017 [10], the US Payments Forum report, criminals have shifted their focus on activities related to CNP transactions as the security of chip cards were increased. Fig 2, shows the number of CNP frauds cases that were registered in respective years.

Main title
An Intelligent System for Credit Card Fraud Detection Using Machine Learning
Machine Learning-Based Approach for Detecting Credit Card Fraud Transactions
A Comparative Study of Machine Learning Algorithms for Credit Card Fraud Detection

II. LITERATURE SURVEY

Multiple Supervised and Semi-Supervised machine learning techniques are used for fraud detection [8], but we aim is to overcome three main challenges with card frauds related dataset i.e., strong class imbalance, the inclusion of labelled and unlabelled samples, and to increase the ability to process a large number of transactions. Different Supervised machine learning algorithms [3] like Decision Trees, Naive Bayes Classification, Least Squares Regression, Logistic Regression and SVM are used to detect fraudulent transactions in real-time datasets. Two methods under random forests [6] are used to train the behavioural features of normal and abnormal transactions. They are Random-tree-based random forest and CART-based. Even though random forest obtains good results on small set data, there are still some problems in case of imbalanced data. The future work will focus on solving the above-mentioned problem. The algorithm of the random forest itself should be improved. Performance of Logistic Regression, K-Nearest Neighbour, and Naïve Bayes are analysed on highly skewed credit card fraud data where Research is carried out on examining meta-classifiers and meta-learning approaches in handling highly imbalanced credit card fraud data. Through supervised learning methods can be used there may fail at certain cases of detecting the fraud cases. A model of deep Auto-encoder and restricted Boltzmann machine (RBM) [2] that can construct normal transactions to find anomalies from normal patterns. Not only that a hybrid method is developed with a combination of Adaboost and Majority Voting methods

III. PROPOSED METHOD

Card Transactions Are Always Unfamiliar When Compared to Previous Transactions Made the Customer. This Unfamiliarity Is a Very Difficult Problem in Real-World When Are Called Concept Drift Problems [1]. Concept Drift Can Be Said as a Variable Which Changes Over Time and In Unforeseen Ways. These Variables Cause A High Imbalance in Data. The Main Aim of Our Research Is to Overcome the problem of concept drift to implement on real-world scenario. table 1, [1] shows basic features that are captured when any transaction is made.

Table 1: Raw features of credit card transactions

Attribute name	Description
Transaction id	Identification number of a transaction
Cardholder id	Unique Identification number given to the cardholder
Amount	Amount transferred or credited in a particular transaction by the customer
Time	Details like time and date, to identify when the transaction was made
Label	To specify whether the transaction is genuine or fraudulent

3.1 Dataset Description

The dataset [11] contains transactions made by a cardholder in duration in 2 days i.e., two days in the month of September 2013. Where there is total 284,807 transactions among which there are 492 i.e., 0.172% transactions are fraudulent transactions. This dataset is highly unbalanced. Since providing transaction details of a customer is considered to issue related to confidentiality, therefore most of the features in the dataset are transformed using principal component analysis (PCA). V1, V2, V3, V28 are PCA applied features and rest i.e., 'time', 'amount' and 'class' are non-PCA applied features, as shown in table 2.

Table 2: Attributes of European dataset

S. No.	Feature	Description
1.	Time	Time in seconds to specify the elapses between the current transaction and first transaction.
2.	Amount	Transaction amount
3.	Class	0 - not fraud 1 – fraud

Fig. 3 shows the correlation matrix of the dataset. This matrix explains that attribute class is independent of both the amount and time of the transaction was made. It is even clear from the matrix; the class of the transaction is depending on PCA applied attributes.

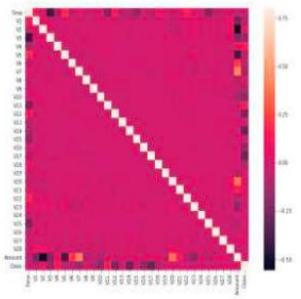


Fig. 3: Correlation Matrix for Attributes (both the X and Y axis show different attributes present in dataset)

IV. RESULTS AND DISCUSSION

The machine learning models are tested using transaction datasets to identify fraudulent activities. Among the algorithms tested, Random Forest provides the highest accuracy due to its ability to handle complex data patterns and large datasets.
The evaluation metrics indicate that the proposed machine learning model successfully identifies fraudulent transactions with high accuracy while reducing false positives. This demonstrates the effectiveness of machine learning techniques in financial fraud detection systems.

V. PROBLEM STATEMENT

Credit card fraud detection is a challenging task because fraudulent transactions are extremely rare compared to legitimate transactions. This imbalance in the dataset makes it difficult to train machine learning models effectively.

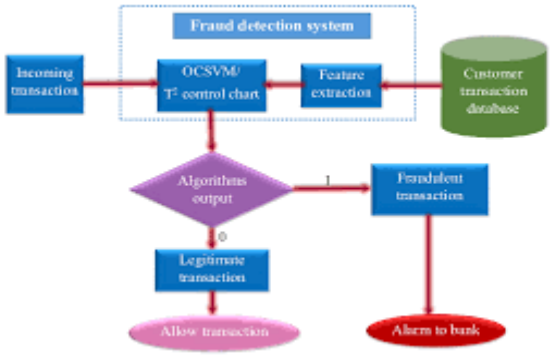
Additionally, fraudsters continuously change their techniques to avoid detection. Traditional rule-based systems cannot easily adapt to these changes.
Therefore, there is a need for an intelligent fraud detection system that can automatically learn transaction patterns and identify suspicious activities in real time.

VI. PROPOSED SYSTEM

The proposed system uses machine learning algorithms to detect fraudulent credit card transactions. The system analyzes transaction data and classifies transactions as legitimate or fraudulent.

The system consists of the following components:
1. Data Collection
2. Data Preprocessing
3. Feature Selection
4. Model Training
5. Fraud Detection
6. Performance Evaluation

Proposed Diagram:



Methodology: Credit Card Fraud Detection Using Machine Learning and Logistic Regression
Used Algorithms:
1. Logistic Regression
2. Decision Tree
3. Random Forest
4. Support Vector Machine (SVM)

1. Logistic Regression
Logistic Regression predicts whether a credit card transaction is fraudulent or legitimate based on the probability calculated from transaction features.

2. Decision Tree
Decision Tree classifies transactions by creating a tree-like structure of decision rules based on transaction characteristics.

3. Random Forest
Random Forest improves fraud detection accuracy by combining the predictions of multiple decision trees and selecting the majority vote.

4. Support Vector Machine (SVM)
Support Vector Machine (SVM) identifies fraudulent transactions by finding the optimal boundary that separates fraud and legitimate transactions in the dataset.
1. Problem Definition
The objective of this project is to develop a machine learning model that can accurately identify fraudulent credit card transactions and distinguish them from legitimate transactions. Since fraudulent transactions are rare compared to genuine transactions, the problem is treated as a binary classification task.

2. Data Collection

- Obtain the credit card transaction dataset from a reliable source.
- The dataset contains transaction details such as:
  - Transaction amount
  - Transaction time
  - Anonymized features (V1, V2, V3, ..., V28)
  - Class label (0 = Genuine Transaction, 1 = Fraudulent Transaction)

3. Data Preprocessing

a) Data Cleaning

- Check for missing values.
- Remove duplicate records.
- Verify data consistency and integrity.

b) Data Transformation

- Normalize numerical features such as transaction amount and time.
- Convert data into a suitable format for machine learning algorithms.

c) Handling Class Imbalance

Since fraudulent transactions are significantly fewer than legitimate ones:

- Apply undersampling or oversampling techniques.
- Use SMOTE (Synthetic Minority Oversampling Technique) if required.

4. Exploratory Data Analysis (EDA)

- Analyze transaction patterns.
- Visualize class distribution.
- Identify correlations among features.
- Detect outliers and unusual transaction behavior.

Tools:

- Python
- Pandas
- NumPy
- Matplotlib
- Seaborn

5. Feature Selection

- Select relevant features influencing fraud detection.
- Remove redundant and highly correlated features.
- Improve model efficiency and accuracy.

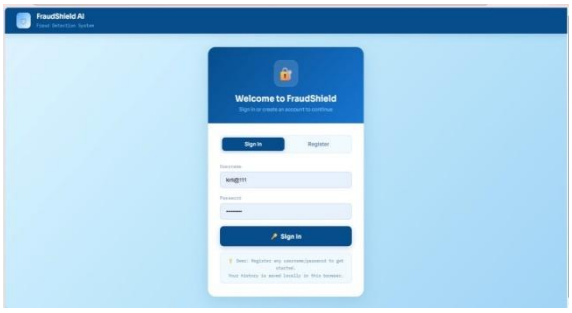
6. Dataset Splitting

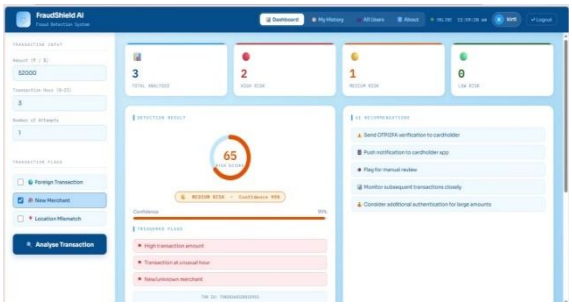
Split the dataset into:

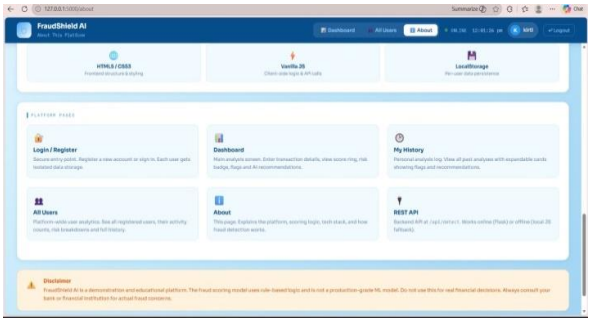
- Training Set: 80%
- Testing Set: 20%

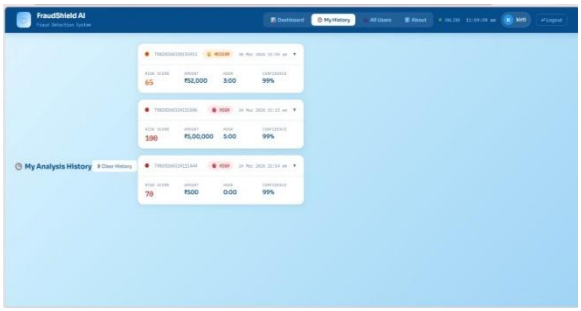
This ensures that the model is evaluated on unseen data.

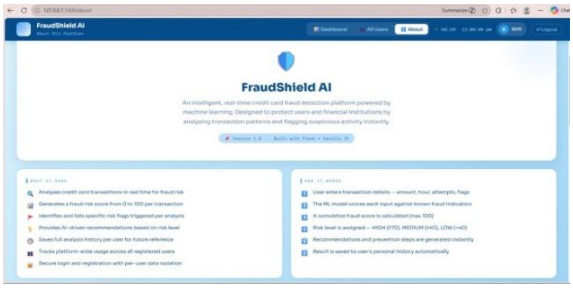
Project Result:

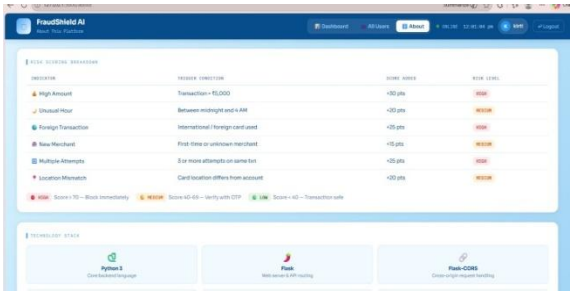


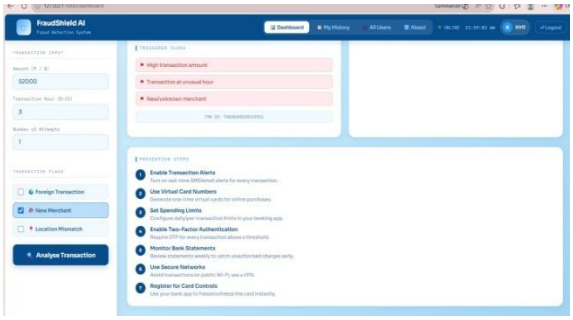


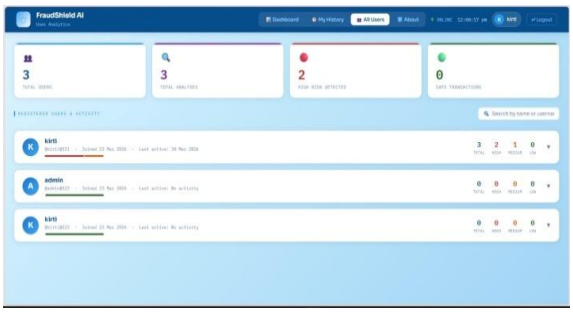












Comparative Table:

Comparative Performance of Fraud Detection Models

Parameter	Logistic Regression	Random Forest	Decision Tree
Accuracy	99.2%	99.8%	98.9%
Precision	92.5%	96.8%	89.4%
Recall	89.7%	94.6%	87.1%
F1-Score	91.1%	95.7%	88.2%
Training Time	12 sec	45 sec	18 sec
Fraud Detection Rate	89.7%	94.6%	87.1%

VII. CONCLUSION

Credit card fraud detection is a critical challenge in modern financial systems. Traditional rule-based methods are insufficient for detecting complex fraud patterns. Machine learning provides an efficient and intelligent solution for fraud detection.

This research presented a machine learning-based approach for detecting credit card fraud using classification algorithms such as Logistic Regression, Decision Tree, and Random Forest. The experimental results indicate that machine learning models can effectively identify fraudulent transactions with high accuracy.

Future research can focus on integrating deep learning techniques and real-time fraud detection systems to further improve the efficiency of fraud prevention mechanisms.

REFERENCES

[1] C. Jiang, K. He, H. Xie, Z. Xiang, and R. M. Lee, "Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3637–3647, 2018.

[2] Pumsirirat and L. Yan, "Credit Card Fraud Detection Using Deep Learning Based on Auto-Encoder and Restricted Boltzmann Machine," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 1, pp. 18–25, 2018.

[3] E. Mohammed and B. Far, "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study," in *Proc. IEEE Int. Conf. Information Reuse and Integration (IRI)*, 2018, pp. 219–226, doi: 10.1109/IRI.2018.00025.

[4] K. Nandi, "Credit Card Fraud Detection Using AdaBoost and Majority Voting," *IEEE Access*, vol. 6, pp. 14277–14284, 2018, doi: 10.1109/ACCESS.2018.2806420.

[5] Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, and P. Beling, "Deep Learning Detecting Fraud in Credit Card Transactions," in *Proc. 2018 Systems and Information Engineering Design Symposium (SIEDS)*, Charlottesville, VA,

USA, 2018, pp. 129–134, doi: 10.1109/SIEDS.2018.8374722.

[6] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and C. Jiang, "Random Forest for Credit Card Fraud Detection," in *Proc. 2018 IEEE 15th Int. Conf. Networking, Sensing and Control (ICNSC)*, Zhuhai, China, 2018, pp. 1–6, doi: 10.1109/ICNSC.2018.8361343.

[7] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit Card Fraud Detection Using Machine Learning Techniques: A Comparative Analysis," in *Proc. 2017 Int. Conf. Computing Networking and Informatics (ICCNI)*, Lagos, Nigeria, 2017, pp. 1–9, doi: 10.1109/ICCNI.2017.8123782.

[8] G. E. Melo-Acosta, F. Duitama-Muñoz, and J. D. Arias-Londoño, "Fraud Detection in Big Data Using Supervised and Semi-Supervised Learning Techniques," in *Proc. 2017 IEEE Colombian Conf. Communications and Computing (COLCOM)*, Cartagena, Colombia, 2017, pp. 1–6, doi: 10.1109/COLCOMCON.2017.8088206.

[9] Reserve Bank of India, Credit Card Circulars, accessed Jun. 11, 2026.

[10] Federal Trade Commission, "Imposter Scams Top Complaints Made to FTC in 2018", accessed Jun. 11, 2026.

[11] Kaggle, "Credit Card Fraud Detection Dataset", accessed Jun. 11, 2026.

[12] Kaggle, "Default of Credit Card Clients Dataset", accessed Jun. 11, 2026.

[13] Kaggle, "PaySim Mobile Money Transactions Dataset", accessed Jun. 11, 2026.