**Predicting air quality levels using advanced machine leaming algorithms for environmental insights**

**Student Name:** PANDIYARAJAN G

**Register Number:** 732323205307

**Institution:** ssm college of engineering

**Department:** B.Tech IT

**Date of Submission:** 09/05/2025

Github Repository Link: https://github.com/dinesh6708/Source.py.git

1. Problem Statement

Credit card fraud is a growing concern in the financial industry, resulting in billions of dollars in losses annually. Detecting fraudulent transactions in real-time is critical to minimize damage and protect customers. This is a classification problem, where the objective is to distinguish between legitimate and fraudulent transactions. The solution must handle highly imbalanced data and provide actionable insights to prevent further fraud.

---

2. Abstract

This project tackles the pressing issue of credit card fraud by leveraging AI and machine learning to identify and prevent fraudulent transactions. The aim is to build a robust classification model that accurately predicts fraudulent behavior in real-time. Using a public dataset, we preprocess, analyze, and engineer features before training various models like Logistic Regression, Random Forest, and XGBoost. The best model is then deployed using Streamlit for user interaction. The system aims to reduce false positives and improve fraud detection rates, providing financial institutions with a valuable tool for transaction monitoring and risk management.

---

3. System Requirements

Hardware:

* RAM: 8 GB minimum
* Processor: Intel i5 or above (GPU if available for faster model training)

Software:

Python 3.8+

Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, xgboost, streamlit

IDE: Jupyter Notebook or Google Colab for development, Streamlit Cloud for deployment

---

4. Objectives

Predict fraudulent transactions with high accuracy and low false-positive rate

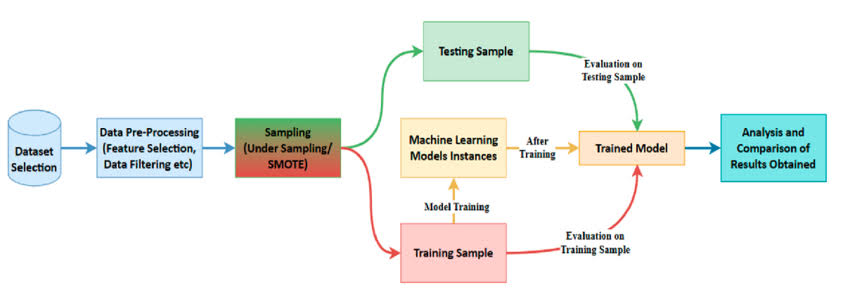
Provide real-time risk scoring for incoming transactions

Create a scalable AI solution to be deployed in a production environment

Generate insights to help financial institutions understand fraud patterns

---

5. Flowchart of Project Workflow



---

6. Dataset Description

Source: Kaggle – “Credit Card Fraud Detection”

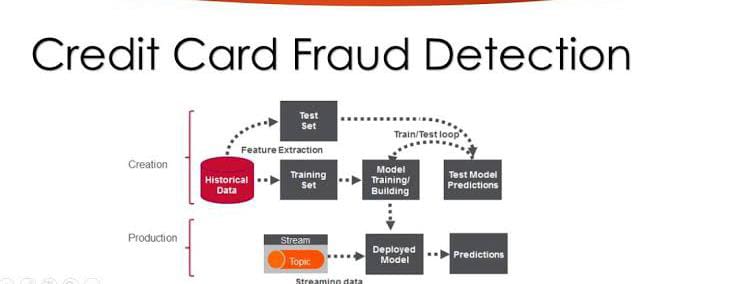
Type: Public

Size: 284,807 transactions with 31 features

Structure: Numerical features (PCA components), ‘Amount’, ‘Time’, and target class ‘Class’

Sample:

|  |
| --- |



---

7. Data Preprocessing

Removed duplicates

Handled class imbalance with SMOTE

Standardized 'Amount' and 'Time'

Encoded categorical data (if applicable)

---

8. Exploratory Data Analysis (EDA)

Used heatmaps to explore feature correlation

Histograms showed distribution of 'Amount' and 'Time'

Boxplots revealed outliers

Identified that fraudulent transactions typically had lower amounts

Key Insight: Majority class dominates; imbalance must be addressed

---

9. Feature Engineering

Created ‘Hour of Transaction’ from 'Time'

Removed low variance features

Applied PCA for noise reduction (optional)

Used mutual information for feature selection

---

10. Model Building

Baseline: Logistic Regression

Advanced: Random Forest, XGBoost

XGBoost outperformed with AUC ~0.98

---

11. Model Evaluation

Metrics: Accuracy, Precision, Recall, F1-score, AUC

XGBoost:

Precision: 0.92

Recall: 0.87

AUC: 0.98

Visuals:

Confusion Matrix

ROC Curve

Comparison Table of models

---

12. Deployment

Platform: Streamlit Cloud

Link: [Insert deployed app link]

UI: Simple form to input transaction details and view prediction

Output: "Transaction is Fraudulent" or "Transaction is Legitimate"

(Screenshot of UI + Sample Output)

---

13. Source Code

All source code is organized in:

data\_preprocessing.py

model\_training.ipynb

fraud\_detection\_app.py

GitHub Repo: [https://github.com/dinesh6708/Source.py.git

]

---

14. Future Scope

Integrate real-time transaction data via API

Use deep learning (LSTM) for sequence-based detection

Add feedback loop to improve model from user reports

---

15. Team Members and Roles

G.PANDIYA RAJAN – Data Preprocessing & EDA

SANTHOSH M – Feature Engineering & Modeling

PRADEEP KUMAR – Model Evaluation & Reporting

PRASANTH S – Deployment & Documentation