Credit Card Fraud Detection Project Report

Date: July 2025

Project Title: Credit Card Fraud Detection using Machine Learning

Team Members:
1. Sai Preetham – Data Analyst & Cloud Deployment
2. Vittal Rao – ML Engineer
3. Lekhana Gowd – Evaluation Lead
4. Sneha – Visualizer
5. Jawahar Reddy – Streamlit UI Developer
6. Abhilash Reddy – Documentation Lead

1. Introduction

This project aims to detect fraudulent credit card transactions using machine learning. Due to the highly imbalanced nature of the dataset, we focused on data preprocessing and class balancing techniques before training models. The final model was deployed as a web application using Streamlit and hosted on AWS EC2.

2. Dataset Description

The dataset used is from Kaggle and contains credit card transaction records. It includes 284,807 transactions, with only 492 labeled as fraud. Features V1 to V28 are anonymized using PCA, along with 'Time', 'Amount', and 'Class' (target variable).

3. Roles and Responsibilities

- Data Analyst: Data loading, cleaning, SMOTE balancing.
- ML Engineer: Model training, evaluation, saving.
- Evaluation Lead: Compared model metrics (accuracy, precision, recall, F1-score).
- Visualizer: Built charts for performance comparison.
- Streamlit UI Developer: UI Design deployment.
- Cloud Deployer: AWS EC2 deployment.
- Documentation Lead: Created slides and this report.

4. Data Analysis and Cleaning

Initial steps involved loading the dataset, checking nulls, and examining class imbalance. The data was split into training and testing sets (80:20). To fix the class imbalance, SMOTE (Synthetic Minority Over-sampling Technique) was applied to the training data.

5. Model Building and Evaluation

Two classification models were trained: Logistic Regression and Random Forest. Random Forest performed better in terms of precision, recall, and F1-score. The model was saved using joblib for use in the web app.

6. Streamlit App and Cloud Deployment

The app was developed using Streamlit and allows users to input transaction data and predict fraud. The model is loaded and used in real-time on user input. Deployment was done on AWS EC2 by transferring files and running the app with open ports for public access.

7. Visualization and Documentation

The Visualizer was responsible for creating a clear and interactive dashboard that presents all the project insights visually. This helps both technical and non-technical viewers understand the results easily. The Documentation Lead compiled and organized the entire project into a structured **report and presentation**. This included updates from all team members and formatted deliverables for evaluation.

8. Challenges Faced

- Handling class imbalance
- Installation issues on EC2
- Streamlit not recognized initially
- SSH and file permission issues during deployment

9. Learnings and Next Steps

We learned the end-to-end process of data analysis, machine learning, and cloud deployment. For future improvements, we plan to add login features, host with a custom domain, and enhance visualizations.

10. Conclusion

This project gave us hands-on experience in solving real-world problems using data science and machine learning technologies. We successfully created and deployed a fraud detection model that can be accessed online.