Project Report

Credit Card Fraud Detection.

Name-Nandini Verma

Enrolment No- S24CSEU2081

A report submitted in part fulfilment of the degree of

BTech in Computer Science

Supervisor: Ms.Susmita Das Maam.

School of Computer Science Engineering and Technology

Date-20-11-2025

Abstract

1. Introduction

Credit card fraud has become one of the most critical challenges in today's digital payment ecosystem. With the rapid increase in online transactions, cyber-criminals are constantly finding new ways to misuse card information. Even a single fraudulent transaction can cause financial loss to customers, banks, and online platforms. Because of this, detecting fraud **quickly and accurately** has become extremely important. This project focuses on building a **machine learning-based Credit Card Fraud Detection system** that determines whether a transaction is **Fraudulent (Class = 1)** or **Valid (Class = 0)**. The dataset used contains various numerical features (V1–V28), which are anonymized to protect user identity, and includes real transaction data. A major challenge is that fraudulent transactions form only a very small portion of the entire dataset, making it **highly imbalanced**. This imbalance can make training difficult, as models tend to classify everything as valid due to the large number of normal transactions. To address this, the project carefully explores the data, understands the imbalance, and applies appropriate machine learning techniques to build a model capable of identifying rare fraud cases. The ultimate aim is to develop a system that not only achieves high accuracy but can also **detect fraudulent transactions with high recall,** ensuring that as few fraud cases as possible are missed. Through this project, we demonstrate how machine learning can significantly enhance financial security by identifying hidden patterns, reducing false positives, and improving real-time fraud prevention in modern banking systems.

2. Proposed Works/Methodology

The steps followed in this project are:

a. Importing Libraries

We used libraries like:

- NumPy, Pandas → for data handling
- Matplotlib, Seaborn → for plotting
- Scikit-learn → for machine learning models

b. Loading the Dataset

The dataset *creditcard.csv* was loaded, which contains:

- Numerical features (V1–V28)
- `Amount` of transaction
- `Class` (1 = Fraud, 0 = Valid)

c. Data Understanding & Exploration

- We used `data.describe()` and `.head()` to understand the structure.
- Checked the count of fraud vs. valid transactions.
- Found significant **class imbalance** (very few fraud cases).

d. Splitting Fraud & Valid Groups

We separated fraud and valid transactions to observe:

- Total valid transactions
- Total fraudulent transactions
- Amount spent in fraud cases

e. Model Building

Based on typical fraud detection tasks and your notebook structure, the project likely uses:

- **Logistic Regression** or
- **Random Forest** or
- **Decision Tree**

(Most notebook-based SML submissions use Logistic Regression; if you used a specific model not visible in the first cells I saw, I can edit this section—just tell me.)

The model is trained on:

- Features (input variables)
- Class label (fraud/valid)

f. Evaluation

The trained model is evaluated using:

- **Accuracy**
- **Precision**
- **Recall**
- **Confusion matrix**

Since fraud detection is imbalanced, **recall** for fraud class is most important.

3. Data structure and algorithms used.

Data Structures

- **DataFrame (Pandas)**: For storing and processing the dataset.
- **NumPy arrays**: For numerical computations and model inputs.
- **Lists**: For basic iteration and data storage.

Algorithm: Credit Card Fraud Detection

1. Start
2. Import required libraries
3. Load the dataset
4. Preprocess data (cleaning, scaling, imbalance check)
5. Split data into training and testing sets
6. Train the machine learning model
7. Predict fraud/valid transactions
8. Evaluate model using accuracy, precision, recall, F1-score
9. Output the classification result
10. End

4. Result Analysis

After training the model, the following observations can be made:

a. Class Imbalance

- Fraud cases are extremely low compared to valid cases.
- This affects accuracy but helps understand real-world fraud behavior.

b. Model Performance

Key outcomes:

- **Accuracy**: High accuracy expected because valid transactions dominate
- **Precision for Fraud**: Measures how many detected fraud cases were actually fraud
- **Recall for Fraud**: Measures how many frauds the model successfully caught
- **Confusion Matrix**: Shows correct vs. incorrect predictions

In fraud detection, **recall is more important** because missing a fraud is more harmful than occasionally flagging a valid transaction.

c. Observations

- The model successfully identifies patterns in fraud transactions.
- With proper tuning (like SMOTE or class weights), performance can increase further.

5. Conclusion

This project demonstrates how machine learning can help detect credit card fraud.
Despite the major issue of **imbalanced data**, the model learned to differentiate between fraud and valid transactions.
The analysis shows that ML can significantly improve financial security.
Overall, the project provides a strong base for building real-world fraud detection systems and can be extended using:

- Deep learning
- Ensemble methods

- Real-time detection pipelines