Credit Card Fraud Detection Using Machine Learning

1. Introduction

Credit card fraud detection is an important real-world problem in the banking and financial sector. The objective of this project is to build a machine learning model that can accurately classify transactions as fraudulent or genuine using historical transaction data.

2. Dataset Description

The dataset used in this project is the Credit Card Fraud Detection dataset. It contains transaction records where most features are anonymized using PCA transformation. The target column 'Class' indicates whether a transaction is fraudulent (1) or genuine (0).

3. Data Preprocessing

- • Converted target column to numeric format and handled missing values.
- • Normalized feature values using StandardScaler.
- • Split the dataset into training and testing sets (80:20 ratio).

4. Handling Class Imbalance

The dataset is highly imbalanced, with very few fraudulent transactions. To handle this issue, SMOTE (Synthetic Minority Over-sampling Technique) was applied to oversample the minority class in the training data.

5. Model Building

Logistic Regression was used as the classification algorithm. It is a simple and effective model for binary classification problems and performs well when combined with proper data preprocessing and balancing techniques.

6. Model Evaluation

The model performance was evaluated using confusion matrix and classification metrics such as precision, recall, and F1-score. These metrics are especially important for imbalanced datasets like fraud detection.

7. Conclusion

The project successfully demonstrates how machine learning techniques can be applied to detect fraudulent credit card transactions. Using data normalization, SMOTE, and logistic regression improved the model’s ability to identify fraud cases effectively.