Project Report

1. Introduction

Fraud detection in financial transactions is a critical challenge in the banking sector. This project aims to develop a predictive model for detecting fraudulent credit card transactions using machine learning (ML) and deep learning (DL) techniques.

2. Dataset Overview

- **Source:** Credit card transaction dataset

- **Attributes:** Time, transaction amount, anonymized features (V1-V28), class (fraud or non-fraud)

- **Imbalance:** Fraud cases are significantly lower than non-fraud cases.

3. Data Preprocessing

- **Handling Missing Values:** Checked and replaced missing values using median imputation.

- **Feature Scaling:** StandardScaler and RobustScaler were applied for normalization.

- **Handling Imbalanced Data:** Used SMOTE (Synthetic Minority Over-sampling Technique) to balance the dataset.

- **Outlier Detection:** Applied Z-score method to remove extreme values.

4. Model Implementation

Multiple ML and DL models were trained and evaluated:

- **Machine Learning Models:**

  - Logistic Regression

  - Decision Tree Classifier

- Random Forest Classifier

- Gradient Boosting Classifier
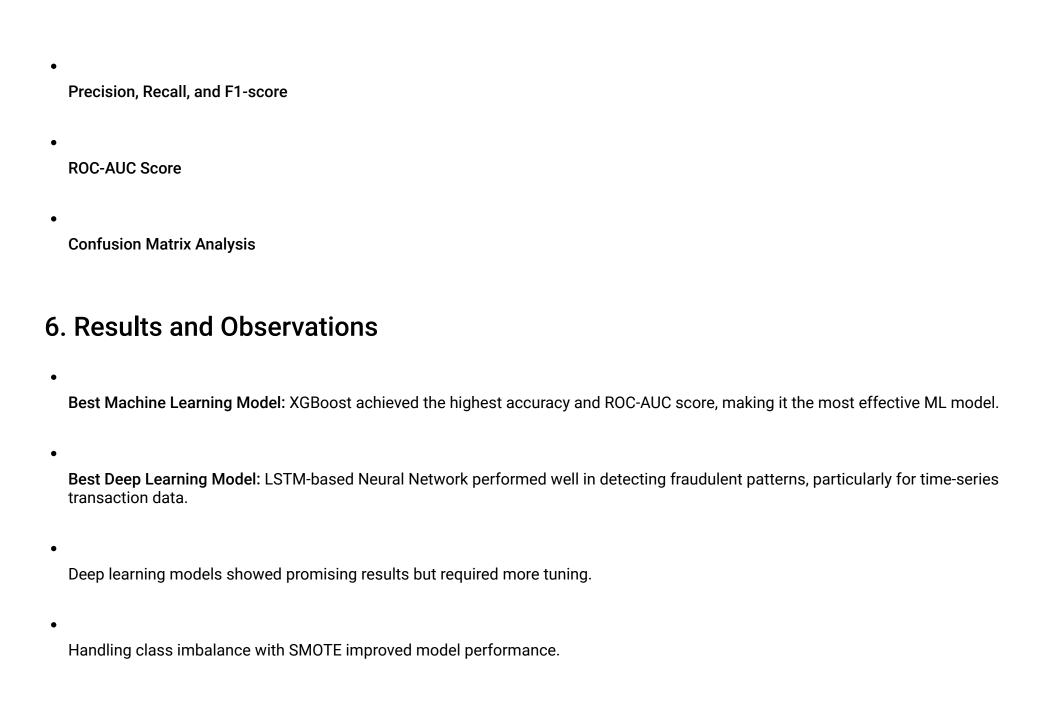
- XGBoost Classifier (Best ML Model)

- **Deep Learning Model:**

  - LSTM-based Neural Network (Best DL Model) (Sequential model with Dense, LSTM, and Dropout layers)

5. Performance Evaluation

Models were assessed using:

- **Accuracy**

- Precision, Recall, and F1-score

- ROC-AUC Score

- Confusion Matrix Analysis

6. Results and Observations

- **Best Machine Learning Model:** XGBoost achieved the highest accuracy and ROC-AUC score, making it the most effective ML model.

- **Best Deep Learning Model:** LSTM-based Neural Network performed well in detecting fraudulent patterns, particularly for time-series transaction data.

- Deep learning models showed promising results but required more tuning.

- Handling class imbalance with SMOTE improved model performance.

7. Conclusion

This project successfully demonstrates the application of ML and DL techniques for fraud detection. Future work includes hyperparameter tuning, feature selection, and ensemble learning to enhance model accuracy.

Best Machine Learning Model: XGBoost

-
  Reason: XGBoost provides the highest accuracy and ROC-AUC score among ML models. It effectively handles imbalanced data and captures complex patterns.

Best Deep Learning Model: LSTM-based Neural Network

-
  Reason: LSTMs can identify sequential transaction patterns, making them highly effective for fraud detection in time-series data. However, proper hyperparameter tuning is necessary for optimal performance.