Credit Card Fraud Detection

SRINIDHI S

Problem Statement

Credit card fraud is a growing concern in the financial industry, leading to billions in losses globally. The challenge lies in detecting fraudulent transactions that occur infrequently among millions of legitimate ones. Manual monitoring is not scalable due to the sheer volume and speed of transactions. This project aims to automate fraud detection using machine learning techniques that can learn from patterns in past data.

Tech Stack

- **Programming Language**: Python
- **Libraries**:
  - ✓ Pandas, NumPy – for data processing
  - ✓ Matplotlib, Seaborn – for EDA and visualization
  - ✓ Scikit-learn – for ML modeling and evaluation
- **Development Environment**: Jupyter Notebook

Solution Approach

1. **Data Preprocessing**:
   - Loaded the anonymized credit card transaction dataset.
   - Detected heavy class imbalance (fraud cases < 1%) and applied resampling strategies.
2. **Exploratory Data Analysis (EDA)**:
   - Visualized feature distributions and class correlations.
   - Identified key differences in patterns between fraudulent and normal transactions.
3. **Class Imbalance Handling**:
   - Used Under-sampling to balance the dataset (reduced majority class).
   - Considered that False Negatives (missing fraud) are more dangerous than False Positives.

4. **ML Algorithms Used**:
    - **Logistic Regression**: Interpretable, used as a baseline model.
    - **Random Forest Classifier**: Captured complex patterns via ensemble learning.
    - **Isolation Forest**: Unsupervised algorithm ideal for detecting outliers and anomalies like fraud.
5. **Evaluation**:
    - Focused on Recall and F1-score since fraud detection needs minimizing false negatives.
    - Compared confusion matrices and ROC-AUC scores across models.