DSCI:6003-03 MACHINE LEARNING PROJECT

TEAM MEMBERS:

1. Veda Prakash Chiliveru (00905759)
2. Pavan Kumar Sunkara (00898969)
3. Vinay Kumar Reddy Punuru (00903155)

Fraudulent Transaction Detection in Financial Systems

Goal: Develop machine learning classification models to identify fraudulent transactions in financial systems. This project will aim to find key factors that distinguish fraudulent transactions from legitimate ones, evaluate the effectiveness of various security measures, and predict the likelihood of fraud in real-time financial operations.

Data: https://www.kaggle.com/datasets/kartik2112/fraud-detection?select=fraudTest.csv

The dataset will consist of transaction records including fields such as transaction ID, amount, merchant details, location, timestamp, user demographics, account history, transaction type (e.g., online, in-person), and whether the transaction was classified as fraudulent. A sample dataset can be sourced from platforms like Kaggle's credit card fraud detection datasets.

Description: The dataset includes both real and fraudulent transaction records, with variables that capture information such as transaction type, amount, location, time, and account details. The dataset is suitable for exploring machine learning classification tasks, anomaly detection, and identifying the characteristics of fraudulent activities. Analyzing this data helps detect fraud patterns and informs better fraud prevention strategies.

Methods:

1. **Exploratory Data Analysis (EDA):**
    - Performing initial data cleaning and analysis to explore key features influencing fraud detection.

- Investigating patterns, correlations, and anomalies in fraudulent vs. legitimate transactions.

2. **Dimensionality Reduction:**
    - We use Principal Component Analysis (PCA) to identify the most important features.
    - Employ Factor Analysis to explore hidden relationships between variables and fraudulent behaviors.
3. **Machine Learning Models:**
    - Build classification models such as:
        - Random Forest Classifier and XGBoost for more complex, high-dimensional data.
        - Support Vector Machines (SVM) for handling non-linear patterns in data.
        - Neural Networks for large datasets with intricate patterns.
4. **Evaluation Metrics:**
    - Use metrics like accuracy, precision, recall, and F1-score to evaluate the model's performance.
    - Focus on reducing false positives and false negatives, particularly in detecting true fraud cases.

Research Contribution: The outcome of this project will help financial institutions and businesses improve fraud detection mechanisms, reduce financial losses, and enhance security. The research can contribute to real-time detection and prevention of fraud, making financial systems safer for consumers and businesses alike. Successful fraud detection will bolster public trust in financial institutions and contribute to a more secure and stable economy.

This project mirrors the structure of the counterfeit currency detection project, focusing on fraud detection in financial systems. It combines similar methods such as PCA, classification models, and addresses the issue of financial security.