Fraud Detection Project Using AI Classification

. This project aims to develop a machine learning system that identifies fraudulent transactions in real-time using a historical dataset of financial transactions for many reasons such as :

Identifying fraudulent transactions early, financial institutions can prevent financial losses associated with fraudulent activities.

Minimizing false positives, legitimate transactions won't be declined, leading to a smoother customer experience.

A robust fraud detection system can deter fraudulent activities and improve overall financial security.

Project Goals:

Build a Robust Model	•Train a model to accurately classify transactions as fraudulent or legitimate based on past data.
Minimize False Positives	•Reduce the number of legitimate transactions flagged as fraudulent to avoid inconveniencing customers.
Maximize Fraud Detection Rate	•Identify a high percentage of actual fraudulent transactions to prevent financial losses.

Database Description:

The CSV database will store information about past transactions, containing features that can potentially indicate fraudulent activities. Here's a possible structure and link:

https://drive.google.com/file/d/1FAn1gAxiWr3Xjt9q-3ttGRRMVGZs6Rvm/view?usp=drive_link

Transaction ID	• Unique identifier for each transaction.
Amount	• Transaction amount in a specific currency.
Time	• Date and time of the transaction (including timestamp for high-resolution analysis).
Cardholder Name	• Name of the cardholder associated with the transaction (consider anonymization for privacy).
Merchant	• Name of the merchant where the transaction occurred.
Card Number (Hashed)	• Hashed version of the card number for security purposes. (Never store the full card number!)
Transition Number	• Transaction Number
category	• Category of Merchant
Label	• This column indicates whether the transaction was legitimate (0) or fraudulent (1).

Project Requirements:

Data Preprocessing

- Clean and prepare the data in the CSV. This might involve handling missing values, converting categorical data (e.g., country codes), and feature scaling.

Feature Engineering

- Create new features from existing ones that might be more informative for fraud detection (e.g., difference between current and average transaction amount, time difference between previous transactions).

Model Selection and Training

Choose a suitable machine learning model for sentiment classification, such as:

- **Logistic Regression:** Good baseline model, interpretable results.
- **Random Forest:** Handles imbalanced data well, robust to outliers.
- **Support Vector Machines (SVM):** Effective for high-dimensional data.
- **XGBoost:** Powerful ensemble method, may require hyperparameter tuning. Train the model on the labeled data in the CSV, splitting it into training and testing sets (consider stratified sampling to maintain class balance).

Model Evaluation

Evaluate the performance of the model on the testing set using metrics like:

- **Accuracy:** Overall percentage of correctly classified sentiment labels.
- **Precision:** Ratio of true positives to all positive predictions (reduces false positives).
- **Recall:** Ratio of true positives to all actual positive examples (reduces false negatives).
- **F1-score:** Harmonic mean of precision and recall, provides a balanced view.

Model Tuning

If the initial model performance is not satisfactory, you can fine-tune hyperparameters of the model or explore different model architectures.

Deployment and Monitoring:

Once satisfied with the model's performance, deploy it into a production environment. Continuously monitor the model's performance and retrain it periodically with new data to adapt to evolving fraud

Mentor

TA. Andrew Magdy

Email: Andrew.magdy@cis.asu.edu.eg