Fraud Detection Project

# Project Title

Develop and Test a Generative AI-Powered Fraud Detection System

# Overview

This project demonstrates the application of generative AI concepts to a real-world fraud detection problem in the financial domain. The objective was to simulate a fraud detection pipeline from synthetic data generation to model training, integration, and reporting.

# Tools & Technologies Used

- Python

- Google Colab

- pandas - for data manipulation

- scikit-learn - for machine learning

- SMOTE (imbalanced-learn) - for handling imbalanced data

- Logistic Regression - for model training

# Dataset Description

A synthetic dataset was generated to mimic real-world financial transactions with fields such as TransactionAmount, BankName, BIN, Location, DeviceID, Timestamp, and IsFraud. Fraud patterns included reuse of devices and larger transaction amounts.

# Model Training

- LabelEncoded categorical data

- Extracted Hour and DayOfWeek from Timestamp

- Split data into training and testing sets

- Used Logistic Regression model

- Applied SMOTE to balance the training data

# Compliance Workflow Integration

Developed a script that:

1. Takes a new transaction

2. Uses the trained model to classify it

3. Flags suspicious transactions

4. Generates a report summarizing flagged entries

# Testing & Optimization

- Improved recall using SMOTE

- Evaluated performance on test data using classification metrics

# Challenges

- Imbalanced dataset (handled with SMOTE)

- Creating realistic fraud patterns

- Effective use of time-based features

# Potential Improvements

- Use of advanced models like XGBoost

- Add more complex fraud scenarios

- Unsupervised learning for anomaly detection

- Real-time monitoring dashboard

# Final Deliverables

- Python scripts for preprocessing, training, and integration

- Synthetic dataset: synthetic\_fraud\_transactions\_5000.csv

- Trained model (Logistic Regression with SMOTE)

- Evaluation reports

- Portfolio documentation (this file)

# Community Sharing

This project will be shared via Google Drive/GitHub with peer feedback and engagement.

# Summary

Demonstrates generative AI application to fraud detection through data engineering, model training, and integration in a compliance setting.