

Financial Risk Prediction and Fraud Detection Application

Overview

This application is designed to detect fraudulent transactions and assess financial risk using machine learning techniques. It leverages various models to predict fraudulent activity based on transactional data and provides in-depth exploratory data analysis (EDA) to derive insights for decision-making. Additionally, users can interact with an AI-powered expert (Gemma 2B) for financial insights or advice.

Architecture

The application is built using **Streamlit**, a Python framework for creating interactive web apps, combined with several machine learning models and libraries for data processing, model training, and evaluation. The key components are:

1. Data Loading and Preprocessing:

- The dataset (`synthetic_financial_data.csv`) is loaded and preprocessed to clean and handle missing values, ensuring it is ready for model training.
- Non-numeric and unnecessary columns are dropped, and missing numeric data is replaced by the column mean.

2. Model Training:

- Multiple machine learning algorithms are employed, including:
 - Logistic Regression
 - Support Vector Machine (SVM)
 - K-Nearest Neighbors (KNN)
 - Decision Tree
 - Random Forest
- A grid search for hyperparameter tuning is used to identify the best performing model based on accuracy and other metrics.

3. Class Imbalance Handling:

- **SMOTE (Synthetic Minority Over-sampling Technique)** is applied to balance the classes, improving model performance on imbalanced datasets.

4. Model Evaluation:

- Models are evaluated using the following metrics: Accuracy, Precision, Recall, F1 Score, and ROC AUC.
- A **GridSearchCV** is used to tune hyperparameters for optimal model selection.

5. Fraud Detection and Risk Assessment:

- After training, users can input transaction data to get a prediction on whether the transaction is fraudulent or legitimate.
- For credit risk assessment, users can provide financial details, and the model will predict the risk based on the trained classifier.

6. Exploratory Data Analysis (EDA):

- A variety of EDA charts are provided to give insights into the data, such as:
 - Correlation heatmaps
 - Class distribution
 - Transaction amount distribution (Fraud vs Non-Fraud)
 - Boxplots for key features like credit score

7. AI-Powered Financial Insights:

- Users can ask financial questions to an AI model (**Gemma 2B** via **Ollama API**). The AI expert provides real-time, context-specific financial advice based on a conversation history.

Technologies Used

- **Streamlit**: For creating an interactive web interface.

- **Pandas & NumPy:** For data manipulation and preprocessing.
- **Matplotlib & Seaborn:** For data visualization.
- **Scikit-learn:** For machine learning model training, evaluation, and hyperparameter tuning.
- **SMOTE (from imbalanced-learn):** For handling imbalanced datasets.
- **Ollama API:** For generating financial insights via the Gemma 2B model.

Features

1. Interactive Data Analysis:

- Users can explore insights related to fraudulent transaction patterns.
- Visualize class distribution, correlation matrices, and financial patterns.

2. Fraud Detection:

- Users can input specific transaction details to get predictions (fraudulent or legitimate).
- Automatically selects the best performing model based on training data.

3. Financial Risk Assessment:

- Credit risk prediction using machine learning models.
- Transaction details are used to assess whether a lending decision should proceed.

4. Expert Financial Advice (Ollama - Gemma 2B):

- Users can ask specific financial questions (e.g., portfolio optimization, risk management, etc.) and receive expert responses from an AI model.

5. Model Performance Insights:

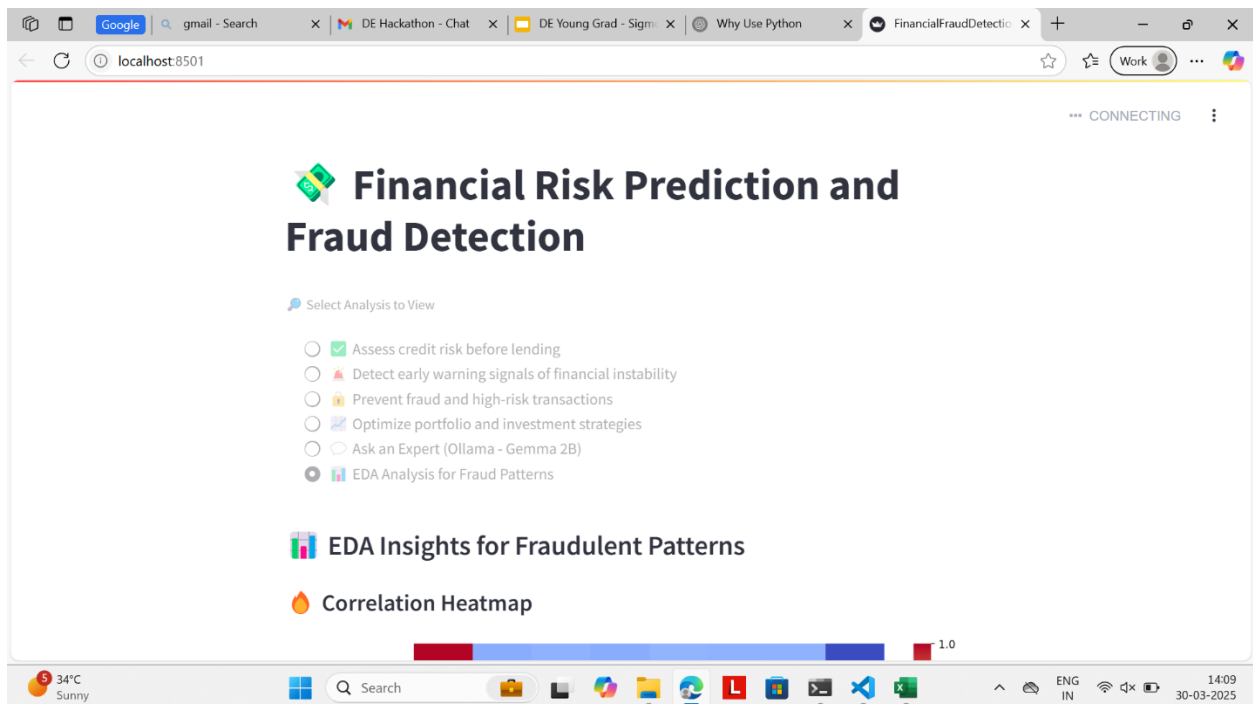
- Displays detailed evaluation results of different models, including accuracy, precision, recall, F1 score, and ROC AUC.

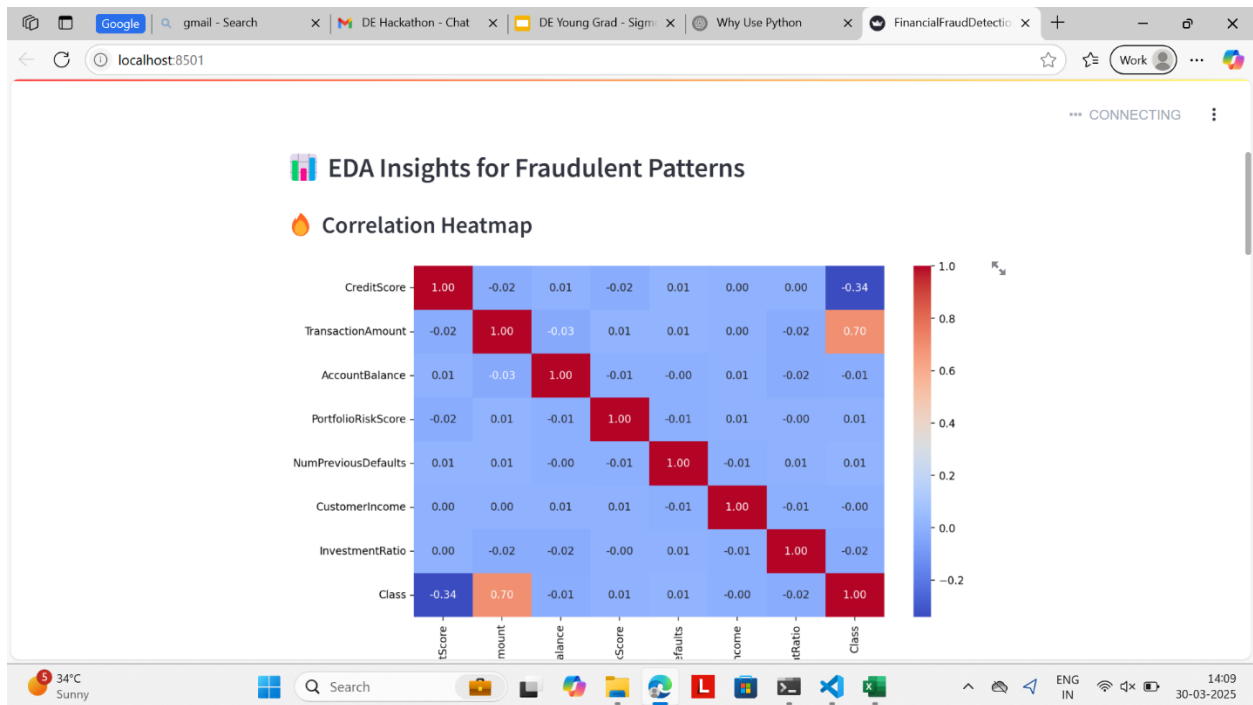
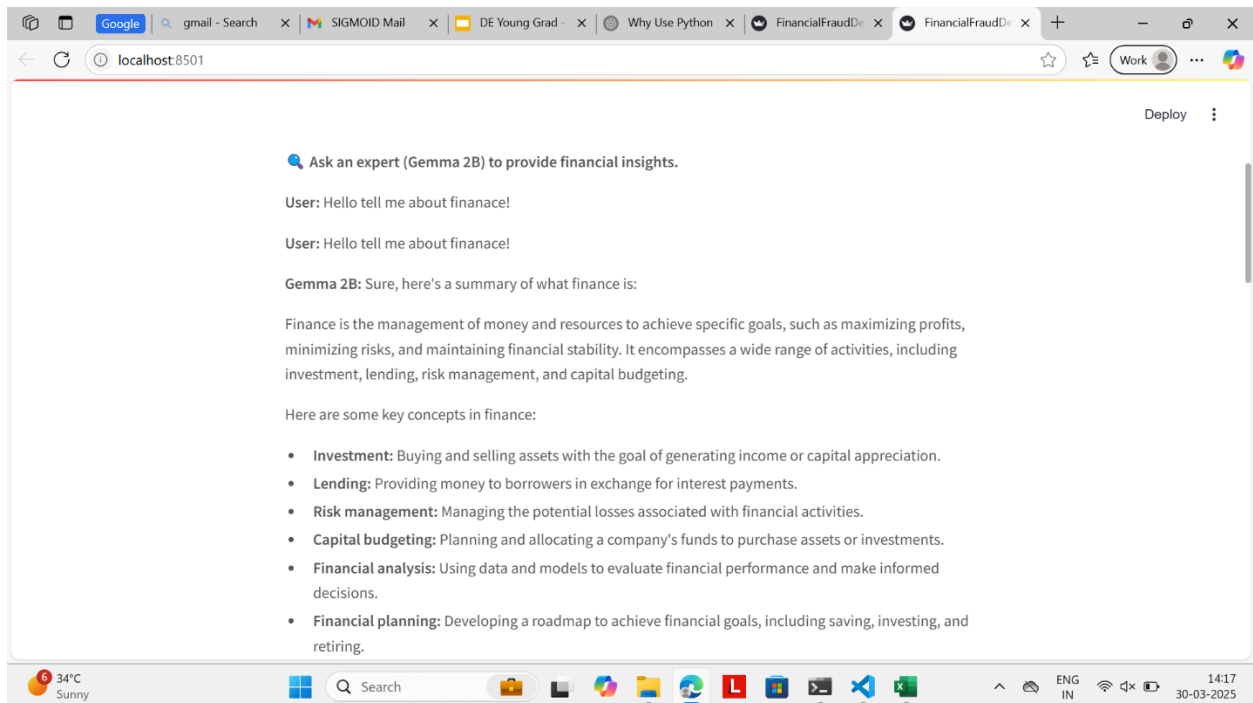
Results and Model Performance

- **Best Model:** The model with the highest accuracy after tuning is automatically selected and used for fraud detection and risk assessment.
- **Evaluation Metrics:** Key metrics like accuracy, precision, recall, F1 score, and ROC AUC are displayed for each model.
- **Insights:**
 - **Fraud Detection:** Higher transaction amounts and low credit scores are strong indicators of fraudulent transactions.
 - **Risk Assessment:** The trained models help in assessing credit risk and whether a transaction should be approved or flagged.

Conclusion

This application provides a comprehensive solution for financial institutions and businesses to detect fraudulent transactions and assess credit risk. It combines traditional machine learning techniques with interactive data analysis and real-time AI-powered insights, offering a robust tool for financial decision-making.



Ask an expert (Gemma 2B) to provide financial insights.

User: Hello tell me about finance!

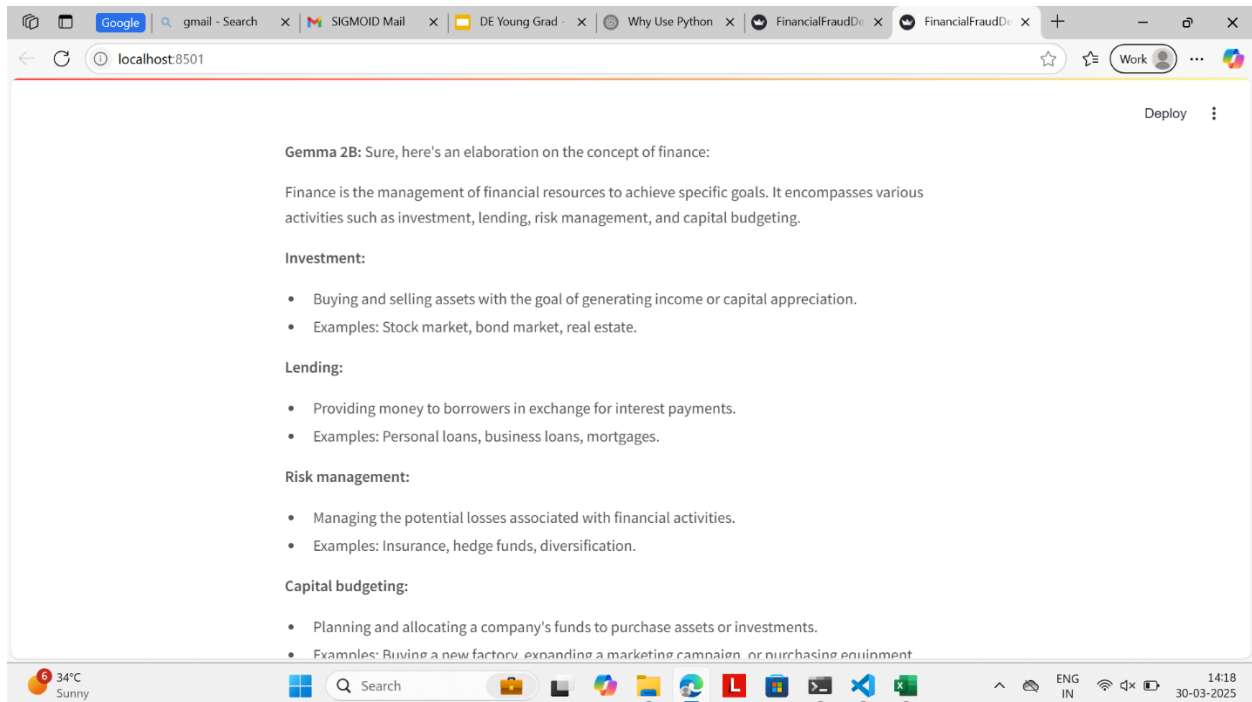
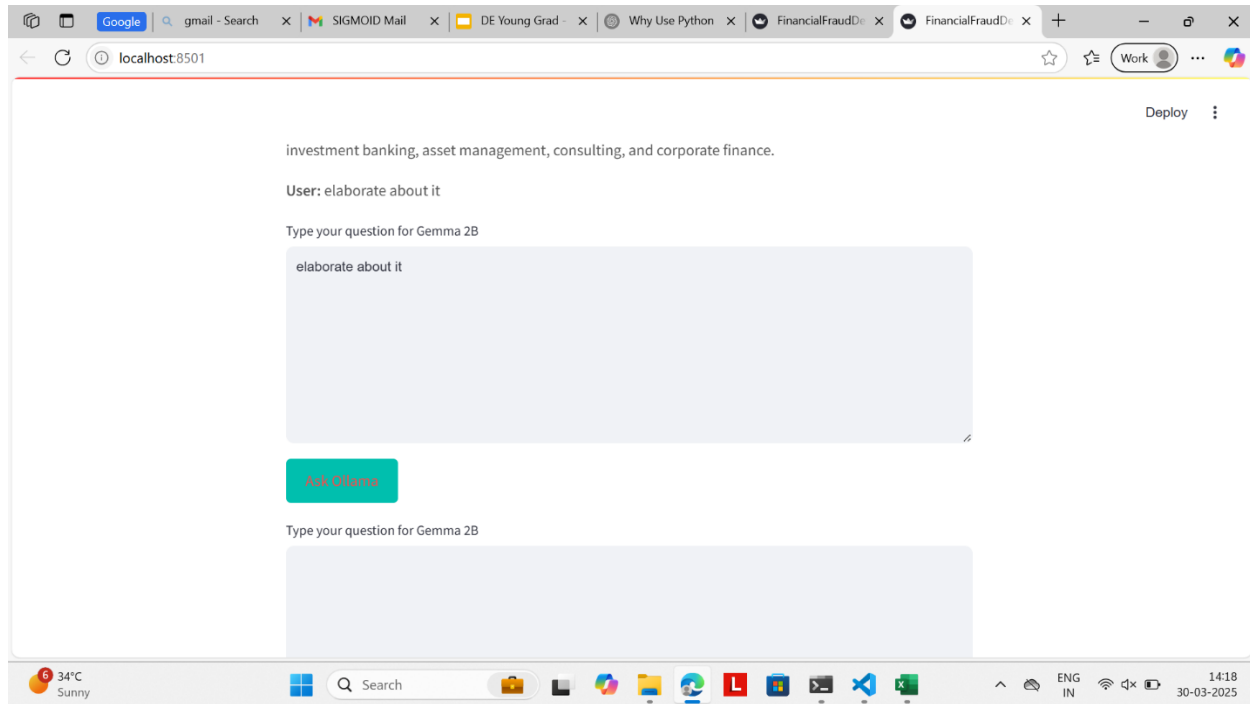
User: Hello tell me about finance!

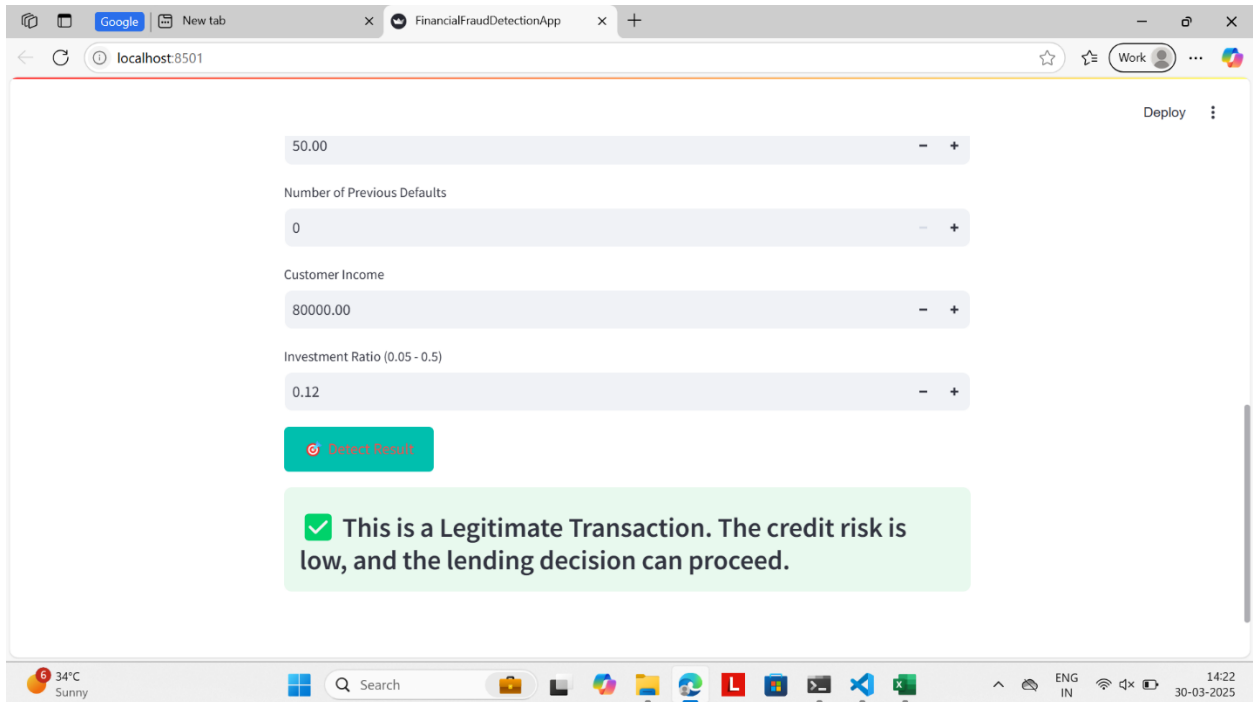
Gemma 2B: Sure, here's a summary of what finance is:

Finance is the management of money and resources to achieve specific goals, such as maximizing profits, minimizing risks, and maintaining financial stability. It encompasses a wide range of activities, including investment, lending, risk management, and capital budgeting.

Here are some key concepts in finance:

- Investment:** Buying and selling assets with the goal of generating income or capital appreciation.
- Lending:** Providing money to borrowers in exchange for interest payments.
- Risk management:** Managing the potential losses associated with financial activities.
- Capital budgeting:** Planning and allocating a company's funds to purchase assets or investments.
- Financial analysis:** Using data and models to evaluate financial performance and make informed decisions.
- Financial planning:** Developing a roadmap to achieve financial goals, including saving, investing, and retiring.



A screenshot of a web browser window showing a financial fraud detection application. The browser has two tabs: "Google" and "FinancialFraudDetectionApp". The address bar shows "localhost:8501". The application interface includes four input fields with labels and values, a "Detect Result" button, and a green feedback message.

50.00

Number of Previous Defaults

0

Customer Income

80000.00

Investment Ratio (0.05 - 0.5)

0.12

Detect Result

✓ This is a Legitimate Transaction. The credit risk is low, and the lending decision can proceed.

34°C Sunny

Search

ENG IN

14:22 30-03-2025