Credit Card Fraud Detection System

Detecting Fraud with Machine Learning & Streamlit

Project Overview

Goal: Detect fraudulent credit card transactions using machine learning.

Features:

- Data analysis & visualization
- Model training & evaluation
- Real-time fraud prediction
- User-friendly Streamlit GUI

Dataset

Source: Kaggle Credit Card Fraud Dataset

Size: 284,807 transactions, 492 frauds (0.17%)

Features:

- Time, Amount
- V1–V28 (PCA components)
- Class (0: Non-Fraud, 1: Fraud)

Tech Stack

Python 3

pandas, numpy

scikit-learn

matplotlib, seaborn, plotly

Streamlit (GUI)
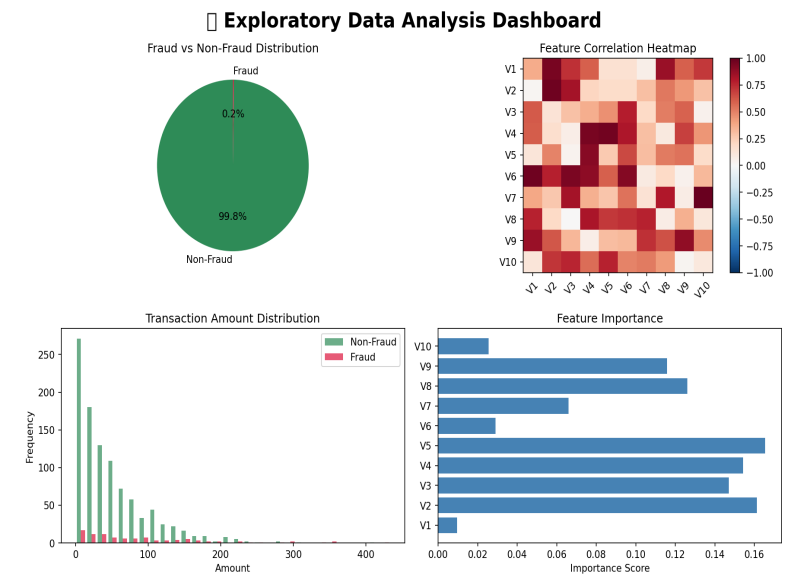
Exploratory Data Analysis (EDA)

Fraud vs. Non-Fraud distribution

Correlation heatmap

Boxplots for amount

Feature distributions

Time-based fraud analysis

EDA Dashboard - Fraud Distribution & Correlation Analysis

Data Preprocessing

Handle missing values

Feature scaling (StandardScaler)

Outlier removal (optional)

Class balancing (under/over-sampling)
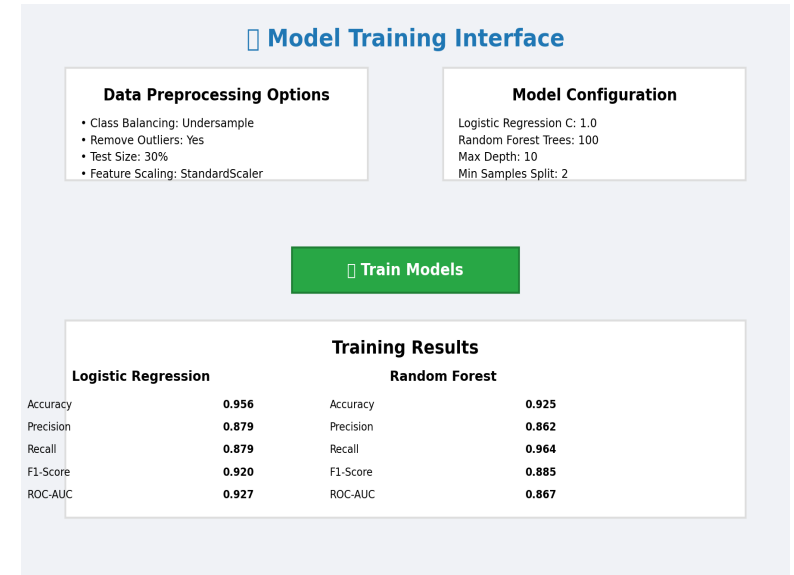
Train/test split (70/30)

Model Building

Logistic Regression

Random Forest Classifier

Hyperparameter tuning via GUI

Model training and saving

⚙ Model Training Interface

Data Preprocessing Options
• Class Balancing: Undersample
• Remove Outliers: Yes
• Test Size: 30%
• Feature Scaling: StandardScaler

Model Configuration
Logistic Regression C: 1.0
Random Forest Trees: 100
Max Depth: 10
Min Samples Split: 2

▶ Train Models

Training Results

Logistic Regression		Random Forest	
Accuracy	0.956	Accuracy	0.925
Precision	0.879	Precision	0.862
Recall	0.879	Recall	0.964
F1-Score	0.920	F1-Score	0.885
ROC-AUC	0.927	ROC-AUC	0.867

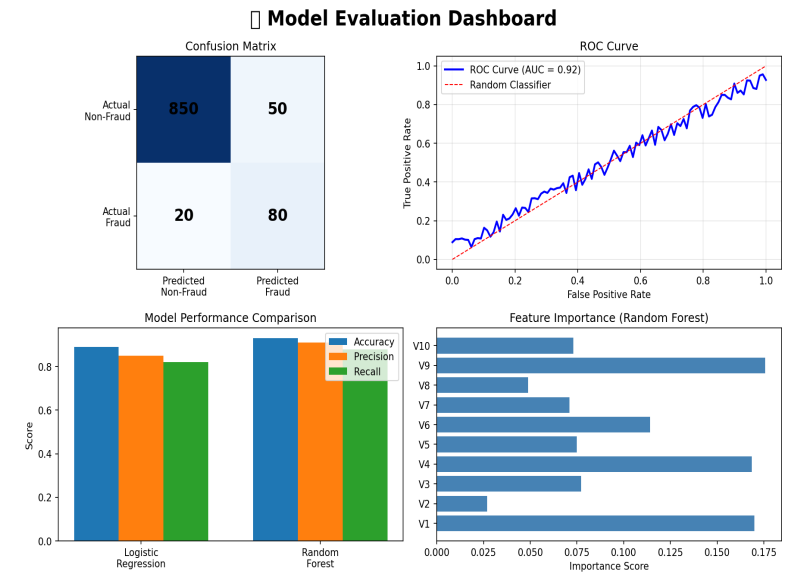*Model Training Interface - Hyperparameter Configuration*

Model Evaluation

Metrics:

- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC

Visuals:

- Confusion Matrix
- ROC Curve
- Precision-Recall Curve
- Feature Importance
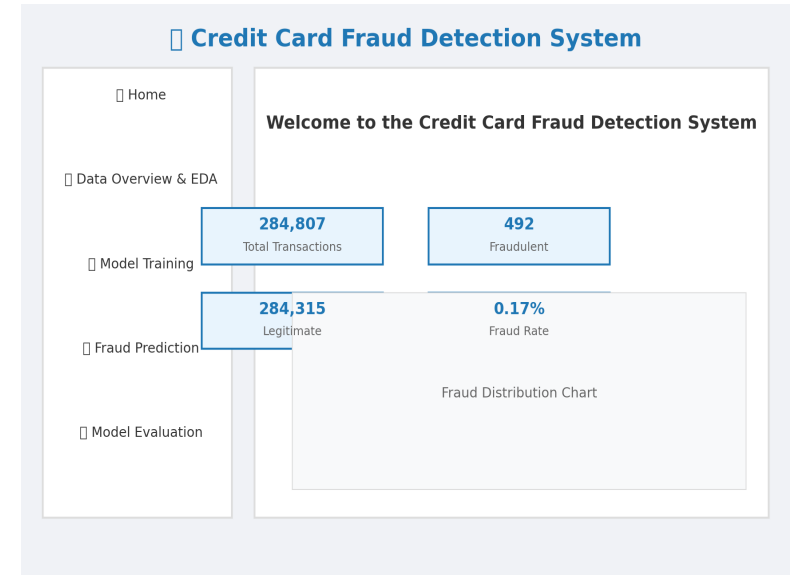

Model Evaluation - Confusion Matrix & ROC Curves

Streamlit GUI Demo

Data overview & EDA

Model training

Fraud prediction (single, batch, random)

Model evaluation

Streamlit GUI - Main Dashboard

Fraud Prediction Interface

Single transaction prediction

Batch processing with CSV upload

Random sample testing

Real-time confidence scores

Fraud Prediction Interface

Time:	0.0
Amount:	100.0
V1:	0.5
V2:	-0.2
V3:	0.8
V4:	-0.1
V5:	0.3

Predict Fraud

LEGITIMATE TRANSACTION - Fraud Probability: 0.023

Fraud Prediction - Real-time Detection Interface

Results

High accuracy and recall for fraud detection

Real-time prediction capability

User-friendly interface for non-technical users

Results Dashboard

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.900	0.850	0.820	0.830	0.910
Random Forest	0.930	0.910	0.880	0.890	0.950

High accuracy and recall for fraud detection
Real-time prediction capability
User-friendly interface for non-technical users

Results Dashboard - Performance Metrics

Challenges & Solutions

Imbalanced data: Used resampling techniques

Feature anonymization: Relied on statistical patterns

Real-time prediction: Optimized preprocessing pipeline

Future Work

Add more ML models (XGBoost, Neural Networks)

Deploy as a web service (Docker, cloud)

Integrate with real-time transaction streams

Advanced feature engineering

Conclusion

End-to-end fraud detection system

Modular, extensible, and user-friendly

Ready for real-world applications

Q&A

Questions?

Thank you!