

Credit Card Fraud Detection



youssef moharm

Bassma ashraf

Mostafa kassab

FraudGuard: Credit Card Fraud Detection System

FraudGuard is a robust Streamlit-based web application designed to detect fraudulent credit card transactions using supervised machine learning algorithms. It provides a user-friendly interface for data upload, exploration, model training, evaluation, and prediction. The app is equipped with user authentication, interactive visualizations, and support for handling imbalanced datasets.

Authentication System

- Users can **register** or **login** using a secure password hashed with SHA-256.
- User credentials are stored in a local users.json file.
- Session state management ensures persistent user sessions and secure access to app functionalities.

Data Handling

- Supports **CSV**, **XLS**, and **XLSX** file formats.
- Uploaded datasets are stored locally and loaded using pandas.
- Missing and duplicate values are automatically removed during preprocessing.
- Users select the **target column** (e.g., 'Class' indicating fraud/non-fraud).

Data Exploration

- Displays summary statistics using pandas.describe().
- Provides target class distribution charts for fraud vs. non-fraud.
- Correlation heatmap (using seaborn) helps identify feature relationships.

Data Preprocessing & Balancing

- Implements basic preprocessing:
 - Null & duplicate removal
 - Target/feature separation
- Supports **undersampling** (RandomUnderSampler) and **oversampling** (SMOTE) to address class imbalance.
- Applies StandardScaler to normalize feature values for model input.

Machine Learning Models

Trains and compares multiple models:

- **Logistic Regression**
- **Random Forest**
- **K-Nearest Neighbors**

Each model is evaluated using:

- Accuracy
- F1 Score
- ROC-AUC
- Precision-Recall Curve
- Confusion Matrix
- Feature Importance (for tree-based models)

Trained model results are stored in session state and visualized via matplotlib/seaborn plots.

Model Evaluation

- Plots **ROC Curves**, **Precision-Recall Curves**, and **Feature Importances**.
- Displays **classification reports** and **confusion matrices**.
- Automatically highlights the best model based on F1 Score.

Live Predictions

- Users can input feature values manually via interactive widgets.
- Uses the selected model to predict fraud risk for a new transaction.
- Shows classification result and fraud probability.

Technologies Used

- **Python 3.x**
 - **Streamlit** (UI/UX)
 - **Scikit-learn** (ML algorithms and evaluation)
 - **Imbalanced-learn** (SMOTE & RandomUnderSampler)
 - **Matplotlib & Seaborn** (visualization)
 - **Pandas & NumPy** (data manipulation)
 - **Joblib** (optional for model persistence)
-

Key Features

- Secure multi-user access with login & registration
 - Customizable preprocessing pipeline
 - Interactive visualization and analytics
 - Real-time predictions using trained models
 - Modular and extensible design
-

Future Enhancements

- Add model persistence and loading (joblib)
- Enable more ML algorithms (e.g., XGBoost, SVM)
- Integrate email alerts for detected fraud
- Deploy on cloud (e.g., Streamlit Cloud, Heroku)