Credit Risk Prediction App

This project is a machine learning-powered web application built using Streamlit that predicts the credit risk (Good/Bad) of loan applicants based on financial and demographic inputs. The model is trained on the German Credit Dataset.

Features

Outlier treatment and data preprocessing

Feature selection using Mutual Information and RFE

Trained models: Decision Tree, KNN, and XGBoost

ROC AUC Curve, Mutual Information, and Correlation heatmap visualizations

Performance evaluation and best model selection

Streamlit-based interactive web UI for predictions

Best model and scaler saved with .pkl files for future use

Quick Start

1. Clone the repository

git clone https://github.com/Harinideepa22/Credit-Risk-Prediction.git

cd Credit-Risk-Prediction

2. Install dependencies

We recommend using a virtual environment:

pip install -r requirements.txt

If you don't have a requirements.txt, install key libraries manually:

pip install streamlit pandas numpy scikit-learn xgboost matplotlib seaborn joblib

3. Prepare data and train the model

Make sure german_credit_data.csv is placed in the data/ folder.

Run the model pipeline:

python model.py

This will:

Preprocess the dataset

Train DecisionTree, KNN, and XGBoost

Evaluate models and choose the best one

Save the best model and scaler in the model/ directory

Generate visualizations in the outputs/ directory

4. Launch the Streamlit App

streamlit run app.py

This opens a browser interface where you can input applicant details and get predictions.

Project Structure

```
Credit-Risk-Prediction/
│
├── data/
│     └── german_credit_data.csv
├── model/
│     ├── best_model_XGBoost.pkl
│     ├── scaler.pkl
│     ├── features.pkl
│     └── label_encoders.pkl
```

```
|
├── outputs/
|   ├── roc_auc_comparison.png
|   ├── mutual_information.png
|   └── correlation_matrix.png
|
├── app.py              # Streamlit application
├── model.py            # Data preprocessing, training and evaluation
├── requirements.txt    # Python dependencies (optional)
└── README.md
```

Input Features

The app uses a subset of important features:

Credit amount

Duration

Sex

Housing

Checking account

These are selected using Recursive Feature Elimination (RFE).

Output

Prediction: Good or Bad Credit Risk

Confidence Score: Probability of good credit risk

Visuals: ROC AUC Curve, Mutual Information Scores, Feature Correlation Heatmap

Models Used

Model GridSearchCV Optimized

Decision Tree

KNN

XGBoost   (Best by default)

The best model is selected based on F1 Score and saved for use in the app.

Evaluation Metrics

Accuracy Score

F1 Score

ROC AUC

Confusion Matrix

Classification Report