Loan Status Prediction Project

Overview

This project aims to predict the loan status (Approved or Rejected) for applicants using **Machine Learning** techniques. The solution is implemented using **KMeans Clustering**, and the user interface is created with **Streamlit** for interactivity.

Project Components

1. Dataset

The dataset contains information about loan applicants, including:

- ApplicantIncome: Monthly income of the applicant.
- LoanAmount: Loan amount requested.
- CreditHistory: Credit history status (1 = Good, 0 = Bad).
- Loan_Status: Target variable (Approved or Rejected).

A synthetic dataset (loan_data.csv) is used for this project.

2. Machine Learning Model

Steps:

1. **Preprocessing**:
    - Encoded the Loan_Status column.
    - Scaled numeric features (ApplicantIncome, LoanAmount, CreditHistory) using **StandardScaler**.
2. **KMeans Clustering**:
    - Trained a KMeans model with 2 clusters (Approved/Rejected).
    - Evaluated clustering performance using **Silhouette Score**.
3. **Pickle File**:
    - Saved the trained model in loan_kmeans_model.pkl for reuse in the Streamlit app.

3. Exploratory Data Analysis (EDA)

Key insights were derived using **Seaborn** and **Matplotlib** visualizations:

- Loan status distribution (Countplot).
- Relationship between income, loan amount, and loan status (Boxplots).
- Correlation heatmap.

4. Streamlit Application

The interactive app includes:

- Sidebar input controls for applicant details.
- Dynamic visualizations (countplot, scatterplot).
- Loan status prediction based on model output.

Requirements

The project requires the following Python libraries:

streamlit==1.25.0

numpy==1.24.4

pandas==1.5.3

scikit-learn==1.2.2

matplotlib==3.7.2

seaborn==0.12.2

Installation Steps

1. **Clone the repository** or create a project folder.
2. **Set up a virtual environment**:

   python -m venv loan_env

   source loan_env/bin/activate # For Mac/Linux

   loan_env\Scripts\activate # For Windows
3. **Install dependencies**:

   pip install -r requirements.txt
4. **Run the Streamlit app**:

   streamlit run app.py

Code Structure

1. app.py

The Streamlit application for user interaction and prediction.

2. loan_kmeans_model.pkl

Saved KMeans model for loan status prediction.

3. loan_data.csv

Synthetic dataset used for training and testing.

Streamlit Application Features

1. **User Input**:
    - Enter applicant details using sliders and dropdowns.
2. **Prediction**:
    - Displays whether the loan is "Approved" or "Rejected."
3. **Visualizations**:
    - Loan status distribution.
    - Income vs. Loan Amount grouped by status.
4. **Dataset Preview**:
    - Option to view the raw dataset in the app.

Example Workflow

1. Open the app by running the command: streamlit run app.py.
2. Enter applicant details (e.g., Income: 5000, Loan Amount: 200, Credit History: 1).
3. Click "Predict Loan Status."
4. View the predicted result and insights through visualizations.

Evaluation

- **Silhouette Score**: Measures the quality of clustering.
- **Classification Report**: Maps predicted clusters to loan status labels and evaluates model performance.

Future Enhancements

1. Use a more robust supervised learning model (e.g., Logistic Regression, Random Forest).
2. Include more features like employment type, loan term, and co-applicant details.
3. Deploy the app using **Streamlit Cloud** or **Heroku**.
