**Loan Prediction Model**

**Overview**

This project aims to predict whether a loan application will be approved or rejected based on various features provided in the dataset. The dataset includes information about applicants' demographics, financial status, credit history, and other factors.

**Dataset**

The dataset was downloaded from Kaggle in csv and contains the following fields:

* loan\_id: Unique identifier for each loan application
* gender: Gender of the applicant (Male/Female)
* married: Marital status (Yes/No)
* dependents: Number of dependents
* education: Applicant's education level (Graduate/Not Graduate)
* self\_employed: Self-employment status (Yes/No)
* applicant\_income: Applicant's income in pounds
* coapplicant\_income: Co-applicant's income in pounds
* loan\_amount: Loan amount in thousands
* loan\_amount\_term: Term of the loan in months
* credit\_history: Credit history (0 - No history, 1 - Has history)
* property\_area: Area type (Rural/Urban/Semiurban)
* loan\_status: Target variable (0 - Approved, 1 - Rejected)

**Approach**

Exploratory data analysis: Loan\_id field was removed from the dataset. Pair plots were created to compare the integer values, with the separation loan\_status. Correlation matrix displayed the correlation among the integer variables. Count plots created for object variables.

Modelling: Model using loan\_amount and loan\_status, test size = 0.3, logistic regression model instantiated.

Feature engineering convert categorical features to binary, logistic regression model instantiated again. The most important features were displayed on a bar plot.

Decision tree classifier: Instantiate a Decision Tree classifier and hyperparameter tuning performed.

Data Preprocessing: Handling missing values, encoding categorical variables, scaling numerical features, and splitting the data.

Model Selection: Trying different classification algorithms (e.g., Logistic Regression, Decision Trees, SVM, KNN) to find the best model.

Training and Evaluation: Training the selected model(s) on the training set, making predictions on the test set, and evaluating performance using metrics like accuracy, precision, recall, and F1-score.

Model Deployment: Deploying the best-performing model for future loan approval predictions.

**Files Included**

loans.csv: The dataset used for training and testing the models.

MM\_LoanAccept.ipynb: Jupyter Notebook containing code for data preprocessing, model training, and evaluation.

requirements.txt: List of Python dependencies needed to run the code.

**Usage**

Setup: Install the required dependencies using pip install -r requirements.txt.

Run the Notebook: Execute the cells in loan\_prediction.ipynb to preprocess data, train models, and evaluate their performance.

**Results**

Model Performance: Detailed analysis and performance metrics of different models tested on the dataset.

Insights: Insights derived from the analysis that can be helpful for loan approval decision-making.

**Conclusion**

This project explores various machine learning models to predict loan approval or rejection based on applicant information. It provides a framework for understanding how different factors influence loan approval decisions.