📄 Loan Default Prediction System – Summary

📌 Project Goal:

The objective of this project is to develop a machine learning model that predicts whether a loan applicant is likely to default or repay the loan.

By analyzing historical data with features such as applicant income, loan amount, credit history, and property area, this system aims to support financial institutions in making more informed and automated lending decisions.

✅ Chosen Model:

Among the models tested — Logistic Regression, Decision Tree, and Random Forest — the Random Forest Classifier was chosen for its superior performance in terms of F1-score and ROC AUC.

🎯 Final Accuracy:

The Random Forest model achieved an accuracy of approximately 82% on the test dataset, demonstrating strong capability in correctly classifying applicants.

📊 Insights from EDA:

- Most applicants with a **Credit_History of 1.0** were more likely to get loan approval.
- The distributions of **ApplicantIncome** and **LoanAmount** were **right-skewed**, indicating some very high-value outliers.
- **Urban areas** had a slightly higher proportion of approved loans compared to rural areas.
- **Graduate applicants** tended to have better approval rates.
- The **correlation heatmap** showed moderate correlation between **income** and **loan amount**, but no severe multicollinearity.