■ Problem Statement — Loan Approval Prediction Web App

Objective: Build a machine learning model to predict whether a loan application will be approved. Use historical loan data to train and test your model, then deploy it as a simple Streamlit or Flask web app where users can enter applicant details and get a real-time prediction.

Dataset: Kaggle — Loan Approval Prediction Dataset

1. Exploratory Data Analysis (EDA)

- What is the shape of the dataset? Any missing values?
- What is the class distribution of loan status (Approved vs Not Approved)?
- Summary statistics of key numeric features (Income, LoanAmount).
- Relationship of Credit_History and Loan_Status.
- Which categorical features (Gender, Education, Property_Area) seem to influence loan approval?

2. Machine Learning

- Preprocess: handle missing values, encode categoricals, scale if needed.
- Train at least two models (e.g., Logistic Regression & Random Forest).
- Evaluate using Accuracy, F1-score, Confusion Matrix.
- Choose the best model and save it (joblib or pickle).

3. Deployment (Flask OR Streamlit)

- Take user inputs for applicant details.
- Return prediction (Approved / Not Approved) + probability.
- Have a clean, simple UI.
- (Optional bonus: log inputs + predictions in a CSV.)

■ Timeline

- Week duration: 1 week
- Deliverables:
- - Jupyter Notebook (EDA + ML)
- - Saved trained model (model.pkl)
- - Web app code (Flask/Streamlit)