

# Loan Approval Predictor

---

## 1. Project Overview

This group project involves building a machine learning model to predict whether a loan application should be approved based on basic applicant information.

You will work in your groups to complete the ML lifecycle from data exploration to model testing. Deployment is optional.

## 2. Project Goal

Predict loan approval using a binary classification model (Approved or Not Approved).

## 3. Dataset

Name: Loan Approval Dataset

Source: <https://www.kaggle.com/datasets/granjithkumar/loan-approval-data-set/data>

Files: Loan\_train.csv

## 4. Tasks to Complete

1. Understand the problem and define objectives.
2. Load the dataset from Kaggle.
3. Perform Exploratory Data Analysis (EDA):
  - Identify patterns and distributions
  - Check class balance for the target variable
4. Clean and preprocess the data:
  - Handle missing values
  - Encode categorical variables
  - Normalize/scale features (optional)
5. Perform feature selection (optional).
6. Split the dataset into training and testing sets.
7. Train two models:
  - Logistic Regression
  - Decision Tree

8. Evaluate the models using:

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix

9. Test the models on the reserved test set.

10. Draw conclusions and give recommendations based on your findings.

## 5. Tech Stack

Python, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Jupyter Notebook

## 6. Submission Requirements

Each group must submit Github links with :

1. Jupyter Notebook (.ipynb or .py) with clean code and markdown
2. Project Report (.pdf or .docx), summarizing your process, findings, and conclusion
3. A README.md file

## 7. Evaluation Rubric (Out of 100)

| Criteria                                | Marks     |
|---|-----------|
| <b>Exploratory Data Analysis (EDA)</b>  | <b>10</b> |
| - Overview of dataset and key features  |           |
| - Target variable class balance checked |           |
| - Visualizations or summaries included  |           |
| <b>Data Preprocessing</b>               | <b>15</b> |
| - Missing values handled appropriately  |           |
| - Categorical variables encoded         |           |
| Scaling/normalization (if needed)       |           |

|  |            |
|--|------------|
| <b>Model Development</b>   | <b>20</b>  |
| - Both Logistic Regression and Decision Tree built and trained   |            |
| - Training and test split handled correctly                      |            |
| <b>Model Evaluation</b>  | <b>15</b>  |
| - Use of accuracy, precision, recall, F1-score, confusion matrix |            |
| - Clear comparison of model performance                          |            |
| <b>Code Quality</b>  | <b>10</b>  |
| - Code is clean, modular, and well-commented                     |            |
| - Logical structure in notebook or script                        |            |
| <b>Report &amp; Conclusions</b>                                  | <b>20</b>  |
| - Concise summary of approach and findings                       |            |
| - Justified conclusions and recommendations                      |            |
| <b>Bonus: Presentation/Visuals</b>                               | <b>+10</b> |
| - Group includes presentation slides or dashboard                |            |