

Project Summary: Loan Approval Prediction using Machine Learning

1. Understanding the Dataset

- **Load the data:** We start by reading the training dataset into a DataFrame using `pandas`.
- **Shape and basic info:** Check how many rows and columns the dataset has (`shape`), and inspect its general structure (`info()`).
- **Statistics:** Use `describe()` to get insights like mean, median, and standard deviation for numerical columns.

2. Exploratory Data Analysis (EDA)

- **Categorical relationships:** Use `crosstab()` to see how variables like `Credit_History` relate to `Loan_Status`.
- **Boxplots and Histograms:**
 - Boxplots help visualize income ranges and detect outliers for columns like `ApplicantIncome` and `LoanAmount`.
 - Histograms display the distribution of values.

3. Data Cleaning and Missing Value Handling

- Use `isnull().sum()` to check missing values.
- Fill missing values:
 - Categorical variables are filled with their mode (most frequent value).
 - Numerical variables like `LoanAmount` are filled with their mean.
- Transform skewed numerical data:
 - Use `np.log()` to reduce skewness in columns like `LoanAmount`.

4. Feature Engineering

- Combine `ApplicantIncome` and `CoapplicantIncome` into `TotalIncome` for better predictive power.
- Apply logarithmic transformation (`TotalIncome_log`) to normalize the data.

5. Preparing Data for Machine Learning

- **Feature Selection:** Extract relevant columns (independent variables `x` and dependent variable `y`).
- **Encoding Categorical Variables:** Use `LabelEncoder` to convert text categories into numbers.

- **Splitting Data:** Divide data into training and testing sets with `train_test_split()`.

6. Data Standardization

- Use `StandardScaler` to standardize features (scale them to have a mean of 0 and a standard deviation of 1) for better model performance.

7. Machine Learning Models

- **Decision Tree Classifier:**
 - Train a decision tree model on the training data.
 - Evaluate its accuracy on the test data.
- **Naive Bayes Classifier:**
 - Train a Naive Bayes model.
 - Evaluate its accuracy on the test data.

8. Making Predictions on Test Data

- Clean and preprocess the test dataset (similar to training data).
- Make predictions using the trained Naive Bayes model.