Project Questions

1.Display the Top 5 Rows Of the Dataset

2.Display the last 5 Rows Of the Dataset

3. Find Shape of The Dataset (Number of Rows And Number of Columns)

4. Get Information About Our Dataset Like Total Number Rows, Total Number of Columns, Datatypes of Each Column

5. Check Null Values In The Dataset

6.Displaying the missing values in percentage

7. Drop columns with missing values of less than 5%

8. Replace the missing values in columns Self\_employed and Credit\_history using the mode

9. Handling Categorical Columns

10. Creating The Feature Variable X and Target Variable y

11.Feature Scaling

12. Splitting The Dataset Into The Training Set And Test Set & Applying K-Fold Cross Validation

13.Logistic Regression

14.SVC- Support Vector Classifier

15.Decision Tree Classifier

16.Random Forest Classifier

17.Gradient Boosting Classifier

18.Hyperparameter Tuning

19. Save the model

### **Steps Followed**

1. **Data Loading and Initial Exploration**
   * Imported necessary libraries, including pandas and google.colab.drive.
   * Loaded the dataset from Google Drive.
   * Displayed the first and last 5 rows of the dataset.
   * Checked the shape of the dataset.
   * Obtained information about the dataset, including the number of rows, columns, and data types of each column.
   * Identified columns with missing values and calculated the percentage of missing values.
2. **Data Cleaning**
   * Dropped the Loan\_ID column, as it was not required.
   * Dropped columns with missing values less than 5%.
   * Filled remaining missing values with appropriate strategies (mean, median, or mode).
3. **Data Preprocessing**
   * Converted categorical variables into numerical values using label encoding.
   * Checked for and addressed multicollinearity using correlation matrix and VIF (Variance Inflation Factor).
4. **Model Building**
   * Split the data into training and testing sets.
   * Trained and evaluated several models, including Logistic Regression, Decision Tree Classifier, Random Forest Classifier, K-Nearest Neighbors, and Support Vector Classifier.
   * Used GridSearchCV for hyperparameter tuning of the best performing models.
5. **Model Evaluation**
   * Evaluated models using accuracy scores before and after hyperparameter tuning.
   * Chose the Random Forest Classifier based on performance metrics.
6. **Model Deployment**
   * Saved the trained model using joblib.
   * Loaded the saved model and made predictions on new data.
   * Displayed the loan approval status for each prediction.