NAME: **PARNA MISHRA** SAP ID: **60009220201**

COURSE NAME: **Machine Learning - I Laboratory**  BATCH: **D2-1**

# Mini Project

**Task 4**

**(Model)**

**Title of Your Project: Predicting Loan Approval**

**Aim of the Project:**

The primary objective of this project is to build a predictive model to determine the approval status of a loan application. This model will assist banks and financial institutions in making decisions regarding loan approvals based on applicant data, thereby reducing financial risks and enhancing decision-making efficiency.

**Data Description:**

The dataset comprises several attributes associated with loan applicants, as follows:

* Loan\_ID: A unique identifier for each loan application (Categorical)
* Gender: Male or Female (Categorical)
* Married: Indicates if the applicant is married (Yes/No) (Categorical)
* Dependents: Number of dependents reliant on the applicant (Categorical)
* Education: Education level of the applicant (Graduate/Not Graduate) (Categorical)
* Self\_Employed: Whether the applicant is self-employed (Yes/No) (Categorical)
* ApplicantIncome: Monthly income of the applicant (Numerical)
* CoapplicantIncome: Monthly income of the coapplicant (Numerical)
* LoanAmount: Total loan amount requested (Numerical)
* Loan\_Amount\_Term: Term of the loan in months (Numerical)
* Credit\_History: Credit history meets guidelines (0/1) (Categorical)
* Property\_Area: Urban, Semi-Urban, Rural (Categorical)
* Loan\_Status: Loan approved (Y/N) (Categorical)

**Data Preprocessing:**

* The preprocessing steps taken to prepare the data include:
* Handling Missing Values: Missing data were filled using appropriate strategies like median for numerical features and mode for categorical features.
* Encoding Categorical Data: Categorical features including Gender, Married, Education, Self\_Employed, and Property\_Area were encoded using one-hot encoding to convert them into a format suitable for modeling.
* Feature Scaling: Numerical variables such as ApplicantIncome, CoapplicantIncome, and LoanAmount were scaled using standardization to normalize their range.

**Data Modeling:**

The project evaluated multiple machine learning models including Logistic Regression, Decision Trees, Random Forest, and SVM. Each model was trained on the dataset and their hyperparameters were optimized using techniques like GridSearchCV to ensure optimal performance.

**Performance Evaluation:**

Model performances were compared using metrics such as accuracy, precision, recall, and F1 score. The models' ability to accurately predict loan approval was further analyzed using confusion matrices. The Random Forest model performed the best in terms of accuracy and overall balance between sensitivity and specificity, making it the model of choice for predicting loan approval.