**1. Project Title**

**Loan Approval Prediction Using Machine Learning**

**2. Team Members**

* [Nitin Kumar - 524110051]
* [Karan Kanoujia - 524110011]

**3. Objective**

The objective of this project is to develop a machine learning-based model that can accurately predict whether a loan application will be approved or not, based on historical data. This will assist financial institutions in automating the loan eligibility process and reducing human error and bias.

**4. Problem Statement**

Financial institutions receive thousands of loan applications every month. Evaluating each application manually is time-consuming and prone to inconsistencies. A reliable and data-driven solution can help predict the likelihood of loan approval, based on key applicant details such as income, credit history, employment, and more.

**5. Dataset**

We will be using the **Loan Prediction Dataset from Kaggle**:

* Link: https://www.kaggle.com/datasets/altruistdelhite04/loan-prediction-problem-dataset
* Dataset contains:
  + 614 rows and 13 features (e.g., Gender, Married, Education, ApplicantIncome, LoanAmount, Credit\_History, etc.)
  + Target variable: **Loan\_Status** (Y/N)

**6. Methodology**

**a) Data Preprocessing**

* Handling missing values
* Encoding categorical variables
* Normalizing/Scaling numerical features
* Feature selection/engineering

**b) Exploratory Data Analysis (EDA)**

* Visualizing distributions
* Understanding feature relationships
* Identifying patterns in approvals

**c) Model Development**

We plan to implement and compare various classification algorithms:

* Logistic Regression
* Decision Tree
* Random Forest
* Support Vector Machine (SVM)
* XGBoost
* K-Nearest Neighbors (KNN)

**d) Model Evaluation**

* Using accuracy, precision, recall, F1-score, ROC-AUC
* Confusion matrix for binary classification

**e) Hyperparameter Tuning**

* Grid Search / Randomized Search to optimize models

**7. Expected Outcomes**

* A trained and validated machine learning model that predicts loan approval status
* Comparative analysis of different models' performance
* A dashboard or UI (optional) for loan prediction based on user input

**8. Tools & Technologies**

* Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, XGBoost)
* Jupyter Notebook / Google Colab
* GitHub for version control
* (Optional) Streamlit or Flask for web interface

**9. References**

* Kaggle Dataset: Loan Prediction Problem
* Scikit-learn documentation
* Research papers on loan prediction models
* Blogs and tutorials related to machine learning in finance