

# Machine Learning Project Proposal

## Title: Loan Prediction System for Banks

### Project Description:

The project aims to develop a robust and accurate loan prediction system for banks using machine learning algorithms. The system will analyze historical customer data, such as credit history, income, loan amount, loan term, and employment status, to predict whether a loan applicant is likely to receive loan or not. The goal is to assist banks in making informed and reliable lending decisions, reduce the risk of financial losses, and streamline the loan approval process.

The successful implementation of the loan prediction system will enable banks to make data-driven decisions, minimize loan default risks, optimize their lending processes, and ultimately enhance their overall financial performance.

### Data Description:

For this problem, two CSV files: train and test are available. Train file will be used for training the model, i.e., the model will learn from this file. It contains all the independent variables and the target variable. Test file contains all the independent variables, but not the target variable. The model will be applied to predict the target variable for the test data. Given below is the description for each variable with its data type.

Variable	Description	Data Type
Loan_ID	Unique Loan ID	Object
Gender	Gender of the applicant	Object
Married	Applicant Marital Status	Object
Dependents	Number of Dependents	Object
Education	Applicant Education	Object
Self_Employed	Whether applicant is self-employed or not	Object
ApplicantIncome	Income of applicant	Int
CoapplicantIncome	Income of coapplicant	Float
LoanAmount	Required Loan Amount in thousands	Float
Loan_Amount_Term	Term of loan in months	Float
Credit_History	Credit history of applicant	Float
Property_Area	Urban/Semi Urban/ Rural	Object
Loan_Status	Loan Approved	Object

## **Objectives:**

Comfort Zone company deals in all home loans. They have presence across all urban, semi urban, and rural areas. Customer first apply for home loan after that company validates the customer eligibility for loan. Company wants to automate the loan eligibility process (real time) based on customer detail provided while filling online application form.

To automate this process, they have given a problem to identify the customers' segments, those are eligible for loan amount so that they can specifically target these customers. The goal of this project is to predict whether a loan would be approved or not.

## **Problem Statement:**

Loans are the core business of banks. The main profit comes directly from the loan's interest. The loan companies grant a loan after an intensive process of verification and validation. However, they still do not have assurance if the applicant is able to repay the loan with no difficulties. The two most critical questions in the lending industry are:

1. How risky is the borrower?
2. Given the borrower's risk, should we lend him/her?

In the modern era, the data science teams in the banks build predictive models using machine learning to predict how likely a client is going to default the loan when they only have a handful of information. Loan Prediction is a very common real-life problem that each retail bank faces at least once in its lifetime. If done correctly, it can save a lot of man hours at the end of a retail bank.

## **Methodology/Analysis Plan:**

The system will involve the following key components:

1. Data Collection: Gathering historical loan applicant data, including both approved and denied loan applications, from the bank's existing database.
2. Data Preprocessing: Cleaning and preparing the data for analysis by handling missing values, encoding categorical variables, and normalizing features.
3. Feature Selection: Identifying the most relevant features that significantly impact loan approval and default rates.
4. Model Training: Developing and training machine learning models, such as logistic regression, decision trees, k-means clustering, using the historical data to predict loan approval outcomes.
5. Model Evaluation: Evaluating the performance of the trained models using metrics such as accuracy, precision, recall, and F1 score to ensure the reliability of predictions.
6. Continuous Improvement: Implementing mechanisms to continuously update and retrain the model with new data to improve its accuracy and adapt to changing lending trends.

## **Conclusion:**

The implementation of this system holds the promise of reducing the risk of financial losses associated with loan defaults, optimizing the loan approval process, and enhancing the overall operational effectiveness of banks. Through the continuous improvement of the model with newly acquired data, the system can adapt to evolving market dynamics and ensure its relevance over time.