Date	9 September 2022
Team ID	PNT2022TMID00759
Project Name	Smart Lender - Applicant Credibility Prediction for Loan Approval
Maximum Marks	4 Marks

Introduction:

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 don't have assurance if the applicant is able to repay the loan with no difficulties.

The process of bank credit risk evaluation is recognized at banks across the globe. As we know credit risk evaluation is very crucial, there is a variety of techniques are used for risk

level calculation. In addition, credit risk is one of the main functions of the banking community.

The prediction of credit defaulters is one of the difficult tasks for any bank. But by forecasting the loan defaulters, the banks definitely may reduce their loss by reducing their non-profit assets, so that recovery of approved loans can take place without any loss and it can play as the contributing parameter of the bank statement.

This makes the study of this loan approval prediction important. Machine Learning techniques are very crucial and useful in the prediction of these types of data.

Understanding the Problem Statement:

They have a presence across all urban, semi-urban and rural areas. The customer first applies for a home loan and after that, the company validates the customer eligibility for the loan.

The company wants to automate the loan eligibility process (real-time) based on customer detail provided while filling out online application forms. These details are Gender, Marital Status, Education, and number of Dependents, Income, Loan Amount, Credit History, and others.

To automate this process, they have provided a dataset to identify the customer segments that are eligible for loan amounts so that they can specifically target these customers.

We need to predict our Target label which is "Loan Status". Loan status can have two values: **Yes** or **NO**.

Yes: if the loan is approved

NO: if the loan is not approved

So using the training dataset we will train our model and try to predict our target column that is "Loan Status" on the test dataset.

Project Description:

One of the most important factors which affect our country's economy and financial condition is the credit system governed by the banks.

The process of bank credit risk evaluation is recognized at banks across the globe. "As we know credit risk evaluation is very crucial, there is a variety of techniques are used for risk level calculation.

In addition, credit risk is one of the main functions of the banking community.
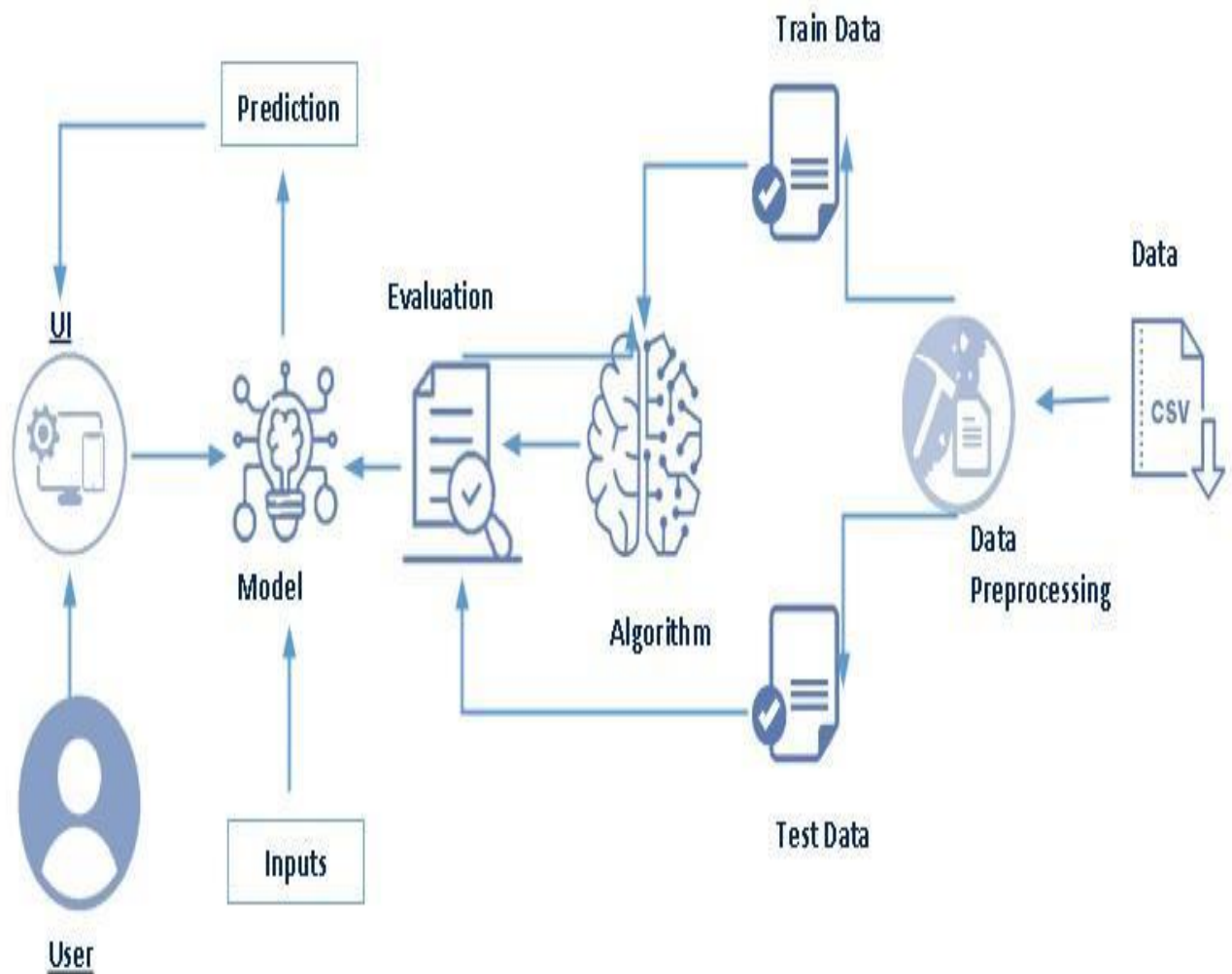
The prediction of credit defaulters is one of the difficult tasks for any bank.

But by forecasting the loan defaulters, the banks definitely may reduce their loss by reducing their non-profit assets, so that recovery of approved loans can take place without any loss and it can play as the contributing parameter of the bank statement.

This makes the study of this loan approval prediction important. Machine Learning techniques are very crucial and useful in the prediction of these types of data.

Architecture:

About the dataset:

So train and test dataset would have the same columns except for the target column that is "Loan Status".

Train dataset:

Variable	Description
Loan_ID	Unique Loan ID
Gender	Male/ Female
Married	Applicant married (Y/N)
Dependents	Number of dependents
Education	Applicant Education (Graduate/ Under Graduate)
Self_Employed	Self employed (Y/N)
ApplicantIncome	Applicant income
CoapplicantIncome	Coapplicant income
LoanAmount	Loan amount in thousands
Loan_Amount_Term	Term of loan in months
Credit_History	credit history meets guidelines
Property_Area	Urban/ Semi Urban/ Rural
Loan_Status	(Target) Loan approved (Y/N)

Decision Tree:

Decision tree is a type of supervised learning algorithm (having a pre-defined target variable) that is mostly used in classification problems. In this technique, we split the population or

sample into two or more homogeneous sets (or sub-populations) based on the most significant splitter/differentiator in input variables.

Decision trees use multiple algorithms to decide to split a node into two or more sub-nodes. The creation of sub-nodes increases the homogeneity of resultant sub-nodes. In other words, we can say that purity of the node increases with respect to the target variable.

Conclusion:

In this tutorial, we learned how to create models to predict the target variable, i.e. if the applicant will be able to repay the loan or not.