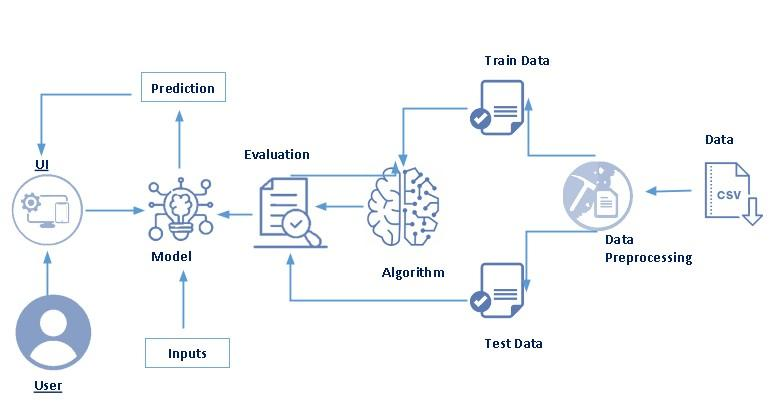
**Smart Lender Loan Approval Using Machine Learning**

Loans are like financial tools that allow people and businesses to borrow money, which they must then repay over time, usually with some extra money added as interest. There are many types of loans out there, like personal loans, mortgages, car loans, student loans, and business loans. These loans are offered by banks, credit unions, and other financial institutions, each with its own unique rules and requirements. Things like interest rates, how long you have to pay the loan back, and any additional charges can all be different depending on the lender and the specific type of loan.

Now, let's focus on personal loans. Think of them as a flexible way to borrow money without needing to put up any collateral. People use personal loans for all sorts of things, like fixing up their homes, covering medical bills, or consolidating their debts. How much money you can borrow, the interest rate you'll be charged, and how long you have to repay the loan will depend on where you get the loan and how trustworthy you appear to the lender. Generally, to get a personal loan, you need to show that you have a regular income and a good credit history.

Now, here's where it gets interesting. Machine learning, a type of computer technology, can be used to predict whether someone will be approved for a personal loan. It does this by crunching numbers and analyzing the person's financial info and credit history. This helps banks and other money-lending places make smarter choices about which loan applications they should say "yes" to and which ones they should turn down. So, in a nutshell, it's a bit like having a virtual financial expert to help decide who should get a personal loan and who shouldn't

**Technical Architecture**:



**Project Flow:**

● User interacts with the UI to enter the input.

● Entered input is analysed by the model which is integrated.

● Once model analyses the input the prediction is showcased on the UI

To accomplish this, we have to complete all the activities listed below

● Define Problem / Problem Understanding

○ Specify the business problem

○ Business requirements

○ Literature Survey

○ Social or Business Impact.

●Data Collection & Preparation

○ Collect the dataset

○ Data Preparation

● Exploratory Data Analysis

○ Descriptive statistical

○ Visual Analysis

● Model Building

○ Training the model in multiple algorithms

○ Testing the model

● Performance Testing & Hyperparameter Tuning

○ Testing model with multiple evaluation metrics

○ Comparing model accuracy before & after applying hyperparameter tuning

● Model Deployment

○ Save the best model

○ Integrate with Web Framework

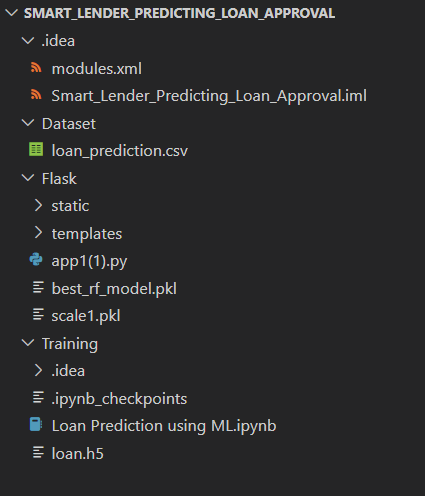
● Project Demonstration & Documentation

○ Record explanation Video for project end to end solution

○ Project Documentation-Step by step project development procedure

**Project Structure:**

Create the Project folder which contains files as shown below



● We are building a flask application which needs HTML pages stored in the templates folder and a python script app.py for scripting.

● rdf.pkl is our saved model. Further we will use this model for flask integration.

● Training folder contains a model training file

**Milestone 1: Define Problem / Problem Understanding**

**Activity 1: Specify the business problem**

Refer Project Description

**Activity 2: Business requirements**

The business requirements for a machine learning model to predict Smart lender predicting loan approval include the ability to accurately predict loan approval based on applicant information,Minimise the number of false positives (approved loans that default) and false negatives(rejected loans that would have been successful).Provide an explanation for the model's decision, to comply with regulations and improve transparency.

**Activity 3: Literature Survey**

As the data is increasing daily due to digitization in the banking sector, people want to apply for loans through the internet. Machine Learning (ML), as a typical method for information investigation, has gotten more consideration increasingly. Individuals of various businesses are utilising ML calculations to take care of the issues dependent on their industry information. Banks are facing a significant problem in the approval of the loan. Daily there are so many applications that are challenging to manage by the bank employees, and also the chances of some mistakes are high.Most banks earn profit from the loan, but it is risky to choose deserving customers from the number of applications.There are various algorithms that have been used with varying levels of success. Logistic regression, decision tree,random forest, and neural networks have all been used and have been able to accurately predict loan defaults. Commonly used features in these studies include credit score, income, and employment history, sometimes also other features like age, occupation, and education level.

**Activity 4: Social or Business Impact.**

Social Impact :- Personal loans can stimulate economic growth by providing individualswith the funds they need to make major purchases, start businesses, or invest in their Education

Business Model/Impact :- Personal loan providers may charge fees for services such as loan origination, processing, and late payments.Advertising the brand awareness and marketing to reach out to potential borrowers to generate revenue.

**Milestone 2: Data Collection & Preparation**

ML depends heavily on data. It is the most crucial aspect that makes algorithm training possible. So this section allows you to download the required dataset.

**Activity 1: Collect the dataset**

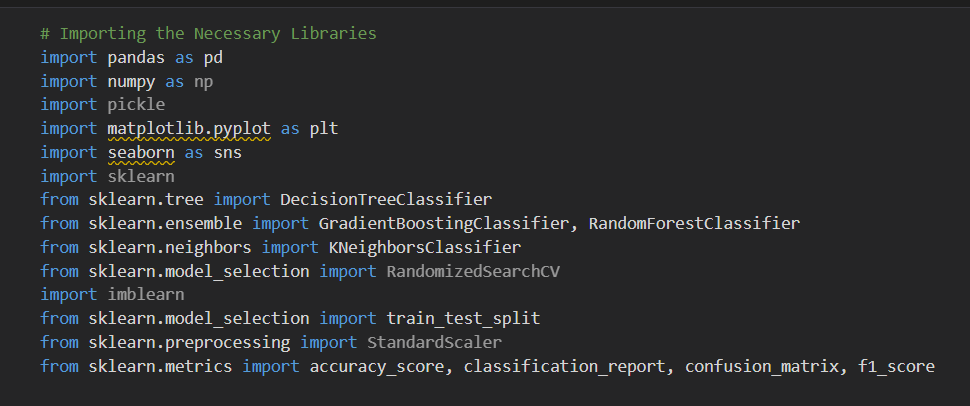
There are many popular open sources for collecting the data. Eg: kaggle.com, UCI repository, etc.In this project we have used .csv data. This data is downloaded from kaggle.com. Please refer to the link given below to download the dataset.

Link: https://www.kaggle.com/datasets/altruistdelhite04/loan-prediction-problem-dataset

As the dataset is downloaded. Let us read and understand the data properly with the help of some visualisation techniques and some analysing techniques.

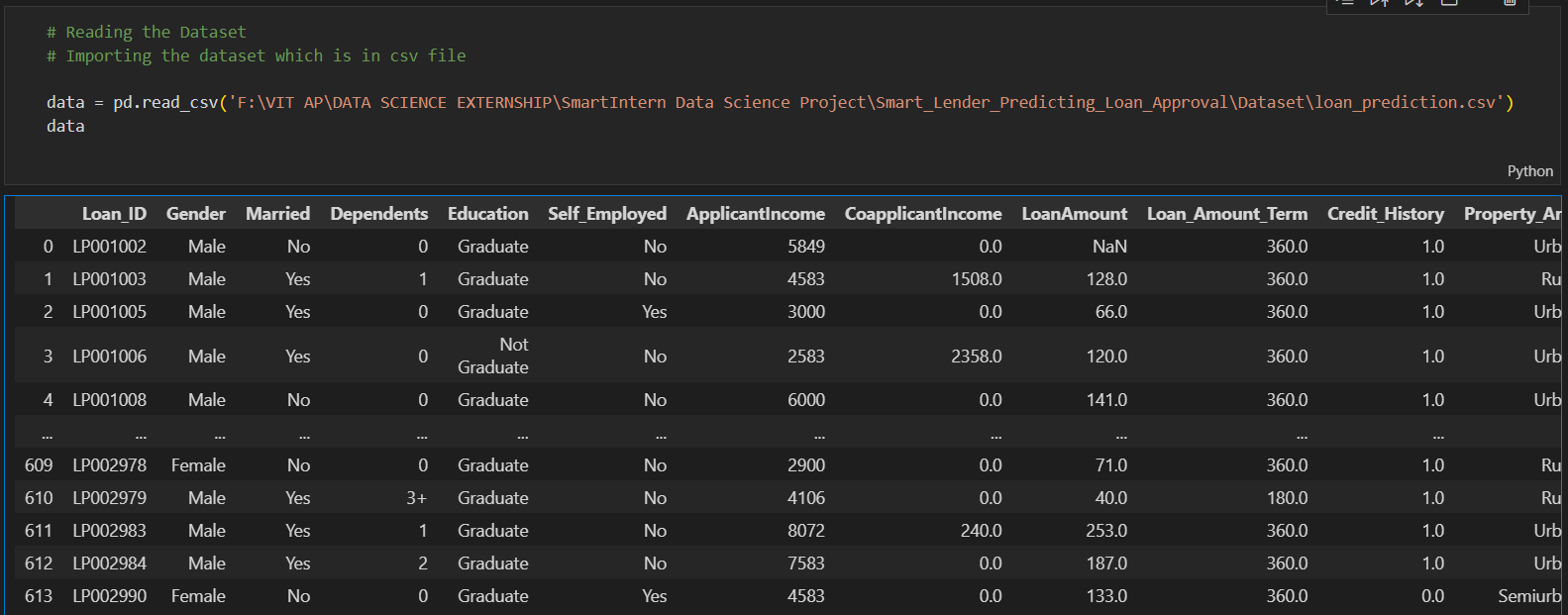
Note: There are a number of techniques for understanding the data. But here we have used some of it. In an additional way, you can use multiple techniques.

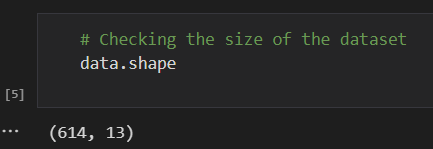
**Activity 1.1: Importing the libraries**



**1.2: Read the Dataset**

Our dataset format might be in .csv, excel files, .txt, .json, etc. We can read the dataset with the help of pandas.In pandas we have a function called read\_csv() to read the dataset. As a parameter we have to give the directory of the csv file.





**Activity 2: Data Preparation**

As we have understood how the data is, let's pre-process the collected data.

The download data set is not suitable for training the machine learning model as it might have so much randomness so we need to clean the dataset properly in order to fetch good results. This activity includes the following steps.

● Handling missing values

● Handling categorical data

● Handling Imbalance Data

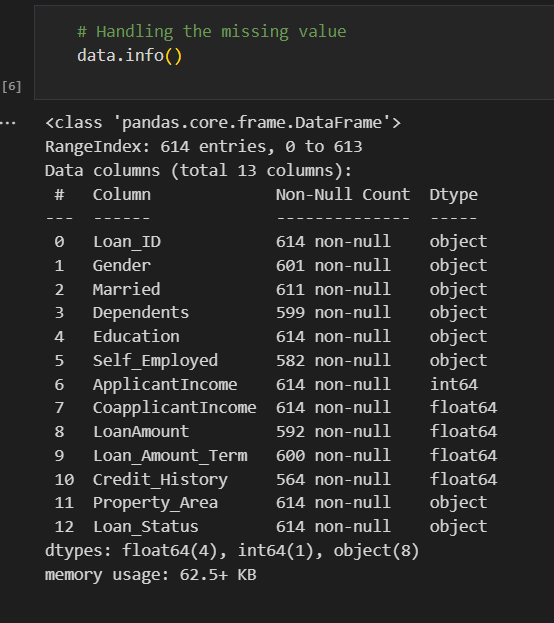
Note: These are the general steps of pre-processing the data before using it for machine

learning. Depending on the condition of your dataset, you may or may not have to go through all these steps.

**Activity 2.1: Handling missing values**

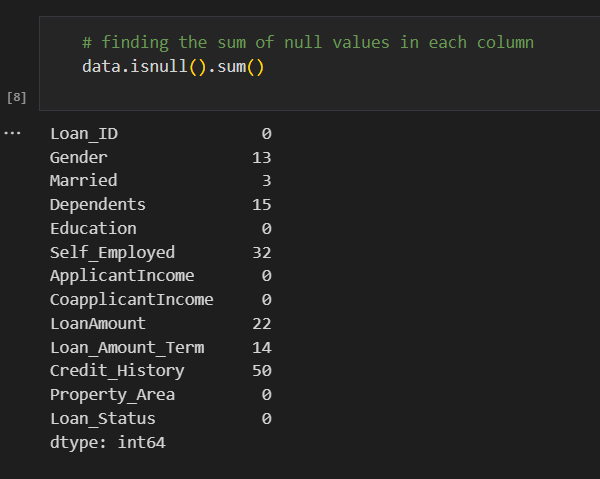
● Lets find the shape of our dataset first. To find the shape of our data, the

df.shape method is used. To find the data type, df.info() function is used.



● For checking the null values, df.isnull() function is used. To sum those null

values we use .sum() function. From the below image we found that there are no null values present in our dataset. So we can skip handling the missing values step.



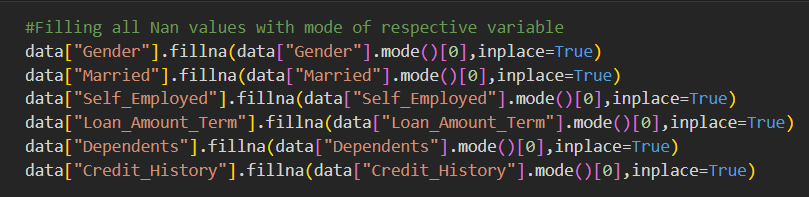
● From the above code of analysis, we can infer that columns such as

gender ,married,dependents,self employed ,loan amount, loan amount

term and credit history are having the missing values, we need to treat

them in a required way.

● We will fill in the missing values in the numeric data type using the mean value of that particular column and categorical data type using the most repeated value



**Activity 2.2: Handling Categorical Values**

As we can see our dataset has categorical data we must convert the categorical data to integer encoding or binary encoding.

To convert the categorical features into numerical features we use encoding techniques.There are several techniques but in our project we are using manual encoding with the help of list comprehension.

● In our project, Gender ,married,dependents,self-employed,co-applicants

income,loan amount ,loan amount term, credit history With list comprehensionencoding is done.