Data Pre-Processing

Splitting Data Into Train And Test

Date	7 November 2022
Team Id	PNT2022TMID10172
Project Name	Smart Lender- Applicant Credibility Prediction for Loan Approval

Now let's split the Dataset into train and test sets

Changes: first split the dataset into x and y and then split the data set

Here x and y variables are created. On the x variable, df is passed by dropping the target variable. And on y target variable is passed. For splitting training and testing data, we are using the train_test_split() function from sklearn. As parameters, we are passing x, y, test_size, and random_state.

In [38]:
```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, OneHotEncoder,StandardScaler
le = LabelEncoder()
oneh = OneHotEncoder()
sc = StandardScaler()
data['Gender'] = le.fit_transform(data['Gender'])
data['Loan_ID'] = le.fit_transform(data['Loan_ID'])
data.head()
x = data.iloc[0:5, 0:2]
x_scaled = sc.fit_transform(x)
x_scaled
x_train, x_test, y_train, y_test = train_test_split(x_scaled, y, test_size = 0.1, random_state = 0)
x_train
```