Data Pre-Processing

Scaling The Data

Date	11 November 2022
Team Id	PNT2022TMID25822
Project Name	Smart Lender- Applicant Credibility Prediction for Loan Approval

Scaling is one the important process, we have to perform on the dataset, because of data measures in different ranges can leads to mislead in prediction

Models such as KNN, Logistic regression need scaled data, as they follow distance based method and Gradient Descent concept.

```
In [34]: 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

Out[34]: array([[-1.41421356,  0.        ],
                [-0.70710678,  0.        ],
                [ 0.        ,  0.        ],
                [ 0.70710678,  0.        ],
                [ 1.41421356,  0.        ]])
```

We will perform scaling only on the input values

Once the dataset is scaled, it will be converted into array and we need to convert it back to dataframe.