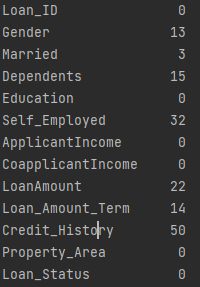
Loan Prediction Project

Artificial Intelligence

**- Reading data and Preprocessing :**

Contains 514 row each row consist of 12 independent features

After we load the data in our model, we see that our data contains null values

 So, first thing we did that we filled the

null values with :

Mode : with column that has object as

data type.

Mean : with column that has non object

data type   
  
we also used method named “IterativeImputer” to fill the null values in selected column .

After we deal with the null values data

We load all the features into “ x ” variable

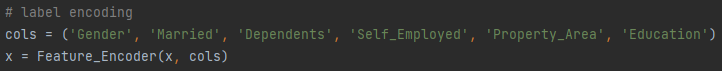
We load Target “Loan\_Status” into “ y ” variable



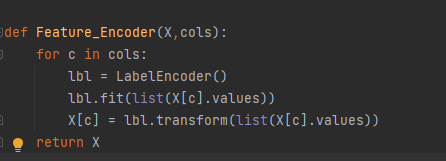
Also, we remove all the duplicates data from all the features



According to the categorical data

We use label encoding to convert them 

Feature\_Encoder :

****

“Credit\_History” column we convert it direct into int data type

Dropping column :

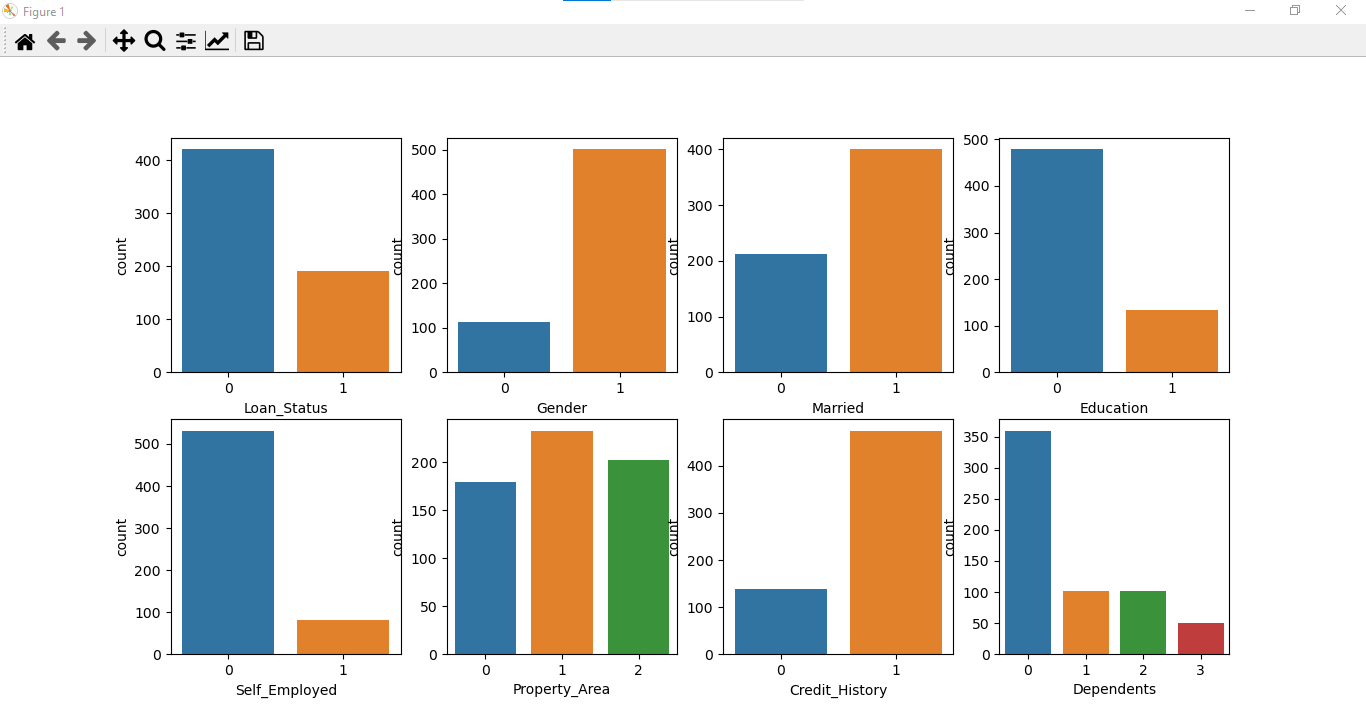
We dropped one column "Loan\_ID"

Last column we worked with Is target column “Loan\_Status”



Every “N” = 0 and Every “ Y” = 1

**Visualization of data**

****

Finally, we normalize the range of independent variables or features of data, and we used “Feature scaling” method to do that .

We used normalization method with min and max values for each feature .

Text

Description automatically generated

So now all the features have values from 0 to 1.

**Classification Part :**

We used four types of models :

We Train data using

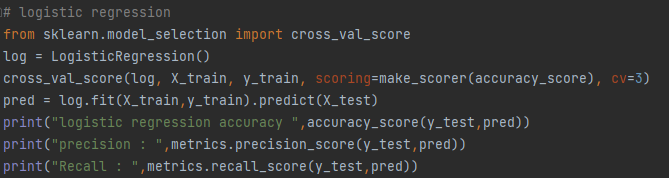
1. Logistic Regression model
2. SVM model
3. Decision tree (ID3) model
4. Random Forrest model

We Are splitting data to 20% test and 80 % train



First, **The Logistic Regression model**

Logistic regression is a classification algorithm used to assign observations to a discrete set of classes



This algorithm had :

|  |  |
| --- | --- |
| logistic regression accuracy | 0.8048780487804879 |
| precision | 0.8333333333333334 |
| Recall | 0.9239130434782609 |

**SVM model** :

**Support vector machine** produces significant accuracy with less computation power. It can be used for both regression and classification tasks, but it is widely used in classification objectives

**Kernel Functions**

Text

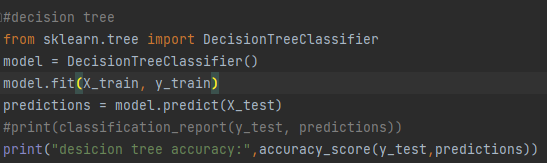
Description automatically generated

This algorithm had :

|  |  |
| --- | --- |
| Svm accuracy | 0.8048780487804879 |
| svm precision | 0.8333333333333334 |
| svm recall | 0.9239130434782609 |

**Decision tree (ID3) model :**

In decision tree learning, ID3 (Iterative Dichotomiser 3) is an algorithm used to generate a decision tree from a dataset



**Decision tree accuracy : 0.6910569105691057**

**Random Forrest model :**

Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time.

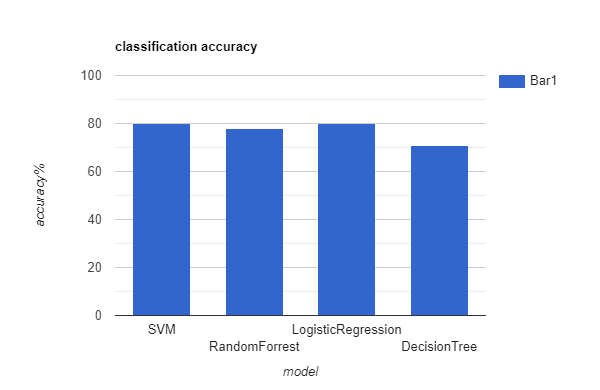
**Text

Description automatically generated**

**Random Forrest accuracy : 0.7804878048780488**

**Now this figure shows the difference between each model**

**Accuracy.**

****