Big Data & Data Analytics-2

Mohan Akshay Bhogadi
170049

WEEK-8
Project Activity - 2

Section – B

GitHub Link:

https://github.com/MohanAkshay/170049_BD2_ProjectActivity2.git

```
shape: Test dataset   (367, 12)
 shape: Train dataset  (614, 13)
Null values in Train dataset
Null values in Train data set
Null values in Test data set
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 614 entries, 0 to 613
Data columns (total 13 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Loan_ID            614 non-null    object
 1   Gender             614 non-null    object
 2   Married            614 non-null    object
 3   Dependents         614 non-null    object
 4   Education          614 non-null    object
 5   Self_Employed      614 non-null    object
 6   ApplicantIncome    614 non-null    int64
 7   CoapplicantIncome  614 non-null    float64
 8   LoanAmount         614 non-null    float64
 9   Loan_Amount_Term   614 non-null    float64
 10  Credit_History     614 non-null    float64
 11  Property_Area      614 non-null    object
 12  Loan_Status        614 non-null    object
dtypes: float64(4), int64(1), object(8)
memory usage: 62.5+ KB
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 367 entries, 0 to 366
Data columns (total 12 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
```

Big Data & Data Analytics-2

Mohan Akshay Bhogadi
170049

```
0    Loan_ID            367 non-null     object
1    Gender             367 non-null     object
2    Married            367 non-null     object
3    Dependents         367 non-null     object
4    Education          367 non-null     object
5    Self_Employed      367 non-null     object
6    ApplicantIncome    367 non-null     int64
7    CoapplicantIncome  367 non-null     int64
8    LoanAmount         367 non-null     float64
9    Loan_Amount_Term   367 non-null     float64
10   Credit_History     367 non-null     float64
11   Property_Area      367 non-null     object
dtypes: float64(3), int64(2), object(7)
memory usage: 34.5+ KB
Encoding categrical variable
Split data Features and Target Varible
Splitting into train and test Data
handling Missing values
Training Data Set Accuracy:  1.0
Training Data F1 Score  1.0
Validation Mean F1 Score:  0.6742937089861218
Validation Mean Accuracy:  0.7393320964749537
Test Accuracy:  0.8536585365853658
Test F1 Score:  0.903225806451613
Confusion Matrix on Test Data
```

Out[1]:

Predicted	0	1	All
True			
0	21	17	38
1	1	84	85
All	22	101	123

Big Data & Data Analytics-2

Mohan Akshay Bhogadi
170049