

**Dataset :**

## **Telco Customer Churn**

**Logistic Regression implementation**

Performance Measure	Training	Test
Accuracy	0.652	0.3784
True Positive Rate	0.9204	0.986
True Negative Rate	0.1822	0.033
Positive Predictive Value	0.3216	0.262
False Discovery Rate	0.6784	0.738
F1 score	0.44	0.414

**Dataset :**

## **Credit Card Fraud Detection**

**Logistic Regression implementation**

Performance Measure	Training	Test
Accuracy	0.998	0.998
True Positive Rate	0.00	0.00
True Negative Rate	1.00	1.00
Positive Predictive Value	0.00	0.00
False Discovery Rate	0.00	0.00
F1 score	nan	nan

Dear Sir,

You have to comment out line:344 and line:345 to see the adult dataset,comment out line: 344 and line:346 to see the credit card dataset. And finally comment out line : 345 and line:346 to see the telecom customer dataset.

Observation:

For all datasets , I have found a random number of accuracy between 0.44 to 0.74 for the train dataset and 0.36 to 0.38 for the test dataset.

We know the range of sigmoid function is  $[-1,1]$ . My predicted y value also lies between  $[-1,1]$ . I have taken the average of maximum and

minimum y predicted value. Then I have set  $y_{\text{predicted}}=0$  if the return value of  $y_{\text{predicted}}$  is less than average. And set  $y_{\text{predicted}}=1$  if the return value of  $y_{\text{predicted}}$  is greater than average.

For credit card dataset, there are a huge number of negative values and a very few positive values. My model always predicts "the same class i.e class 0". So the F1 score is NaN.