

Report

The data set we used was from Kaggle, titled *Covid-19 Temperature, Oxygen, & Pulse Rate Readings*. It consisted of those 3 attributes, and then labelled if that person tested positive or negative with Covid-19.

Not much data preprocessing steps were taken. Some possible problems that a dataset may have missing data points, contain duplicate values, or have a bias. As far as we were concerned, there were no missing data points or duplicate values. Regarding having a bias, that is not something we would know.

The only step taken was splitting up the given dataset into two, 80% for training and 20% for testing our model.

NBC Metrics

Training dataset

Precision score: 0.8897597977243995
Recall score: 0.8837267704671019
Specificity score: 0.8837267704671019
Accuracy score: 0.887625
Log Loss score: 3.8813385754380336
F1 score: 0.8867330225526017
R2 score: 0.5504908974406731

Test dataset

Precision score: 0.8906560636182903
Recall score: 0.8888888888888888
Specificity score: 0.8888888888888888
Accuracy score: 0.889
Log Loss score: 3.833848157693747
F1 score: 0.8897715988083416
R2 score: 0.5559715821812596

Logistic Regression

Training dataset

Precision score: 0.9681335356600911
Recall score: 0.8011049723756906
Specificity score: 0.8011049723756906
Accuracy score: 0.887875
Log Loss score: 3.8726707979956316
F1 score: 0.8767349182355366
R2 score: 0.5514909176910832

Test dataset

Precision score: 0.9806295399515739
Recall score: 0.8035714285714286
Specificity score: 0.8035714285714286
Accuracy score: 0.893
Log Loss score: 3.6956554710348852
F1 score: 0.8833151581243185
R2 score: 0.57197260624679