MULTICLASS CLASSIFICATION

ON MOTOR DRIVE

BY -

Ashish Sancheti ME19B084

AIM OF PROJECT

Our aim is to predict the defect class(“class”) multiclass attribute of 11 defect classes to accurately predict the actual reason of failure

The benefit of knowing the defect type help us in eliminating the need for unnecessary or preventative repairs on the motordrive, allowing the teams to focus only on the defected parts.

It can benefit the organization/team to save on time, labor, cost, etc.

DATASET EXPLORING

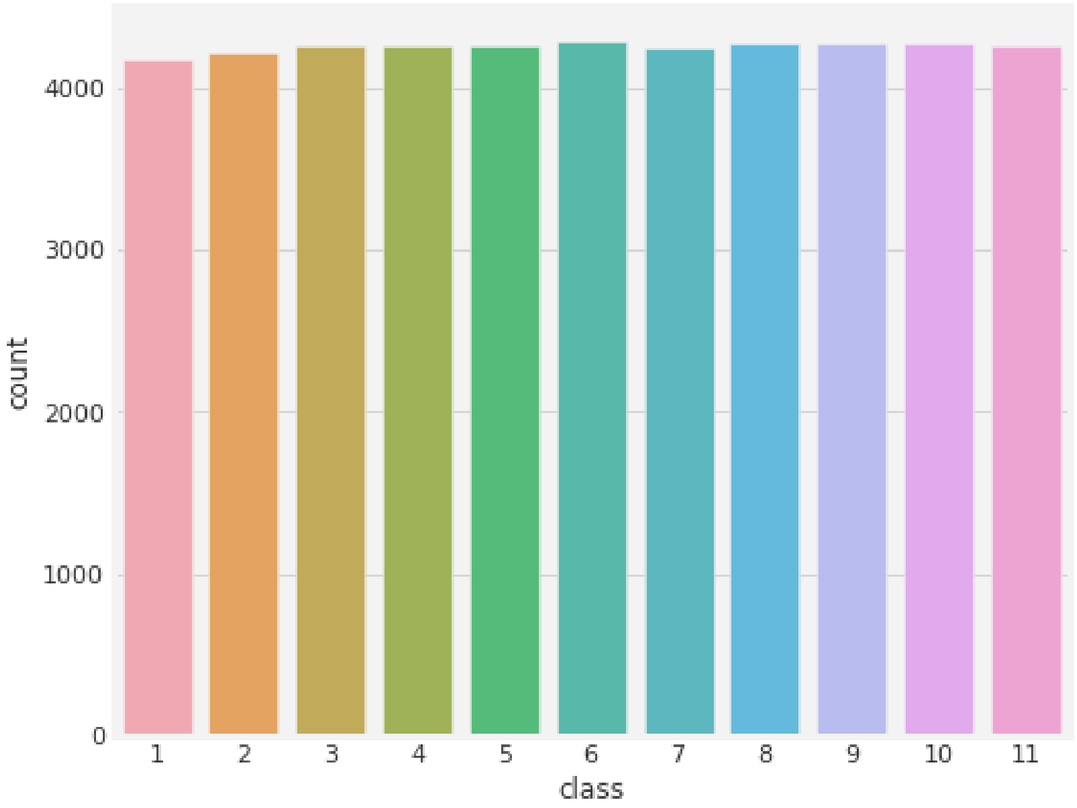
DATASET EXPLORING RAW DATA

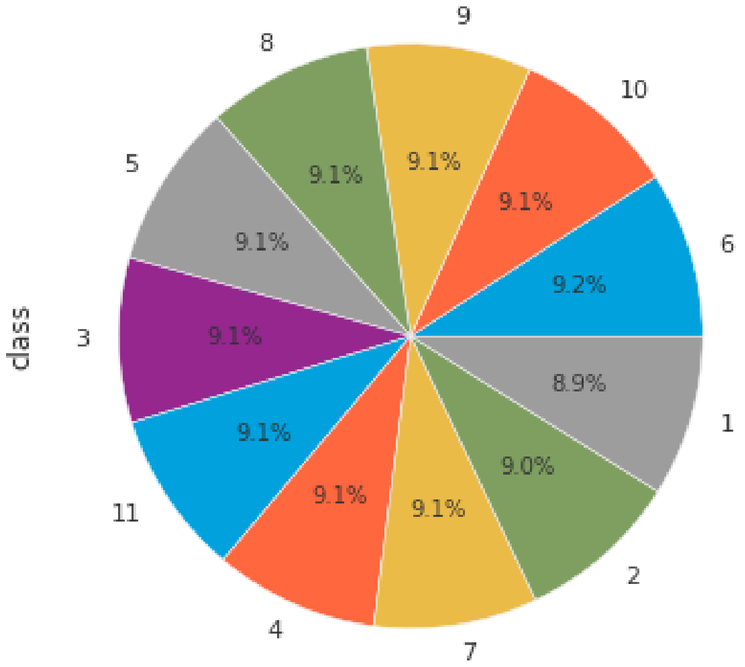
1. It is a Multi-class classification problem
2. There are 11 different type of classes (i.e, type of defects)
3. train dataset consist - (46807 rows, 49 column)
4. test dataset consist - (11702 rows, 48 column)

Dataset Features comprise of -

- 48 different op. conditions

VISUALIZING TRAINING DATASET DISTRIBUTION OF DATA IN DIFFENT CLASSES



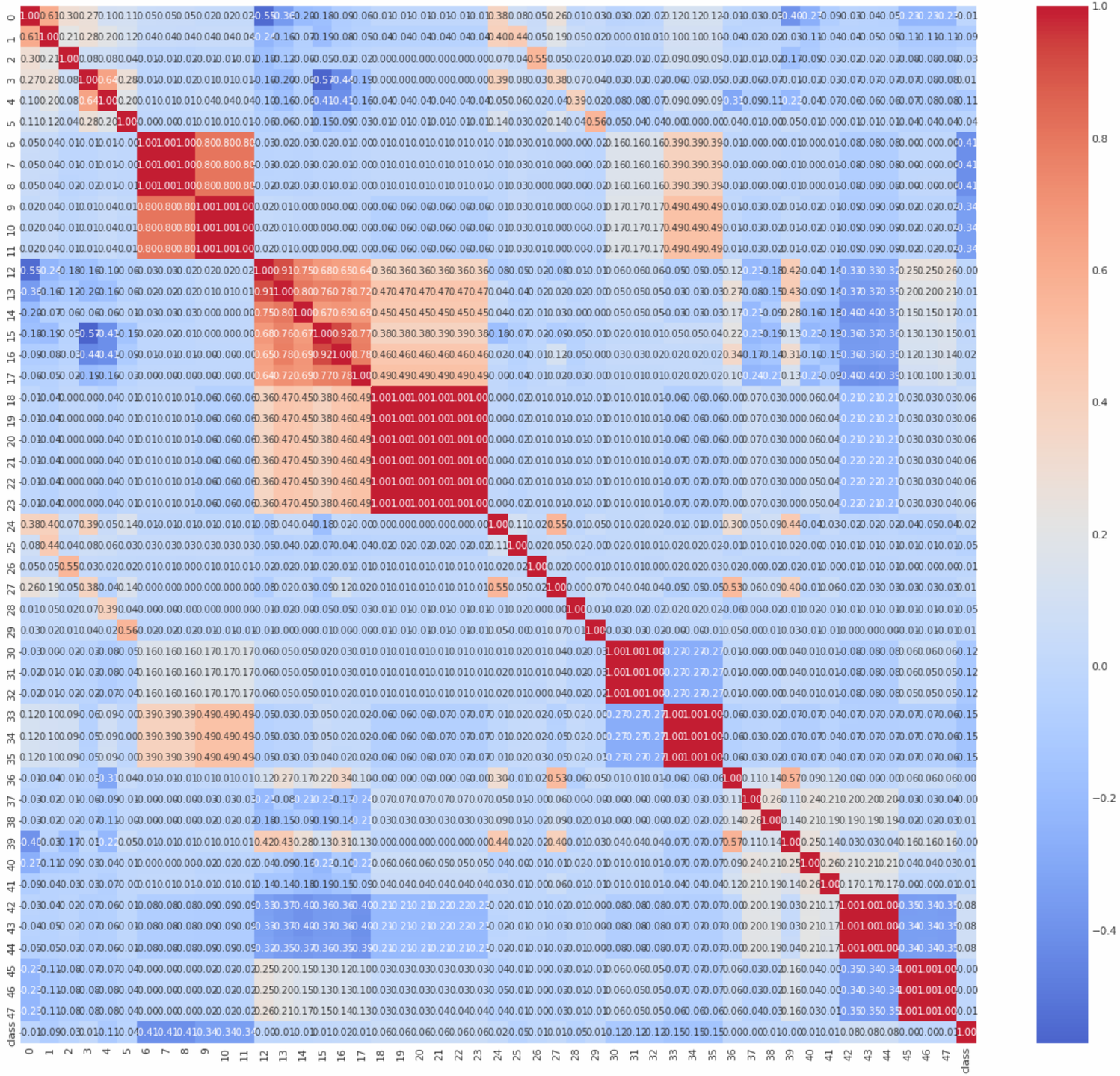


CORRELATIONS MATRIX

Eventhough, there was no need of removing any operation settings or column because there was a very less overall impact on the prediction accuracy.

If needed we can remove -

- 6,7 out of 6,7,8 column
- 30,31 out of 30,31,32
- 33,34 out of 33,34,35
- 42,43 out of 42,43,44
- 45,46 out of 45,46,47



MODEL

FEATURE SCALING

since, the dataset was already scaled. Therefore feature scaling was skipped otherwise we could have used normalization using minmaxscaler fun

Used – MinMaxScaler function

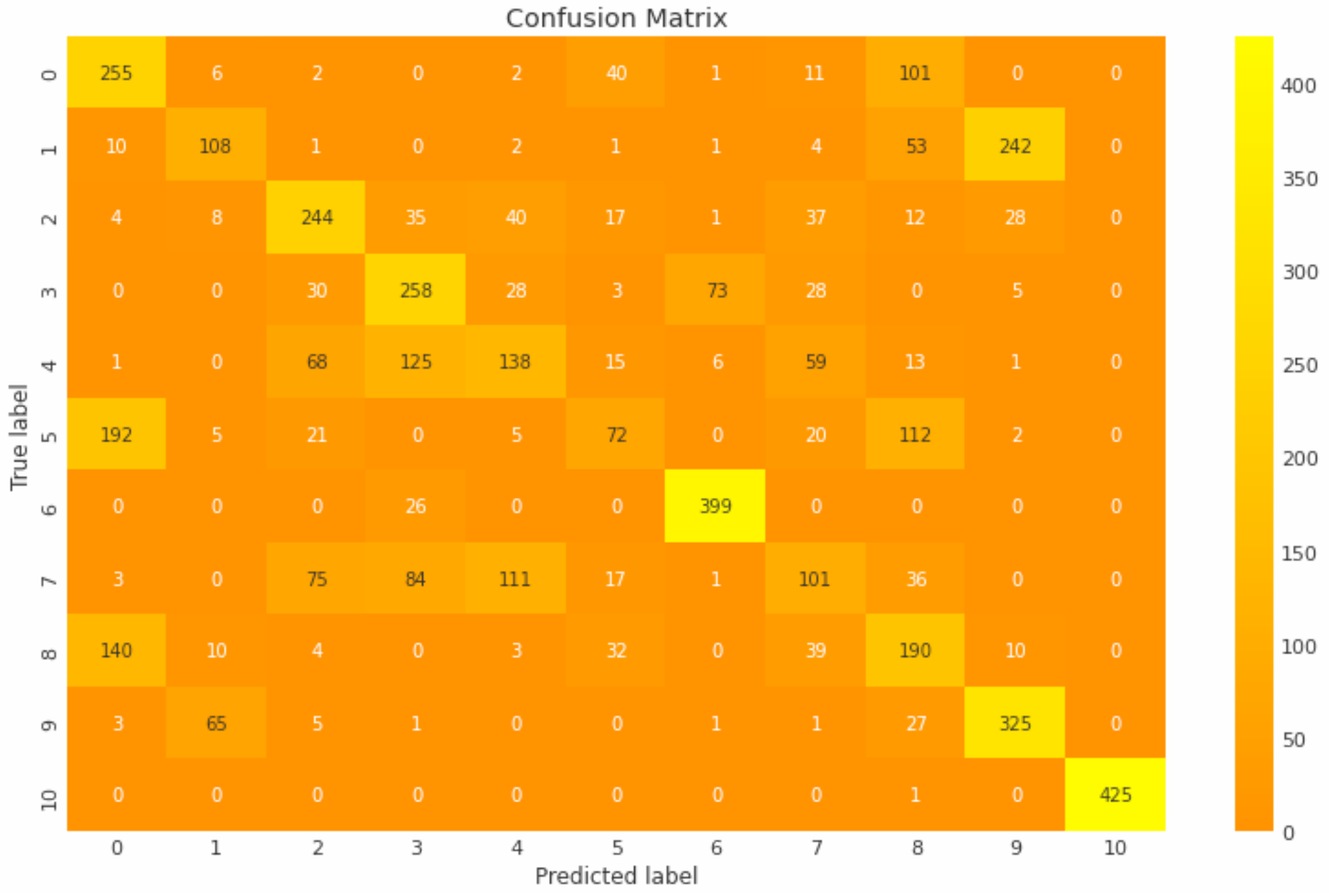
LOGISTIC REGRESSION

train_accuracy_score

0.537838864359303 1

test_accuracy_score

0.535996581926938 7



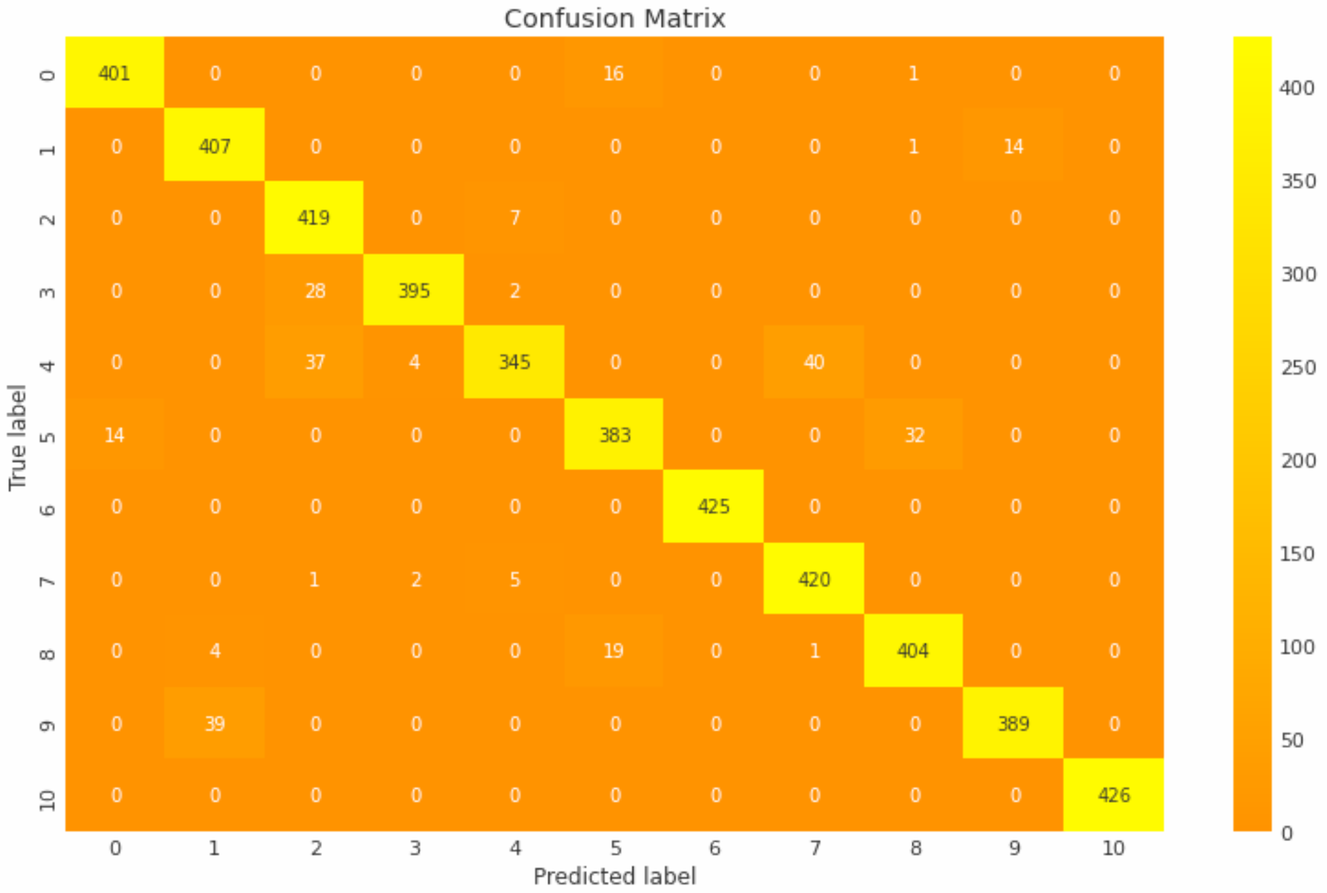
RANDOMFOREST CLASSIFIER

train_accuracy_score

0.946612543322413

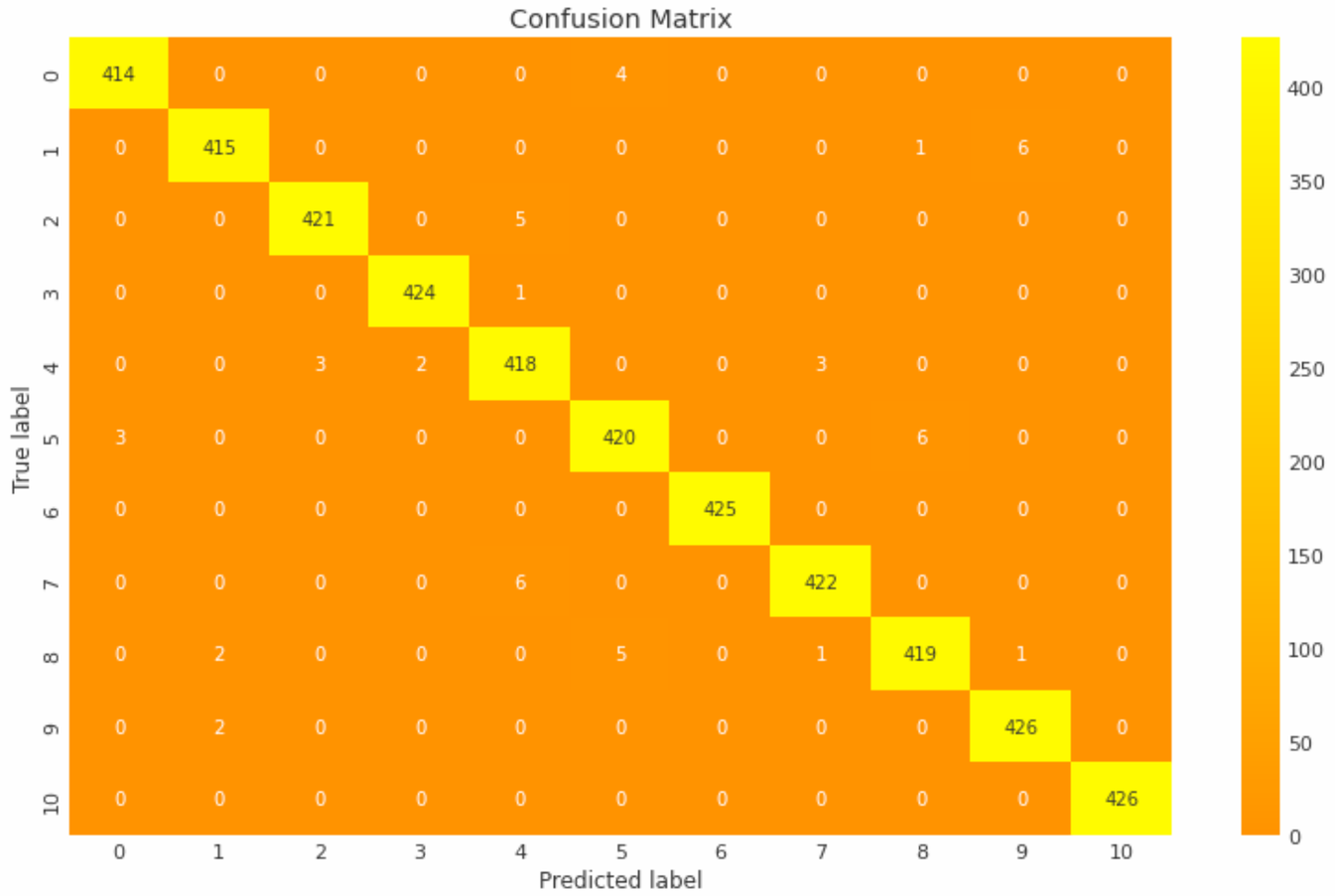
test_accuracy_score

0.942960905789361 2



XGBOOST ALGORITHM

XGBoost is essentially the same thing as Gradient Boost, but the main difference is how the residual trees are built. With XGBoost, the residual trees are built by calculating similarity scores between leaves and the preceding nodes to determine which variables are used as the roots and the nodes.



METRICS CALCULATION

Accuracy	train_accuracy_score - 0.9933295352039121 test_accuracy_score - 0.989104892117069
Precision Score	Macro-Averaged Precision score: 0.9891468211740734 Micro-Averaged Precision score: 0.989104892117069
Recall Score	Macro-averaged recall score: 0.9891159021850776 Micro-Averaged recall score: 0.989104892117069

Thank you !!

Code file - https://colab.research.google.com/drive/13aAU2x0-o4zNzqAZQgWGyNxQmG57ELAI#scrollTo=_MXYJRhkVak0

Reference file -

- https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
- https://www.geeksforgeeks.org/xgboost/