Name: Rupesh VijayKumar Karbhari
Program-Data Science with ML and Python Internship
Batch-Apr 2022-Jun 2022
Certificate ID : TCRIB3R27
Date of Submission : 20/07/2022

Technical Coding Research Innovation, Navi Mumbai,
Maharashtra, India-410206

Product subscription on Bank-Institution using ML and Python

A Case-Study Submitted for the requirement of
Technical Coding Research Innovation

For the Internship Project work done during
DATA SCIENCE WITH MACHINE LEARNING AND PYTHON INTERNSHIP PROGRAM

by
RUPESH VIJAYKUMAR KARBHARI(TCRIB3R27)

Rutuja Doiphode
CO-FOUNDER &CEO
TCR innovation.

Abstract – Supervised Machine learning is a part of Machine learning and Artificial intelligence. It is defined by its use of labelled dataset to train algorithms that aims to classify data from previous information. Classification is very commonly performed in data science problems. Various successful methods have been proposed to solve those problems. This research paper gives you insights into applying a classification algorithm to a dataset. When
the dataset is huge then Random Forest classifier is an algorithm for generating a multiple decision tree.

Index Terms –

1.Problem statement

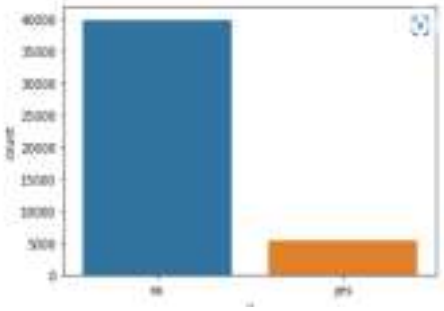
The data is related to direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to assess if the product (bank term deposit) would be ('yes') or not ('no') subscribed (Col -21).

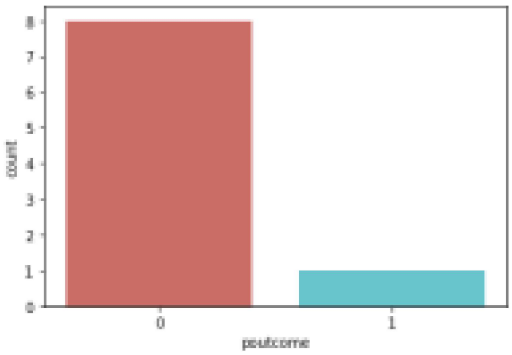
2.Introduction to data set

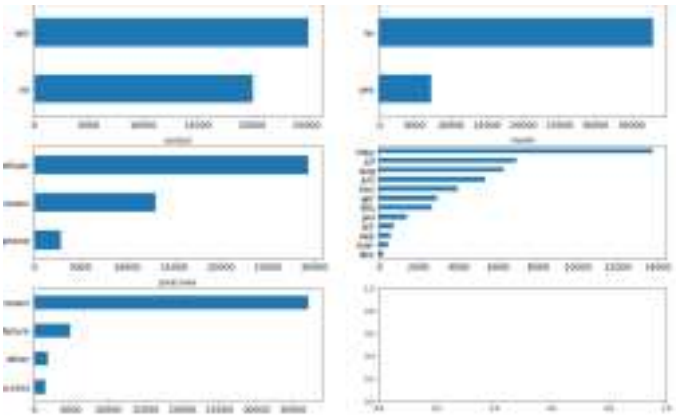
I am using a dataset Bank-full where it as 16 features and 1 target as y and has 45211 rows.there are both numerical and object in dataset

3.Exploratory Data Analysis

It gives an overview about the dataset that can be used to plot the diagrams. After EDA, we can detect such values which could create a bad impact on our prediction results.we can use sns plot to plot graphical representation between two columns







5.comparison

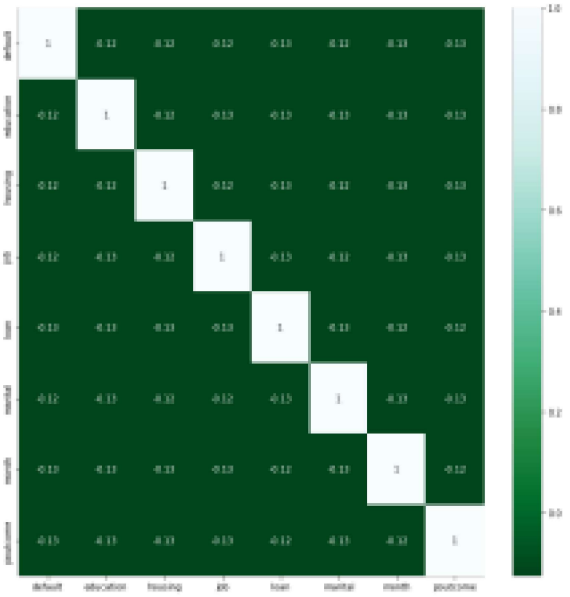
to improve accuracy score we can make report and we can use it for classification .we can use here is the screenshot o accurray score

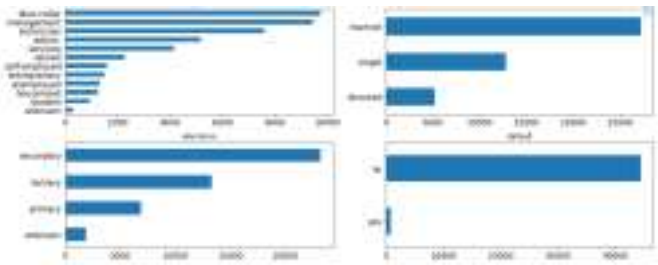
4.Training and prediction of Data

```
preds_test = np.zeros(len(test_X))

from sklearn.metrics import accuracy_score, f1_score
accuracy = accuracy_score(test_y, preds_test)
f1 = f1_score(test_y, preds_test)

print(f'Accuracy score is: {accuracy}')
print(f'F1 score is: {f1}')
```

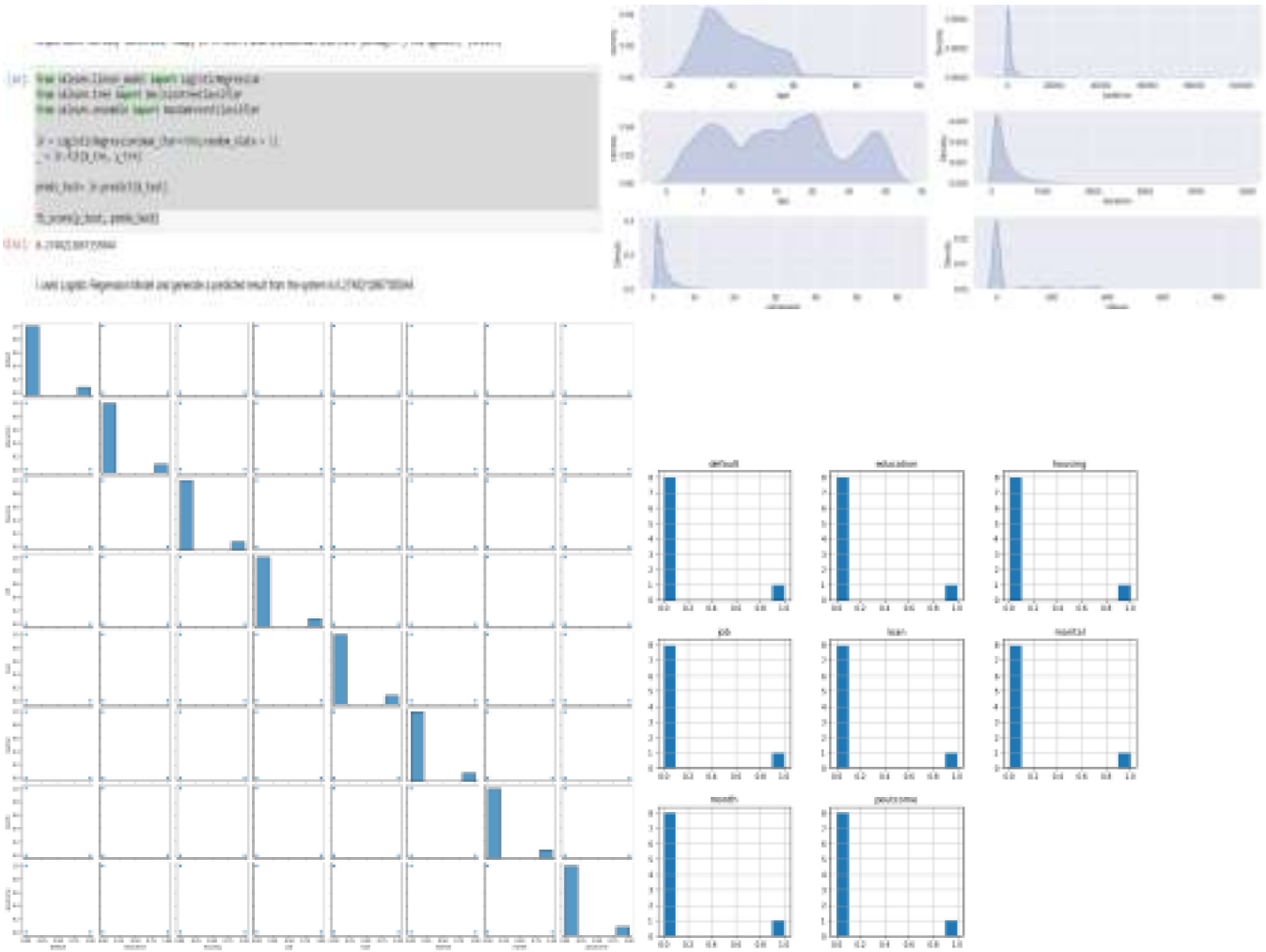
```
Accuracy score is: 0.8874664879356569
F1 score is: 0.0
```

here we are training data and trying to make predictions and





we can also the data that we got from below graphs and we can use those to train data and make predictions and we can use those to determine algorithms which are suitable



6.Conclusion

WWW.TCRINNOVATION.CO.IN INTERNSHIP PROGRAM NAME:

After applying Random Forest classification algorithm, machine learning model is able to predict the results with 90.23 % accuracy. Using

XGboost algorithm, we can predict the same results, but I found that this Random Forest Algorithm provides better results than other classification algorithms.

7.Reference

1. L. Liu and H. Yu, “Efficient feature selection via analysis of Relevance and Redundancy”, JMLR, pp. 1205-1224, 2004.
Internet.

2. K. Wisaeng. “A Comparison of Different classification
Techniques for Bank direct Marketing”. International Journal of Soft Computing and Engineering (IJSCE), vol.3, no. 4, pp. 116-119, 2013.

3. S. B. Kotsiantis, P. E. Pintelas and I. D. Zaharakis.
“Machine Learning: a review of classification and combining techniques”, Artificial Intelligence Rev, vol, 26, pp. 159-190, 2006