**Ids assignment 04**

Name: Ammar Asif

Reg no: FA19-BCS-088

Q1:

1. How many instances does the dataset contain?

* Dataset contains 80 instances.

2. How many input attributes does the dataset contain?

* The dataset contains 7 input attributes which are height, weight, beard, hair length, shoes size, scarf , eye color.

3. How many possible values does the output attribute have?

* The output attribute has possible 2 values which are Male and female.

4. How many input attributes are categorical?

* 4 of the input attributes are categorical which are beard, hair length, scarf and eye color.

5. What is the class ratio (male vs female) in the dataset?

* There are total of 46 males and 34 females in dataset. There class ratio is 23:17.

Q2:

1: No instances are incorrectly classified using Random Forest classifier.

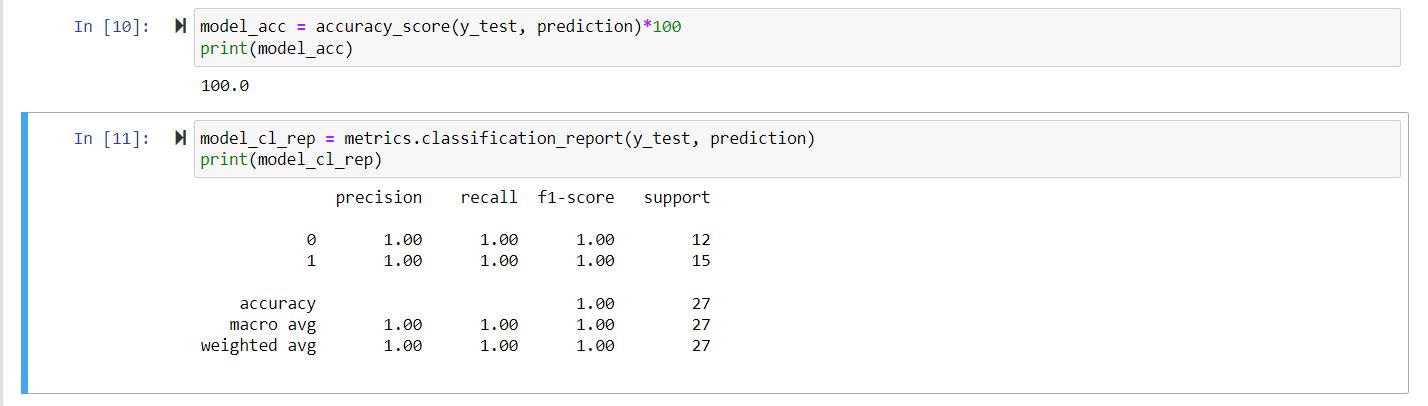


Figure : Using Random Forest

2: 20 instances are incorrectly classified using Support Vector Machine.

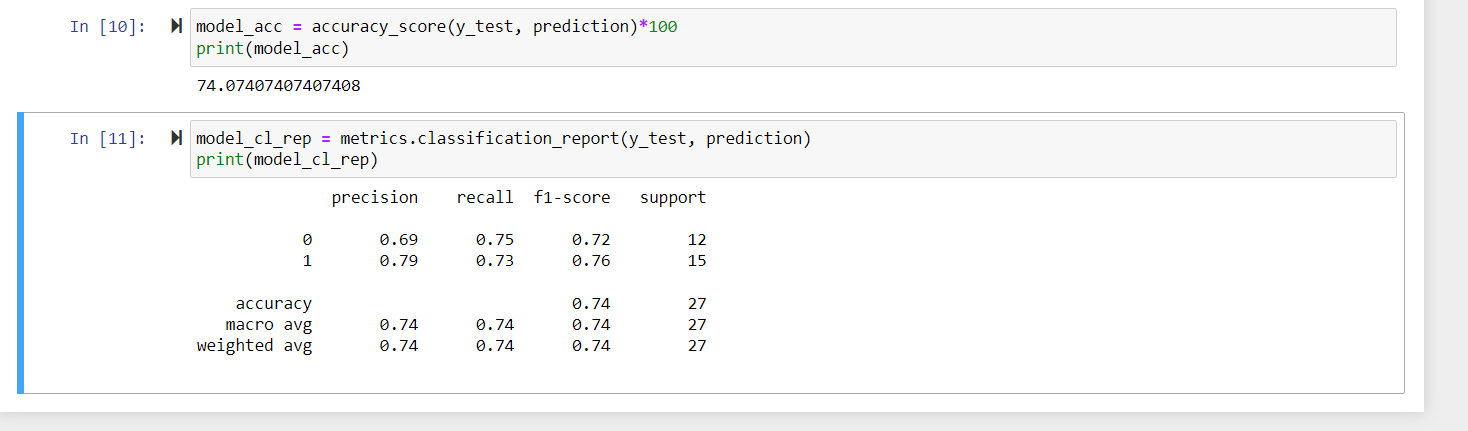
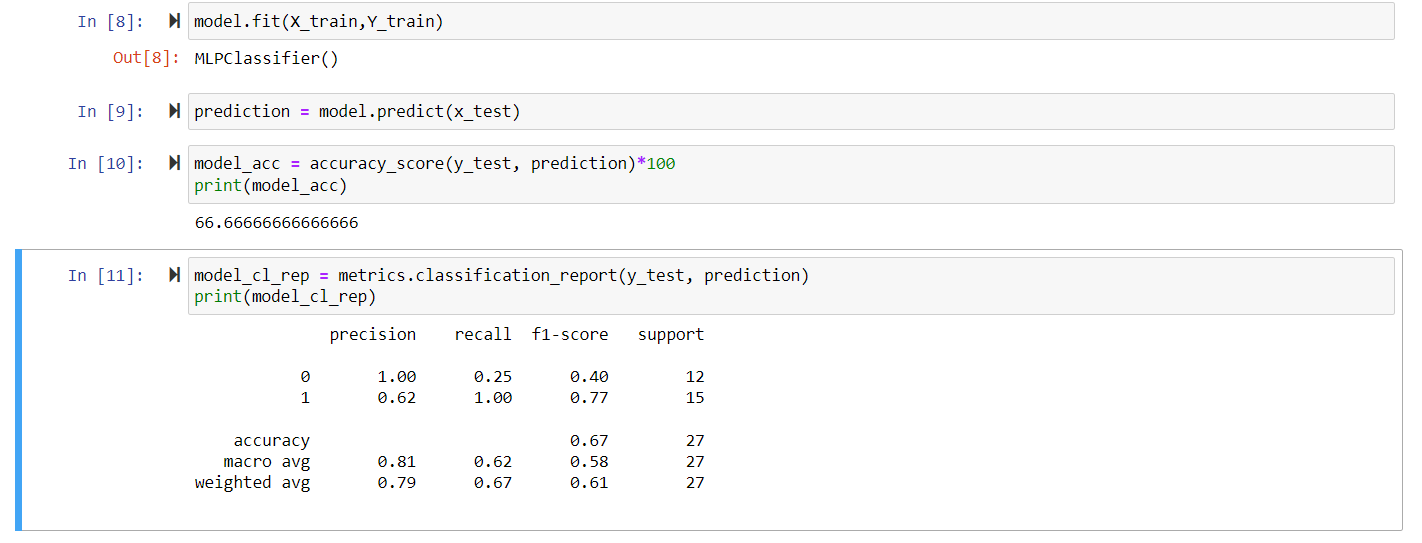


Figure : Using SVM classifier

3: at least 35 instances are incorrectly classified using MLPC.



Q1:

2: There’s no change in Random Forrest classifier after changing to 80/20 train test split ratio. The result is same.

By using SVM classifier the accuracy decreases by approx. 10% after changing to 80/20 train test split ratio.

By using MPLC the accuracy increases by 15% when the train test split ratio is 80/20.

Q1:

3: hair length and beard are 2 most important attributes in classifying this task because these 2 distinguished most between male and female.

Q1:

4:

Yes there’s is a significant decrease in accuracy if we exclude these 2 attributes from dataset.

