Name: Muhammad Hashim

Registration Number: SP20-BCS-062

Assignment #: 4

Group: BCS 6 - G2-A

Submitted to: Prof. Muhammad Sharjeel

Question 1

1. **How many instances does the dataset contain?**

Ans: 80 instances

2. How many input attributes does the dataset contain?

Ans: The dataset contains 7 input attributes

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

Ans: The output attribute have 2 possible values:

- Male
- Female

4. How many input attributes are categorical?

Ans: 4 attributes:

- beard
- hair_length
- scarf
- eye_color

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

Ans: male: 23, female: 17

Question 2

1. **How many instances are incorrectly classified?**
   Ans:
   Random Forest: 0, Support Vector Machine: 6, Multilayer Perceptron: 10
2. **Rerun the experiment using train/test split ratio of 80/20. Do you see any change in the results? Explain**
   Ans:
   Accuracy of the Support Vector Machine and Multilayer Perceptron has increased. While accuracy of Random Forest remains the same.

3. **Name 2 attributes that you believe are the most “powerful” in the prediction task. Explain why?**
   Ans:
   2 most powerful attributes are believed to be "beard" and "scarf". Only males have beard and only females wear scarf.
4. **Try to exclude these 2 attribute(s) from the dataset. Rerun the experiment (using 80/20 train/test split), did you find any change in the results? Explain.**
   Ans:
   The evaluation metrics (accuracy, precision, recall, f1) went up after removing beard and scarf.

Question 3

Monte Carlo Cross Validation
Parameters: n_splits=5, test_size=0.33, random_state=7
F1 Score = 94.94%
Leave p-out Cross Validation
Parameter: LeavePOut(2) //Leave 2 out
F1 Score = 94.19%

Question 4

The accuracy, precision, and recall scores are as follows:

- Accuracy = 100%
- Precision = 100%
- Recall = 100%