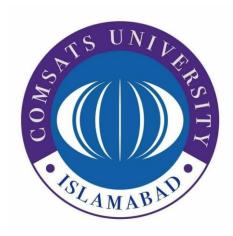
# COMSATS University Islamabad, Lahore Campus



# **Assignment-04- Introduction to Data Science**

Dilawar Hussain SP20-BCS-120 Group -02

**Submitted To: DR. Muhammad Sharjeel** 

#### Q1:

- i. The dataset contains 80 instances.
- ii. The dataset contains 7 input attributes.
- iii. The output attribute has **two possible** values.
- iv. Two input attributes are categorical which are Hair Length and eye color.
- **v.** The class ratio of Male-Female is (46-34) which is (1.4).

#### **Q2**:

#### 1) Standard Train/Test Split

92% accuracy: So 3 instances were incorrectly classified using the Random Forest algorithm.

74% accuracy: So 7 instances were incorrectly classified using the Support Vector Machine algorithm.

**55% accuracy:** So **11** instances were incorrectly classified using the **Multilayer Perceptron** Classifier.

#### 2) 80/20 Train Test Split

100 % accuracy: So 0 instances were incorrectly classified using the Random Forest algorithm.

81 % accuracy: So 5 instances were incorrectly classified using the Support Vector Machine algorithm.

93% accuracy: So 2 instances were incorrectly classified using the Multilayer Perceptron Classifier.

- In 80/20 split the accuracy of the model increases, which can be seen that less number of instances were classified incorrectly.
- 3) **Scarf** and **Beard** are the most powerful attributes in the Prediction task because they have **distinct** values for males and females in almost most of the instances.
- **4**) By removing the above two attributes the accuracy of the system is **reduced.** Like by applying specified classifier algorithms the accuracy compared to previous decreases.

### **Q 3:** Monte Carlo cross-validation:

**F1 Score: 0.94 Parameters:** ShuffleSplit(train\_size=0.66, test\_size=0.34, n\_splits = 5)

**Leave P-Out cross-validation:** 

F1 Score: 0.96 Parameters: p=5

### **Q 4:** 5 instances for testing are:

I create another .csv file named as gender-prediction-test, and add the below 5 instances for testing in that file. After applying the accuracy was 80%

- 1. height: 63, Beard: no, weight: 150, hair length: medium, shoe size: 41, scarf: yes, eye color: blue, gender: female
- 2. height: 81, Beard: no, weight: 182, hair length: Short, shoe size: 38, scarf: NO, eye color: brown, gender: female
- 3. height: 76, Beard: Yes, weight: 170, hair length: Long, shoe size: 37, scarf: NO, eye color: brown, gender: Male
- 4. height: 85, Beard: Yes, weight: 200, hair length: Bald, shoe size: 44, scarf: NO, eye color: Grey, gender: Male
- 5. . height: 79, Beard: NO, weight: 119, hair length: Short, shoe size: 40, scarf: NO, eye color: black, gender: Male

2

## precision recall f1-score support

1.00

0

1 0.7	5 1.0	0 0.8	36	
accuracy		0.0	30 5	
macro avg	0.88	0.75	0.76	5
weighted avg	0.85	0.80	0.78	5

0.50

0.67