12/13/2022

# INTRODUCTION TO DATA SCIENCE

## ASSIGNMENT # 04


**NAME:**

MUHAMMAD SUBHAN KHAN

**REGISTRATION NUMBER:**

SP20-BCS-098

**SECTION:**

(B)

**GROUP:**

4


**SEMESTER:**

6th

**ASSIGNMENT:**

INTRODUCTION TO DATA SCIENCE

# QUESTION NO 1

Provide responses to the following questions about the dataset.

- How many instances does the dataset contain?
- How many input attributes does the dataset contain?
- How many possible values does the output attribute have?
- How many input attributes are categorical?
- What is the class ratio (male vs female) in the dataset?

**A**

There are 18 instances in the dataset.

**B**

There are 7 input instances in the dataset.

**C**

There is 1 possible output instance in the dataset i.e., Gender

**D**

There are 4 categorical input instances in the dataset.

**E**

Ratio of Male: 57.5 %

$$Male\ Ratio = \frac{26}{80} * 100 = 57.5\%$$

Ratio of Female: 42.5 %

$$Female\ Ratio = \frac{34}{80} * 100 = 42.5\%$$

Total Male = 46 MALE

Total Female = 34 FEMALE

# QUESTION NO 2

**A**

**Incorrectly classified instances**

MLP:  12 instances are incorrectly classified.

Random Forest:  1 instance ais incorrectly classified.

SVM:  0 instances are incorrectly classified.

**B**

**WITH 80/20 SPLIT incorrectly classified instances**

MLP:  7 instances are incorrectly classified.

Random Forest:  0 instances are incorrectly classified.

SVM:  1 instance is incorrectly classified.

**C**

**BEARD and Hair Length** are the most powerful attributes in given the dataset.

As these instances can tell us the exact difference between the gender whether it is male or female because females mostly have long hair and no beard whether male have mostly short hair length with beard.

## (D)

## DROPING Hair length and beard

### MLP Confusion matrix Result:

15 instances are correctly classified.

12 instances are incorrectly classified.

### Random Forest Confusion Matrix Result:

15 instances are correctly classified.

1 instance is incorrectly classified.

### SVM Confusion Matrix Result:

16 instances are correctly classified.

0 instances are incorrectly classified.

**EXPLANATION:**

By dropping **hair length** and **beard** instances from the given data set the confusion matrix given best result about correctly and incorrectly classified attributes