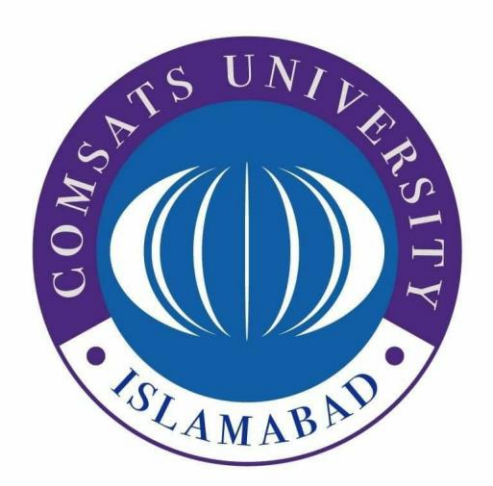
Assignment 3



Name: M. Sarmad Aslam

Roll No: FA21-BSE-093 - C

Subject: Data Science

COMSATS University Islamabad, Lahore Campus

26-November-2023

```
# 26-11-23

# CSC461 – Assignment3 – Machine Learning

# Sarmad Aslam

# Fa21-BSE-093
```

Question No. 1

1. How many instances does the dataset contain?

Ans: The dataset contains 110 instances.

2. How many input attributes does the dataset contain?

Ans: The dataset contains 7 input attributes (height, weight, beard, hair_length, shoe_size, scarf and eye_color)

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

Ans: The output attribute has 2 possible values, male and female.

4. How many input attributes are categorical?

Ans: 4 input attributes are categorical (beard, hair_length, scarf, eye_color)

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

Ans: Count of males: 62

Count of females: 48

The class ratio is 62:48

Question No. 2

2. Rerun the experiment using train/test split ratio of 80/20. Do you see any change in the results? Explain.

Ans: When I changed the size of train/test split, the number of corrects increased and the number of incorrect decreased.

Number of Incorrect before:

```
Incorrect by LR:   5
Incorrect by SVC:  9
Incorrect by MLP:  5
```

Number of Incorrect at 80/20:

```
Incorrect by LR:   0
Incorrect by SVC:  4
Incorrect by MLP:  5
```

Explanation: The reason is that the model got more data for training i.e., 80% of total data which was 67% before, so it got trained better than before and hence the model made more correct predictions, and less incorrect predictions.

3. Name 2 attributes that you believe are the most "powerful" in the prediction task. Explain why?

The attributes **"shoe_size"** and **"height"** are the most powerful attributes in the prediction task because about **44** women out of **48** have shoe size ranging from **35-39** which is about **92%** of total women.

Secondly, about **48** out of **62** males have **height** ranging from **69 – 78** which is about **71%** of total males which makes it more powerful in gender prediction.

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: When I removed the two "powerful" attributes and again Rerun the experiment keeping the train/test split same i.e., 80/20, I saw a slight change.

Number of Incorrect before:

```
Incorrect by LR:   0
Incorrect by SVC:  4
Incorrect by MLP:  5
```

Number of Incorrect after removing “powerful” attributes:

```
Incorrect by LR:   2
Incorrect by SVC:  4
Incorrect by MLP:  4
```

Explanation: The reason is that the model got affected by those two attributes. This can be due to the reason that those two attributes were more helping to distinguish between the two classes “male” and “female”.