**Gender Prediction Report**

**Question:1**

1: Dataset contains 80 Instances.

2: 7 attributes

3: Output attribute has two possible Values.

4: 4 are Categorical attributes

5: 34 are females and 46 are males.

**Question:2**

**1:**

Randomforest accuracy = 100.0

SVC accuracy =77.77777777777779

MLPclassifier accuracy = 88.88888888888889

**2:** After changing the ratio to 80/20 accuracy changes as below

Randomforest accuracy = 100.0

SVC accuracy = 81.25

MLPclassifier accuracy = 100.0

**3:** I think beard and and scarf are the most powerful attributes. Because female don’t have beard at all, so it’s a good attribute to distinguish them from male same as the scarf.

**4:** Lets see, by removing them what is the impact on the prediction.

Randomforest accuracy = 93.75

SVC accuracy = 81.25

MLPclassifier accuracy = 100.0

So, after removing scarf and beard from input, accuracy reduced for the random forest model using 80/20 ratio.

**Question :3**

**Monte Carlo cross validation:**

rs = ShuffleSplit(n\_splits=5, test\_size=.25, random\_state=0)

rs.get\_n\_splits(x)

Cross Validation Scores are [1. 1. 1. 0.95 1. ]

Average Cross Validation score :0.99

**Leave p out cross validation:**

lpo = LeavePOut(p=2)

lpo.get\_n\_splits(x)

**Question 4:**

**Gaussian Naïve Bayes classification algorithm:**

**Test Instances:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| height | weight | beard | hair\_length | shoe\_size | scarf | eye\_color | gender |
| 49 | 179 | Yes | short | 46 | no | brown | male |
| 61 | 127 | No | short | 35 | no | green | female |
| 74 | 139 | No | medium | 37 | yes | gray | female |
| 54 | 125 | no | long | 36 | no | gray | female |
| 72 | 184 | yes | short | 42 | no | gray | male |