

Question No 01:

1. Dataset contains total instances of 80.
2. Dataset has 7 input attributes.
3. Output attribute only has two possible values(male/female).
4. Categorical input attributes are
 - a. Beard **(yes/no)**
 - b. Hair_length **(long/medium/short)**
 - c. Scarf **(yes/no)**
 - d. Eye_color **(black/blue/gray/brown/green)**
5. Male-Female ratio is 57.5%-42.5%.

Question No 02

1. Incorrectly classified instances for each classifier are

- Random Forest = **0(0%)**
- Support Vector Classifier = **~6(22%)**
- Multilayer Perceptron = **~14(52%)**

Exact details are shown in the table below.

2. Using 80-20% split ratio, differences occurred are

- RF – No Difference
- SVC – Accuracy increased from **78%** to **81.25%**
- MLPC – Accuracy increased from **89%** to **81.25%**

	33%-67%			80%-20%		
	RF	SVC	MLPC	RF	SVC	MLPC
Accuracy (%)	100.00%	77.78%	48.96%	100.00%	81.25%	62.96%
Precision(F/M)	1.00/1.00	0.70/0.82	0.00/0.48	1.00/1.00	0.80/0.82	0.00/0.63
Recall(F/M)	1.00/1.00	0.70/0.82	0.00/1.00	1.00/1.00	0.67/0.90	0.00/1.00
F1 Score(F/M)	1.00/1.00	0.70/0.82	0.00/0.49	1.00/1.00	0.73/0.86	0.00/0.77

3. The attributes of **beard** and **scarf** are more effective at predicting tasks.

These are the characteristics that most clearly separate males and females. Women don't grow beards, and males don't wear scarves.

Other traits may be valued equally for men and women.

Both men and women can have short hair. Their height, shoe size, eye color, and weight may all be equal.

4. Excluding two input attributes

- Eye_color
- Shoe_size

The results in accuracy are

- RF – **No Difference**
- SVC – **No Difference**
- MLPC – Accuracy decreased from **62.96%** to **37.50%**

Excluding 5 instances (80%-20%)			
	RF	SVC	MLPC
Accuracy (%)	100.00%	81.25%	62.96%
Precision(F/M)	1.00/1.00	0.80/0.82	0.00/0.63
Recall(F/M)	1.00/1.00	0.67/0.90	0.00/1.00
F1 Score(F/M)	1.00/1.00	0.73/0.86	0.00/0.77

Reason: The reason behind this is, we have less input attributes, and MLPC model underfitted resulting in less accuracy.

Question No 03:

Applying Monte Carlo and leave p-out cross validation on data, resulting F1 score is given in the table below using 80-20% split

Decision Tree Classifier (80%-20%)		
	Monte Carlo	Leave P-Out
F1 Score	0.955%	0.941%

Question No 04:

Below are the five instances added extra in the previous dataset

62	101	no	long	38	yes	black	female
64	103	no	medium	40	no	brown	female
69	173	yes	short	42	no	black	male
72	185	yes	short	43	no	brown	male
65	120	no	long	37	yes	blue	female

Then using the whole dataset for training and using the above five examples for testing and resulting metrics are shown in the table below.

80%-20%	
	Gaussian NB
Accuracy (%)	80.00%
Precision(F/M)	1.00/0.75
Recall(F/M)	0.50/1.00
F1 Score(F/M)	0.67/0.86