Salman Wasi
FA19-BCS-040

Question No. 1

1. The dataset contains 80 instances
2. The dataset contains 7 input attributes
3. Output has 2 possible values (Male / Female)
4. 4 input values are categorical
    a. Beard
    b. Hair Length
    c. Scarf
    d. Eye color
5. The class ratio is 57.5% male - 42.5% female

Question No. 2

1. Incorrectly classified
    a. Random Forest = 0 (0%)
    b. Support Vector Machine = ~6(22%)
    c. Multilayer Perceptron = ~2(8%)
2. Using an 80/20 split, differences are
    a. Random forest - No Change
    b. Support Vector Machine - Accuracy increases from 77.78 to 81.25
    c. Multilayer Perceptron - Accuracy decreases from 92.59 to 62.5
3. The most discriminating features are beard and scarf because males don't wear scarf and females can't have beards.
4. After excluding the beard and scarf, accuracy is
    a. Random forest - Accuracy drops from 100% to `94.11%`
    b. Support Vector Machine - Accuracy increases from 77.78 to 88.23
    c. Multilayer Perceptron - Accuracy increases from 92.59 to 94.11

Question No. 3

Applying Monte Carlo with a 33% test size and Leaving P out with p=2, the F1 score is

1. Monte Carlo - `0.95`
2. `Leave P-out - 0.94`

Question No. 4

Newly added instances are

Salman Wasi
FA19-BCS-040

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

Scores are

1. Accuracy - 0.80
2. Precision (F/M) - 1.00/0.75
3. Recall (F/M) - 0.50/1.00
4. F1 Score (F/M) - 0.67/0.86