***ASSIGNMENT-3***

**Name: Eisha Fatima**

**Class: BSE-C**

**Roll number:084.**

**Question 1:**

**Total Instances:** The dataset comprises 110 records.

**Input Features:** It consists of seven input variables: height, weight, beard, hair\_length, shoe\_size, scarf, and eye color.

**Output Attribute Variability:** The gender attribute as the output can take on two forms: male or female.

**Categorical Input Variables:** In this dataset, the variables classified as categorical are beard, hair\_length, scarf, eye\_color, and gender.

**Gender Class Distribution:** The dataset's gender distribution can be quantified by the ratio of males to females. With 62 males and 48 females, the ratio is computed as 62/48, indicating that there are approximately 1.29 males for every female.

**Question 2:**

**Misclassification Counts**:

**Logistic Regression**: Misclassified 5 instances

**Support Vector Machines**: Misclassified 9 instances

**Multilayer Perceptron:** Misclassified 10 instances

**Impact of 80/20 Train/Test Split:**

While the accuracy of the models remains relatively consistent, the count of misclassified instances varies slightly.

**Logistic Regression:** No change in misclassified instances

**Support Vector Machines**: Reduction of 2 misclassified instances

**Multilayer Perceptron**: Increase of 2 misclassified instances

**Key Predictive Attributes**: Height and weight are deemed the most significant predictors due to their distinct and variable nature.

**Excluding Key Attributes and Rerunning Experiment**:

Logistic Regression's accuracy reaches 1.0.

The accuracy of Support Vector Machines and Multilayer Perceptron stays nearly the same.

Misclassification for Logistic Regression drops to 0, decreases to 2 for Support Vector Machines, and decreases to 6 for Multilayer Perceptron.

**Question 3:**

Parameter Settings:

**Monte-Carlo Experiment:** monte\_carlo\_f1\_scores = cross\_val\_score(rf\_classifier, X, y, cv=5, scoring=f1\_scorer) with n\_estimators set to 100 and random\_state to 42.

**P-Out Cross-Validation:** p\_out\_f1\_scores = cross\_val\_score(rf\_classifier, X, y, cv=leave\_p\_out, scoring=f1\_scorer) with p\_out set to 15.

**Question 4:**

*Testing Instances:*

**Instance 1:** 72, 132, no, medium, 45, no, brown, male

**Instance 2:** 72, 154, yes, medium, 42, no, blue, male

**Instance 3**: 70, 119, yes, medium, 40, no, brown, male