Assignment 4

> **Name: Mohammed Varaliya**

> **Roll No: 54**

Question

1. **What happens if we interchange the places of `X_train`, `X_test`, `y_train` & `y_test`?**

In machine learning, we split the data into training and testing sets: `X_train`, `X_test`, `y_train`, and `y_test`. Typically, we use `X_train` and `y_train` for training the model and `X_test` and `y_test` for evaluating its performance.

If we swap these variables, it disrupts the purpose of training and testing:

1. **Training with `X_test` and `y_test`:**
    a. Training the model with `X_test` and `y_test` means the model is learning from the data it should be tested on. This leads to overfitting, as the model may

memorize the test data, which undermines the purpose of testing on unseen data.

2. **Testing with `X_train` and `y_train`**:
    a. When you test with `X_train` and `y_train` (the data used for training), the evaluation results will be biased and overly optimistic because the model has already seen this data.
3. **Consequences**:
    a. **Overfitting**: The model may perform well on the training data but fail to generalize to new, unseen data.
    b. **Misleading Evaluation**: Metrics like accuracy may be inflated since the model is tested on data it was already trained on.
4. **Conclusion**:
    a. The order of these variables matters for ensuring the model generalizes well. Swapping `X_train`, `X_test`, `y_train`, and `y_test` leads to misleading results and defeats the purpose of having separate training and testing sets.
