Question 2 :

Each symptom can be analyzed to see how often it appears with each disease label

Symptoms like **Fever** and **Shortness of Breath** are more common with **COVID**, while **Runny Nose** and **Sneezing** may be more indicative of **ALLERGY** or **COLD**.

Question 5:

Total Time Complexity:

O(m⋅n⋅k)

Where:

m = number of test entries

n = number of entries in the main dataset

k = number of symptoms (columns)

Question 6:

### **Designing an Efficient Algorithm**

1. **Using KD-Tree or Ball Tree**:
   * Implement a KD-Tree or Ball Tree to store the symptom data, allowing for efficient querying of nearest neighbors.

### **Comparison with Previous Algorithm**

* **Previous Approach**: Naive distance calculation for each test instance led to high time complexity, especially with large datasets.
* **Improved Approach**: Using KD-Tree or Ball Tree reduces query time, allowing the algorithm to scale better with the dataset size. This could lead to faster predictions and potentially higher classification accuracy as a result of efficiently exploring the feature space.