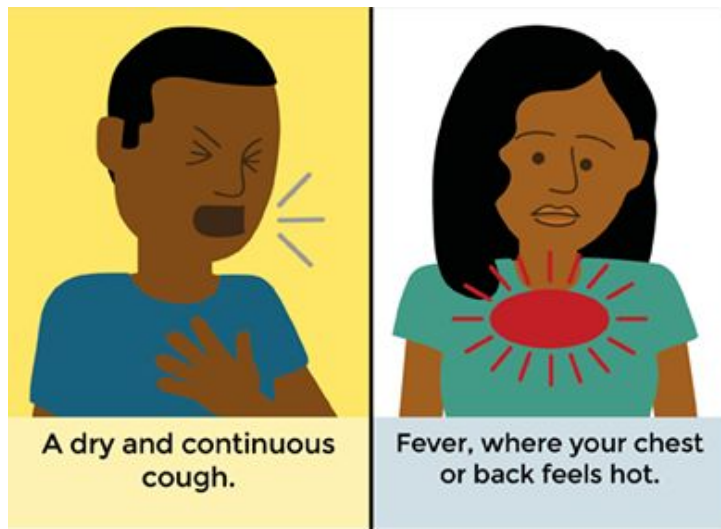


## CENG 484 - Data Mining

### Exercise 3

In this exercise, you will implement **k-nearest neighbors (KNN)** algorithm on **Covid-19** data. It will be detect whether a person has Covid-19 virus according to some features. This data has been simplified to facilitate implementation of KNN algorithm. It contains only two features as cough\_level (0:no cough, 5:severe cough) and fever(35.0-39.9). You can choose Python or R for coding.



Try to implement KNN by writing **your own function**. Complete the stages which are given below.

- Write your **min-max normalization** function and apply on the data.
- Define a knn function as `KNN(train_data, test_data, distance_type, k)`. Distance type can be euclidean or manhattan. This function will return **predicted value**.
- Implement **Euclidean** distance.

Euclidean  
Distance

$$\sqrt{\sum |x_i - y_i|^2}$$

d) Implement **Manhattan** distance.

$$\text{Manhattan Distance} \quad \sum |x_i - y_i|$$

e) Calculate output for this input:

```
test_set = [[5,39.0,1], [4,35.0,0], [3,38.0,0], [2,39.0,1], [1,35.0,0],  
            [0,36.2,0], [5,39.0,1], [2,35.0,0], [3,38.9,1], [0,35.6,0]]
```

f) Try different **k values** and **draw accuracy graph** according to k parameter.

g) What is the **best k value** according to accuracy? Write as a comment in your source code.

**Note:** This exercise will be graded as +10 bonus points for the **Assignment 3**, you can improve your practice and prepare yourself for the assignments. It is optional, however if you decide to submit your solution, you can submit until May 22.