K-Nearest Neighbor (KNN)

“ Now not do WHEN do, We not do WHO do”

For the corporate landscape rotating completing around Data Science, it has been one of the most sought areas of nature. You will learn how the KNN algorithm operates and how it can be applied using Python in this article on KNN algorithm.

KNN concepts can hardly be described in a simpler way. It is an ancient expression that can be used in dozens of languages and traditions. In other terms, it is also said in the ***Bible:*** “He who walks with wise men will be wise, but the compassion of fools will suffer harm.”. It implies the idea of k-nearest neighbor classifiers is part of our everyday existence and judging.

What is KNN Algorithm?

K nearest neighbors or KNN algorithm is a straightforward algorithm that uses the whole dataset in its training dataset. Whenever a prediction is made for an unknown data instance, it looks for the k-most similar across the entire testing dataset, and eventually returns the data with the most similar instances as the predictions. KNN is often used when searching for similar items, such as finding items similar to this one.The Algorithm suggests that you are one of them because you are close to your neighbors.

How does a KNN algorithm work?

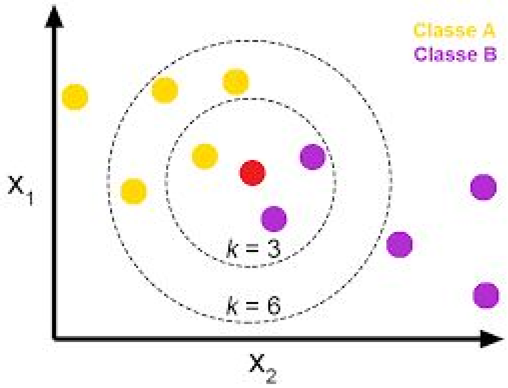
To conduct grouping, the KNN algorithm uses a very basic method to perform classification. When a new example is tested, it searches at the training data and seeks the k training examples which are similar to the new example. It then assigns to the test example of the most similar class label.

What does ‘K’ in KNN algorithm represent?

K in KNN algorithm represents the number of nearest neighboring points that vote for a new class of test data.

If k = 1, then test examples in the training set will be given the same label as the nearest example.

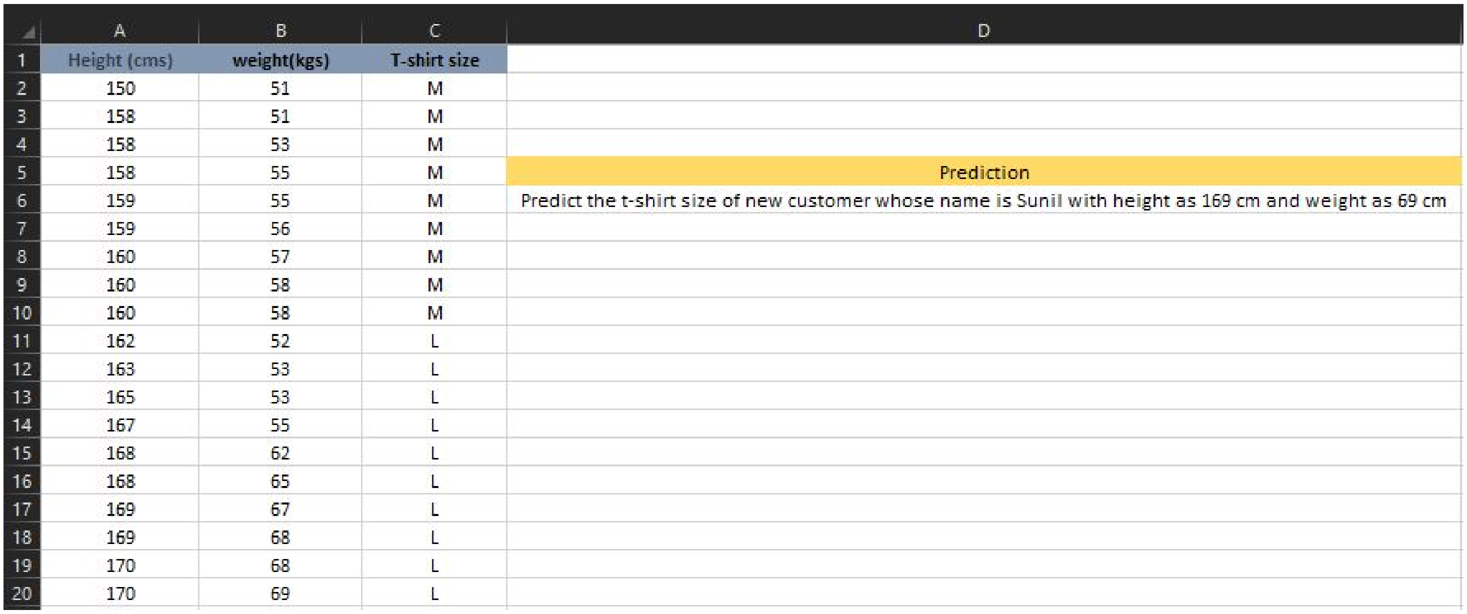
If k = 3 is checked for the labels of the three closest classes and the most common i.e. occurring at least twice, the label is assigned for larger k’s and so on.



Manual Implementation of KNN algorithm

Let's consider an example of height and weight

The below dataset in about height and weight of the customer with corresponding t-shirt size where M represents medium size and L represent large size. Now your task is to predict the t-shirt size for the new customer whose name is Sunil with height as 169 cm and weight as 69 kg.

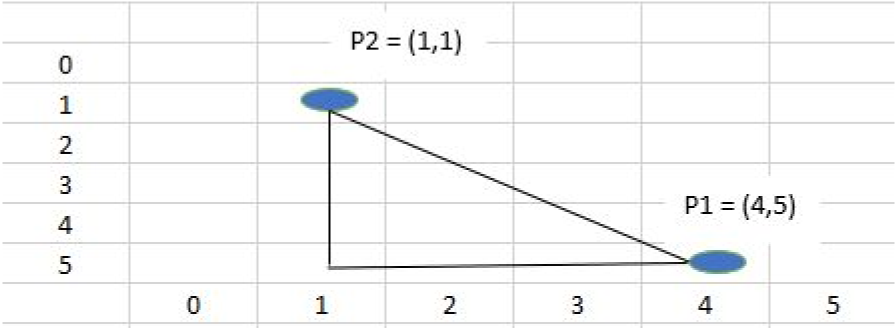


	A	B	C	D
1	Height (cms)	weight(kgs)	T-shirt size	
2	150	51	M	
3	158	51	M	
4	158	53	M	
5	158	55	M	Prediction
6	159	55	M	Predict the t-shirt size of new customer whose name is Sunil with height as 169 cm and weight as 69 cm
7	159	56	M	
8	160	57	M	
9	160	58	M	
10	160	58	M	
11	162	52	L	
12	163	53	L	
13	165	53	L	
14	167	55	L	
15	168	62	L	
16	168	65	L	
17	169	67	L	
18	169	68	L	
19	170	68	L	
20	170	69	L	

Step 1 : The initial step is to calculate Euclidean distance between the existing points and new points. For example the existing point is (4,5) and the new point is (1 , 1).

So, P1 = (4,5) where x1 = 4 and y1 = 5

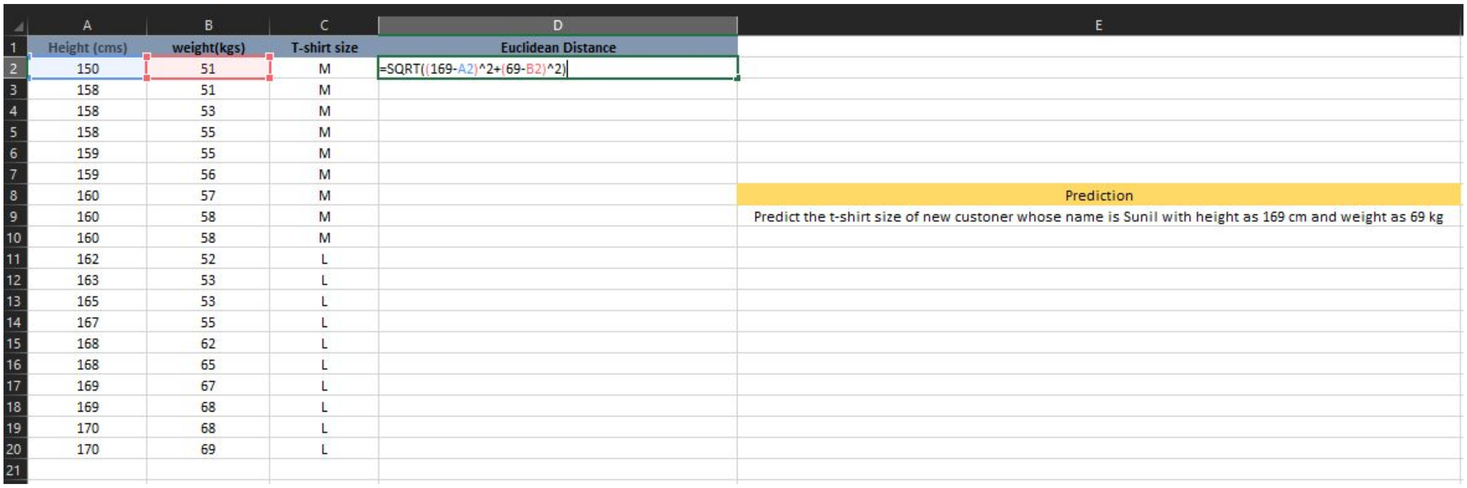
P2 = (1,1) where x2 = 1 and y2 = 1



Now Euclidean distance = $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

$= \sqrt{(1 - 4)^2 + (1 - 5)^2}$

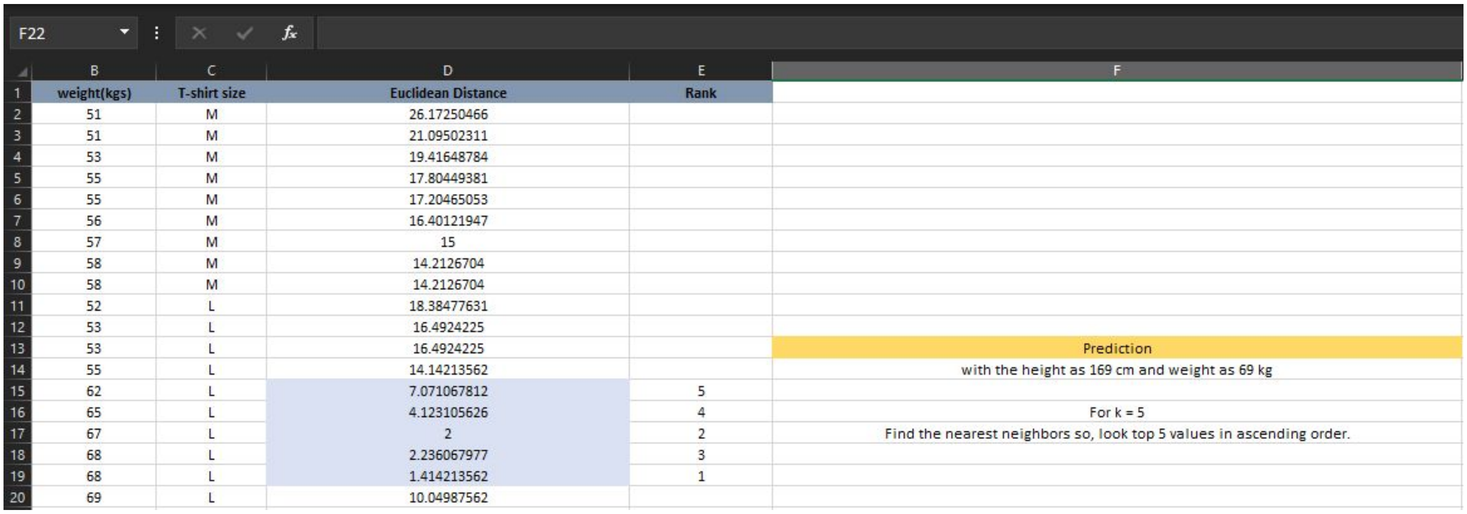
$= 5$



	A	B	C	D	E
1	Height (cms)	weight(kgs)	T-shirt size	Euclidean Distance	
2	150	51	M	=SQRT((169-A2)^2+(69-B2)^2)	
3	158	51	M		
4	158	53	M		
5	158	55	M		
6	159	55	M		
7	159	56	M		
8	160	57	M		Prediction
9	160	58	M		Predict the t-shirt size of new custoner whose name is Sunil with height as 169 cm and weight as 69 kg
10	160	58	M		
11	162	52	L		
12	163	53	L		
13	165	53	L		
14	167	55	L		
15	168	62	L		
16	168	65	L		
17	169	67	L		
18	169	68	L		
19	170	68	L		
20	170	69	L		
21					

Step 2: Second step is to choose the k value and select the closest k neighbors to the new item

So, in our case 5 elements have least Euclidean distance as compared with others.



F22

	B	C	D	E	F
1	weight(kgs)	T-shirt size	Euclidean Distance	Rank	
2	51	M	26.17250466		
3	51	M	21.09502311		
4	53	M	19.41648784		
5	55	M	17.80449381		
6	55	M	17.20465053		
7	56	M	16.40121947		
8	57	M	15		
9	58	M	14.2126704		
10	58	M	14.2126704		
11	52	L	18.38477631		
12	53	L	16.4924225		
13	53	L	16.4924225		Prediction
14	55	L	14.14213562		with the height as 169 cm and weight as 69 kg
15	62	L	7.071067812	5	
16	65	L	4.123105626	4	For k = 5
17	67	L	2	2	Find the nearest neighbors so, look top 5 values in ascending order.
18	68	L	2.236067977	3	
19	68	L	1.414213562	1	
20	69	L	10.04987562		

Step 3: Count the votes of least distance i.e. Euclidean distance of the predicting values to calculate k neighbors

SInce, K = 5, we have 5 t-shirts of size L. So according to this reason new customer name Sunil with height 169 cm and weight as 69 kg will fit into t-shirts of L size.

PYTHON:

Some of the major steps are listed below while implementing KNN algorithm using Python are listed below:

1. Data handling
2. Distance Calculation
3. Finding K nearest point
4. Predict the class
5. Check the accuracy

Step 1: The “iris” dataset is handled in the very first phase where open function opens the data collection and reader function is used to read the data lines available under the CSV module.