

kNN



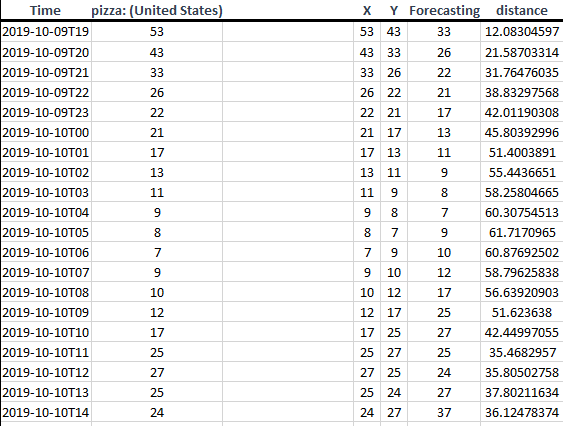
Course: ALY 6020 Predictive Analytics

Instructor: Dr. Marco Montes de Oca

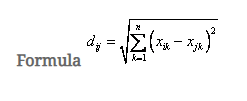
Submitted by: Nithish Saravanan

k-NN is a machine learning algorithm which is used for classification as we as regression problems. This finds the maximum significant and closest data points from the dataset. This model is generally based on two important parameters which are the k value and Euclidean distance.

First we need to create lags for Y and forecasting. Now the Euclidean Distance is calculated between the actual value and the lag1 value using the excel formula of =SQRT((D2-$D$168)^2+(E2-$E$168)^2). This is calculated for all the rows of data in the dataset.

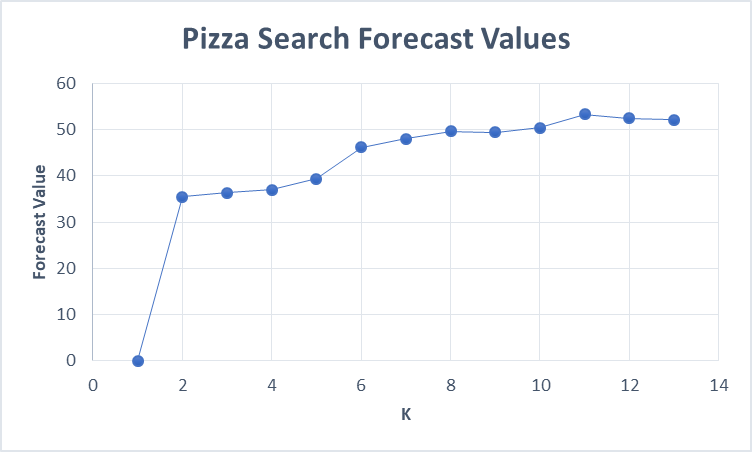


The Euclidean distance is the distance form one point to another point in the graph can be calculated manually using



Now we will create a new column with kNN to find the distance based upon the k values that are substituted. This can be done in excel using =IF(G2<=SMALL($G$2:$G$170,$K$5),F2,"") for all the rows and various other ‘k’ values.

The value for pizza search with k value as 13 give the optimal forecast of the next search value. This can be graphed to look at the values which can be considered optimized.



For the k value at 13 we get the predicted value for pizza search at 07:00 PM as 52.07692 that can be rounded off to 53 searches that can occur for the time which we are predicting.

Reference:

Data Source: Google Trends data on pizza searches in United States. <https://trends.google.com/trends/explore?date=now%207-d&geo=US&q=pizza>