In Homework 2, I created a python program that used the libraries of Pandas and Numpy to read columns from two csv files containing data of hand-written digits converted into pixel mapping of 28 x 28 (784 pixels in total). On of the data sets is reserved for training data to calculate the Euclidean Distance from the points in the second dataset, the test data set, and stores it into a list of values. From this, the nearest neighbors are calculated by sorting the values of the Euclidean distances by the max values. Using the given K value, the majority of the K-nearest neighbors are decided and used to predict the values of the test data set. When the predictions are made for each of the points, an accuracy measurement is calculated by comparing the test data row’s actual value with the predicted value and counting all the correct predictions and dividing by the total number of data rows in the test dataset.

Execution of python code:

Using Python 3.9, this code can be executed by calling

python KNN\_project.py

From this the script will run and will calculate the KNN of the datasets with a K of 50 and print:

“Main Function is Running ... “

Estimated Accuracy: \*Accuracy of predictions\*