Handwritten Digit Recognition (E4CO02 – 2223Evesem)

Aim:

Through this activity titled 'Handwritten Digit Recognition', we aim to learn how to identify handwritten digits present in the input images using Machine Learning/Deep Learning.

Objectives:

Part 1: is to train the model for values of 'K' as 3, 5, 7, 9, 11 or any 5 'K' values using KNN classifier on the train.csv training dataset.

Part 2: aims to test the trained model on the testing dataset test.csv, creating the confusion matrix for the same and to evaluate the testing performance of the model using precision, recall and F1 score.

The solution to the problem of handwritten digit recognition was tackled based on the prototype provided to us.

Solution:

The following steps account for the pre-processing section of the solution:

The first and foremost step is to load in the dataset available to us. The next step is to apply some pre-processing to the data to ease the data-reading and model-building process.

Then we check how well-balanced our dataset is i.e., all the possible values of the output are present in adequate amounts which we check using seaborn's 'countplot' function and with that, the pre-processing of the dataset is finished.
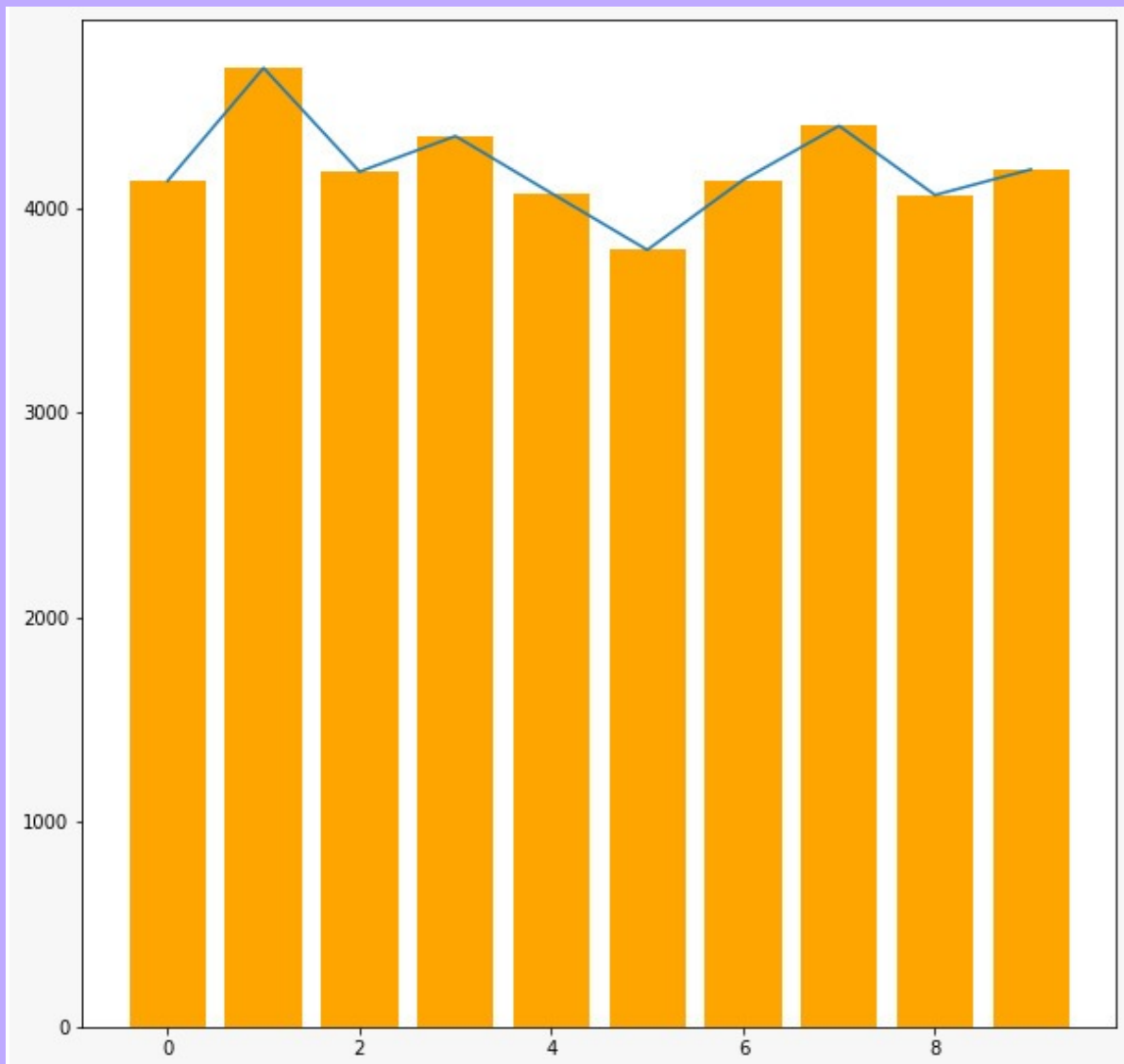
This plot show the amount of data entries pertaining to each of the 10 Arabic numerals used in the world. The x-axis is the value of the digit, and the y-axis shows the number of entries.

These steps were followed for the creation of the ML model which tackles the issue of handwritten digit recognition:

First, we import the necessary python libraries for ML model creation which is 'sklearn' and since we are using the KNN model to tackle our problem, we also import 'KNeighboursClassifier'.

Finally, to make the various modes for different K values, we make a loop and plug the required value of 'K' per iteration of the loop.