

CS 412 HW-1 SUMMARY REPORT**Link to the Colab Notebook:**

<https://colab.research.google.com/drive/1OsHY8zCbU0LYfMNIzjmhxXlu5jDzsoSx?usp=sharing>

This report is about the development of a kNN classifier for the MNIST dataset. It also includes hyperparameter tuning with some given values of kNN classifier parameter k. With the k value giving the best accuracy score, a final classifier was implemented.

To be able to train the kNN classifier provided by the Scikit-learn library, the MNIST dataset has been loaded to Colab. This dataset comes in two parts: train and test, including 60000 and 10000 samples respectively, which means roughly 85% of samples are to be used for training and 15% for testing.

After reshaping these two sets from 2D image arrays to 1D feature arrays, so that they can be used with the kNN classifier, and shuffling them, the training set was split into train/development and validation sets. The aim of this split is to be able to perform hyperparameter tuning. 80% of the MNIST training set was separated for train/development and the other 20% for validation. Split was performed with the *train_test_split* method provided by the Scikit-learn library.

Next, hyperparameter tuning took place. For different values of meta-parameter k that can be seen in the table below, a kNN classifier was trained with the train/development set and tested with the validation set. Corresponding accuracy scores of different k values can also be seen in the table below.

K VALUES	ACCURACY SCORES
K = 1	0.971166...
K = 3	0.970833...
K = 5	0.9699166...
K = 7	0.96766...
K = 9	0.96625
K = 11	0.9655
K = 13	0.96475

We have obtained the best results on the validation set by using a value of 1 for meta-parameter k.

After the hyperparameter tuning, the chosen value of k was used to implement the final version of the kNN classifier. To train the final model, train/development and validation sets were combined to be able to train the final model with a larger training set. Then, the final kNN classifier was trained with this enhanced set and tested with the MNIST testing set. The accuracy score of the final model is recorded as 0.9691.