Handwritten Digit Recognizer Using MNIST Dataset Report

Name: Ali Amr Nour Mokhtar

ID: 20190339

Outline:

- Data and libraries.
- Preprocessing.
- Model.

Data and Libraries:

- That Data used is the MNIST data set obtained from this [Kaggle Competition](#).
- Libraries used are NumPy, Pandas, Sci-Kit Learn and the python math module.

Preprocessing:

The Data used is in form an excel sheet, where each row is an image and the rows contain the pixels of the image, each image is 28x28 pixels.

Each image is divided into sixteen sub-images, and for each sub-image the centroids are calculated, then all centroids are compiled into a single 32-dimensonal feature vector for the image.

The entire data set is preprocessed in such manner, and the results are compiled into a single .csv file.

Model:

The model used is the Kth Nearest Neighbor Classifier, implemented from scratch.

Results:

The model gave an accuracy of 88.5%.