Final Project Report

CMSC 6950

MNIST Hand Written Digit Classification using KNN

Submitted By: Shemonto Das

ID: 202193149

Date: 17th June, 2022

Description of the Dataset: I have used the MNIST csv dataset for classification using KNN. The training data contains 60,000 rows and 785 columns among which a single row specifies the original label. A sample of the training data is as follows

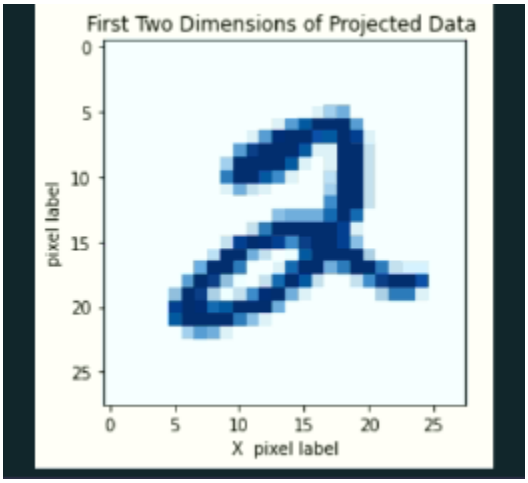
	label	1x1	1x2	1x3	1x4	1x5	...	28x23	28x24	28x25	28x26	28x27	28x28
0	5	0	0	0	0	0	...	0	0	0	0	0	0
1	0	0	0	0	0	0	...	0	0	0	0	0	0
2	4	0	0	0	0	0	...	0	0	0	0	0	0
3	1	0	0	0	0	0	...	0	0	0	0	0	0
4	9	0	0	0	0	0	...	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
59995	8	0	0	0	0	0	...	0	0	0	0	0	0
59996	3	0	0	0	0	0	...	0	0	0	0	0	0
59997	5	0	0	0	0	0	...	0	0	0	0	0	0
59998	6	0	0	0	0	0	...	0	0	0	0	0	0
59999	8	0	0	0	0	0	...	0	0	0	0	0	0

[60000 rows x 785 columns]

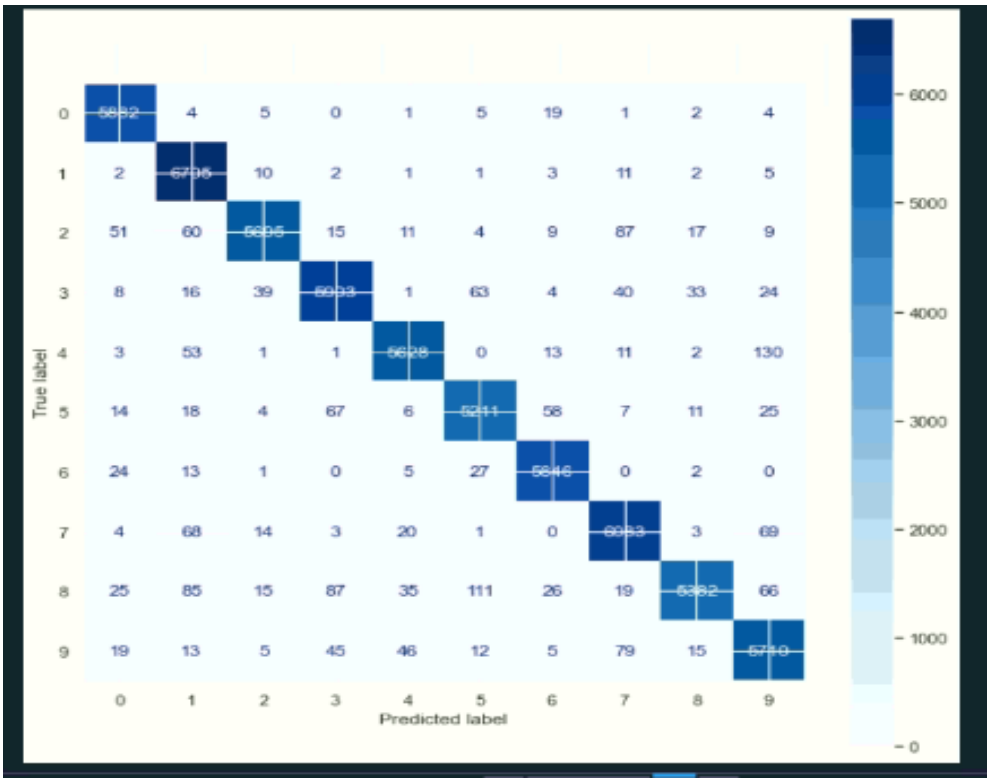
Project Orientation: I have 5 classes and a main class that corresponds to the goal of this project and they are

- Load_data.py
- Visual_data.py
- Training.py
- Augmentation.py
- Evaluate_data.py
- Main.py

Visualizing the Data: The digit is being plot using matplotlib by reshaping the array.



Training and Evaluating Model: The model is being trained using the KNN classifier from Sklearn. The evaluation is done by creating a confusion matrix.



As we can see from the classification report the prediction accuracy through different metrics.

	precision	recall	f1-score	support
0	0.975133	0.993078	0.984023	5923
1	0.953092	0.994512	0.973361	6742
2	0.983762	0.955858	0.969609	5958
3	0.964070	0.962812	0.963441	6131
4	0.978102	0.963369	0.970680	5842
5	0.958786	0.961262	0.960022	5421
6	0.977102	0.987834	0.982438	5918
7	0.959766	0.970950	0.965326	6265
8	0.984092	0.919843	0.950883	5851
9	0.945051	0.959825	0.952381	5949
accuracy			0.967417	60000
macro avg	0.967896	0.966934	0.967216	60000
eighted avg	0.967676	0.967417	0.967345	60000