KNN FOR HANDWRITTEN DIGIT RECOGNITION:

FIVE-FOLD CROSS VALIDATION

- Five-fold cross validation was performed by taking K = 1, 3 and K = 5. The results were similar for all K values.
- At each fold, the training data is divided into two sets.
- The first set has 33600 images for training purpose
- The second set has 8400 images for testing.

```
*** FOLD  1  ***

Training_index:  [ 8400  8401  8402 ... 41997 41998 41999]
Testing_index:  [   0    1    2 ... 8397 8398 8399]
Size of the training data:  33600
Size of the testing data:  8400


*** FOLD  2  ***

Training_index:  [    0     1     2 ... 41997 41998 41999]
Testing_index:  [ 8400  8401  8402 ... 16797 16798 16799]
Size of the training data:  33600
Size of the testing data:  8400


*** FOLD  3  ***

Training_index:  [    0     1     2 ... 41997 41998 41999]
Testing_index:  [16800 16801 16802 ... 25197 25198 25199]
Size of the training data:  33600
Size of the testing data:  8400


*** FOLD  4  ***

Training_index:  [    0     1     2 ... 41997 41998 41999]
Testing_index:  [25200 25201 25202 ... 33597 33598 33599]
Size of the training data:  33600
Size of the testing data:  8400


*** FOLD  5  ***

Training_index:  [    0     1     2 ... 33597 33598 33599]
Testing_index:  [33600 33601 33602 ... 41997 41998 41999]
Size of the training data:  33600
Size of the testing data:  8400
```

Evaluation Metrics over Five-Folds

- Accuracy, Precision, Recall and F1 score over five-folds when K = 1, 3 and 5

K = 1

KNN Evaluation Metrics:

Accuracy	Precision	Recall	F1 Score
0.966548	0.966738	0.966066	0.966304
0.967738	0.968017	0.967327	0.967535
0.9625	0.962569	0.962128	0.962246
0.966667	0.967023	0.966421	0.966583
0.967381	0.96785	0.966757	0.967165

Average metrics over five folds:

The average accuracy is: 0.9662
The average precision is: 0.9664
The average recall is: 0.9657
The average f1_score is: 0.9660

K = 3

KNN Evaluation Metrics:

Accuracy	Precision	Recall	F1 Score
0.97	0.970315	0.9696	0.969857
0.967262	0.967701	0.966725	0.966979
0.962738	0.963169	0.9623	0.962517
0.96381	0.964318	0.963501	0.963685
0.969524	0.970346	0.968673	0.969279

Average metrics over five folds:

The average accuracy is: 0.9667
The average precision is: 0.9672
The average recall is: 0.9662
The average f1_score is: 0.9665

K = 5

KNN Evaluation Metrics:

Accuracy	Precision	Recall	F1 Score
0.967857	0.968478	0.967389	0.967724
0.965238	0.965893	0.964614	0.965009
0.963571	0.964114	0.963134	0.963401
0.963214	0.963774	0.962926	0.96307
0.969048	0.969859	0.968245	0.968836

Average metrics over five folds:

The average accuracy is: 0.9658
The average precision is: 0.9664
The average recall is: 0.9653
The average f1_score is: 0.9656

Digit Frequencies

K = 1

Actual Digit Frequencies

Digit	Count
0	4132
1	4684
2	4177
3	4351
4	4072
5	3795
6	4137
7	4401
8	4063
9	4188

Predicted Digit Frequencies

Digit	Count
0	4184
1	4811
2	4094
3	4347
4	4001
5	3799
6	4172
7	4476
8	3864
9	4252

K = 3

Actual Digit Frequencies

Digit	Count
0	4132
1	4684
2	4177
3	4351
4	4072
5	3795
6	4137
7	4401
8	4063
9	4188

Predicted Digit Frequencies

Digit	Count
0	4213
1	4869
2	4094
3	4383
4	4000
5	3790
6	4172
7	4466
8	3790
9	4223

K = 5

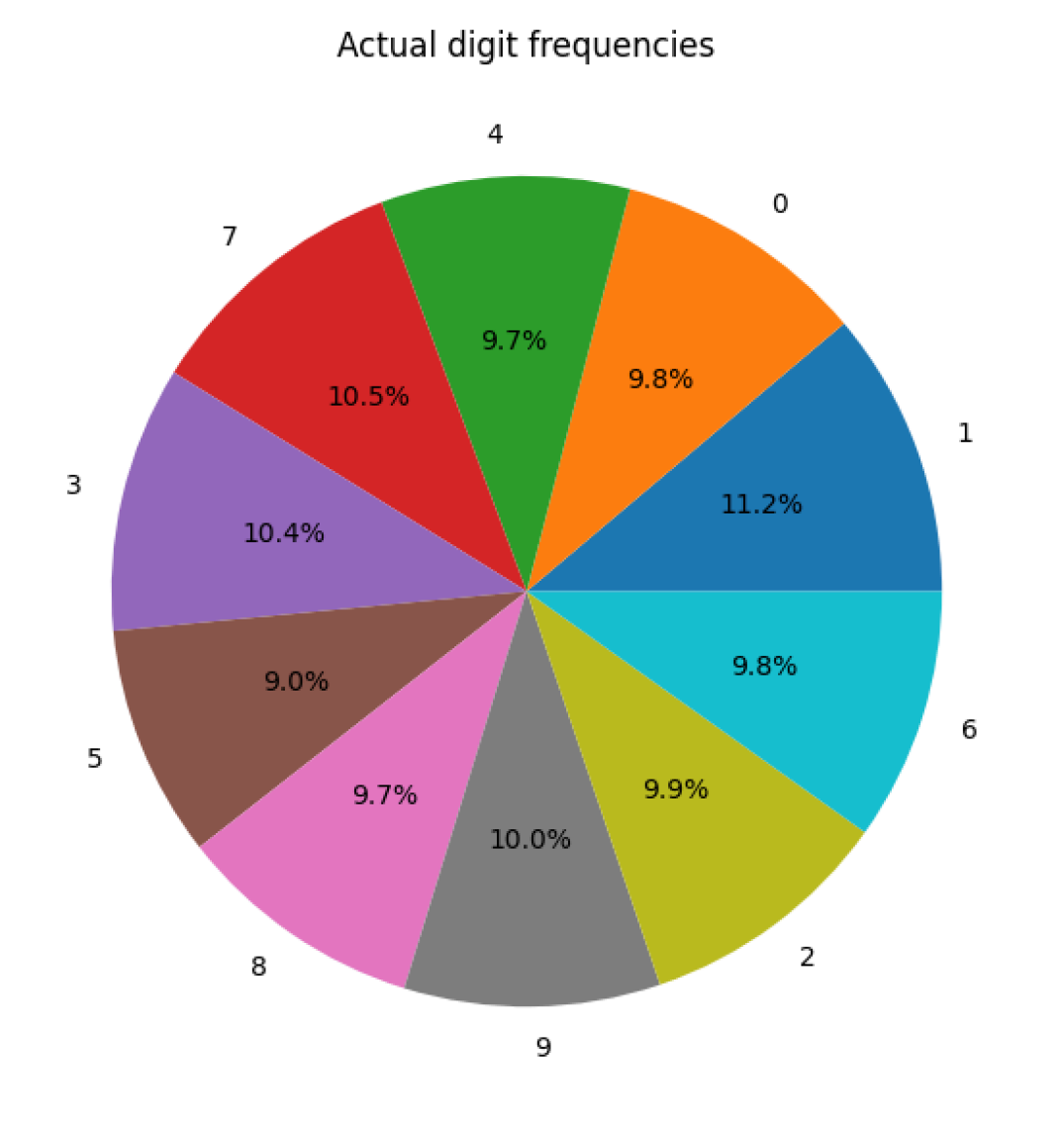
Actual Digit Frequencies

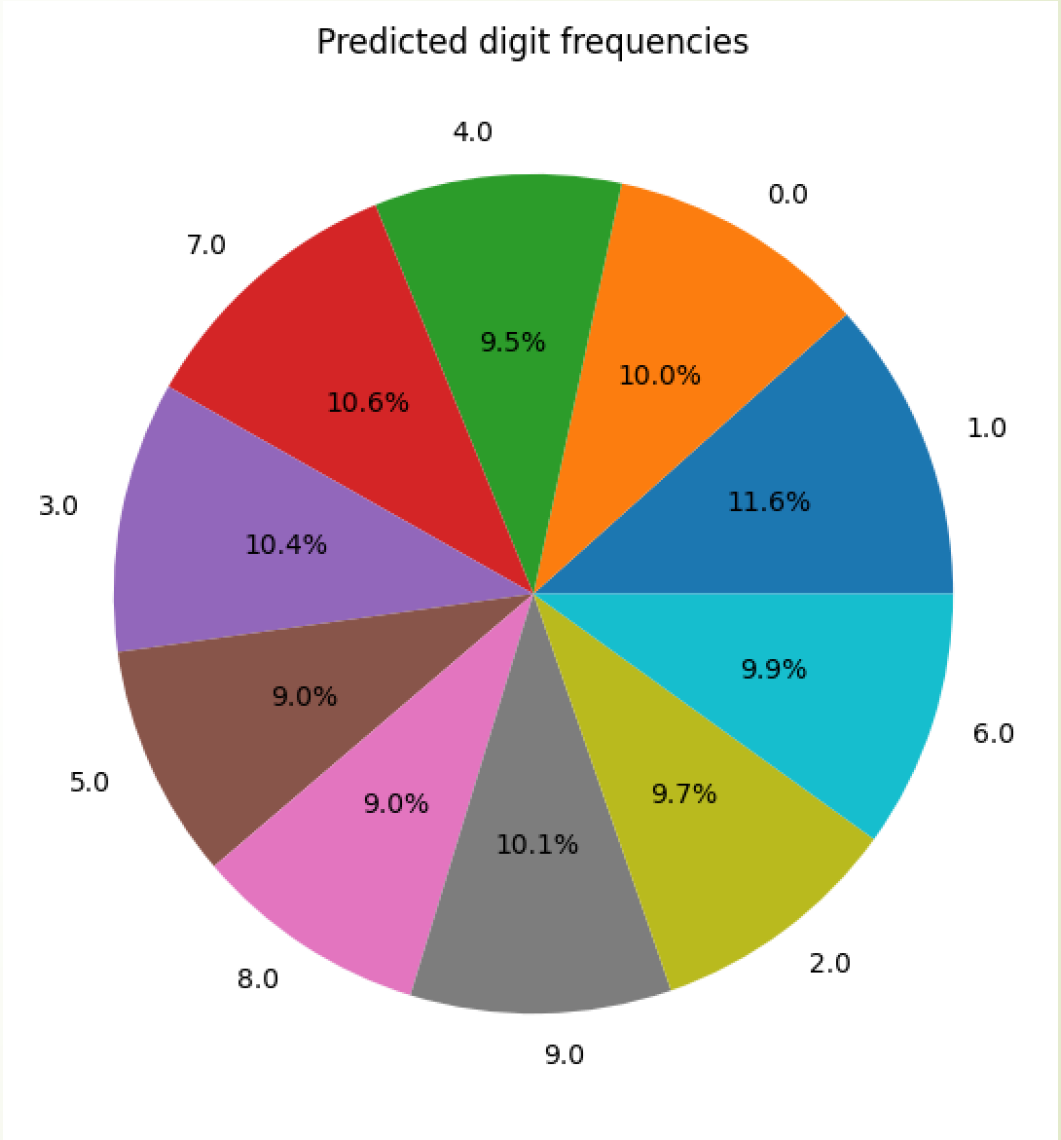
Digit	Count
0	4132
1	4684
2	4177
3	4351
4	4072
5	3795
6	4137
7	4401
8	4063
9	4188

Predicted Digit Frequencies

Digit	Count
0	4212
1	4905
2	4055
3	4379
4	4002
5	3786
6	4185
7	4470
8	3780
9	4226

Digit Frequencies when K = 3





Predictions made on the Testing data:

- Screenshots from the output CSV file:
- These screenshots show the first few and last few predictions from the output CSV file when K = 3.
- Comparing my result with 100% output accuracy, I got an accuracy of 97% (K = 1, K = 3 and then K = 5 respectively) as shown in the screenshot below.

```
Accuracy of the testing dataset: 97.0143 %
```

```
Accuracy of the testing dataset: 96.8036 %
```

```
Accuracy of the testing dataset: 96.7000 %
```

My GitHub link: Project_Checkpoint_2_submission:

GitHub Link

ImageId	label
0	2
1	0
2	9
3	9
4	3
5	7
6	0
7	3
8	0
9	3
10	5
11	7
12	4
13	0
14	4
15	3
16	3
17	1
18	9
19	0
20	9
21	1
22	1
23	5
24	7
25	4
26	2
27	7
28	4
29	7
30	7

27969	3
27970	5
27971	0
27972	4
27973	8
27974	0
27975	3
27976	6
27977	0
27978	1
27979	9
27980	3
27981	1
27982	1
27983	0
27984	4
27985	5
27986	2
27987	2
27988	9
27989	6
27990	7
27991	6
27992	1
27993	9
27994	7
27995	9
27996	7
27997	3
27998	9
27999	2