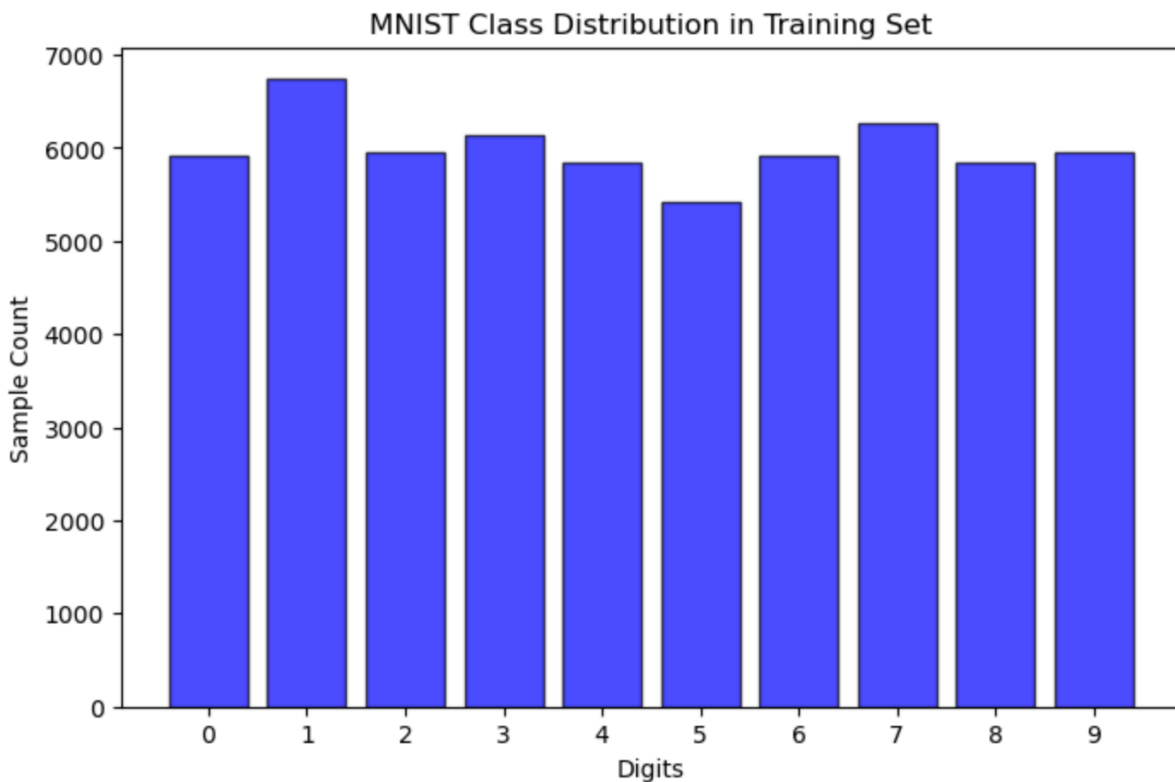


CS 412 HW1 FINAL REPORT

Ataberk Çakır - 30839

Link to the Colab Notebook: <https://github.com/ataberkcakir15/CS-412/blob/main/CS412-HW1-AtaberkCakir.ipynb>

This report is about classifying the MNIST handwritten digit dataset using two different machine learning algorithms: kNN and Decision Trees. The data underwent initial data loading to Colab and reshaping from 2D image arrays to 1D feature arrays to make them compatible with our k-NN classifier. This dataset comes with training and testing sets, 60000 and 10000 respectively. In order to split the data into training and validation subsets, the data randomly shuffled and used `train_test_split` from sklearn library. Specifically, 80% of the training samples were allocated for model development and 20% for validation, ensuring the dedication of the set for hyperparameter tuning before final evaluation. The class distribution proved relatively balanced, with each digit (0-9) represented roughly 5000 - 7000 examples.



Then calculated the mean (~ 33.32) and standard deviation (~ 78.57) of the pixel values. These processes are followed by Data Preprocessing. Each 28x28 image is

flattened into a 784-dimensional vector and normalized the pixel values from [0-255] to [0-1].

Next, hyperparameter tuning is handled. To find the best k value, multiple candidates of meta-parameter k, a k-NN classifier was trained with the training set and measured performance with validation set. The k values and corresponding accuracy scores can be found in the table below.

Validation Accuracies Table:

K Value	Validation Accuracy
1	0.9700
3	0.9707
5	0.9681
7	0.9682
9	0.9666

With the result obtained, it can be concluded that the best result on the validation set by using a value of 3 for meta-parameter k.

Following the performance measurement, final k-NN model is retrained with the combination of training and validation sets. Then evaluate it on the test set to measure its performance on unseen data. The table for each metric and corresponding values can be found below.

Final Model Evaluation (k-NN)

Metric	Value
Test Accuracy	0.9705
Test Precision	0.9709
Test Recall	0.9701
Test F1-score	0.9704

Furthermore, misclassifications are spotted using the confusion matrix. Misclassifications occur when the model incorrectly labels a digit and after training the k-

NN model, it is important to assess where exactly the model is failing. Confusion matrix helped finding which digits are confused with others. Those misclassified samples are located by checking where `pred != true`. Furthermore, matrix aided to spot frequent misclassifications as well. This means that certain digits might be predicted incorrectly more frequently due to their visual similarities like confusing “3” with “5”.



Then, similar experimental approach is applied to decision tree classifier to the one used for k-NN. Decision tree model is trained with MNIST dataset while tuning key hyperparameters. Key parameters, including the maximum depth (e.g., 2, 5, and 10) and the minimum number of samples required to split a node (e.g., 2 and 5), are systematically varied. Each configuration is evaluated using validation metrics such as accuracy, precision, recall, and F1-score, thereby enabling the identification of an optimal balance between model complexity and generalization.

```
max_depth= 2 | min_samples_split= 2 | Validation Accuracy: 0.3367
max_depth= 2 | min_samples_split= 5 | Validation Accuracy: 0.3367
max_depth= 5 | min_samples_split= 2 | Validation Accuracy: 0.6693
max_depth= 5 | min_samples_split= 5 | Validation Accuracy: 0.6693
max_depth=10 | min_samples_split= 2 | Validation Accuracy: 0.8435
max_depth=10 | min_samples_split= 5 | Validation Accuracy: 0.8428
```

```
=== Best Configuration ===
Max Depth:          10.0
Min Samples Split:  2.0
Validation Accuracy: 0.8435
```

Once the best-performing hyperparameters are determined, the decision tree is retrained on the full training set and its performance is subsequently assessed on the test set.

=== Final Model Evaluation ===

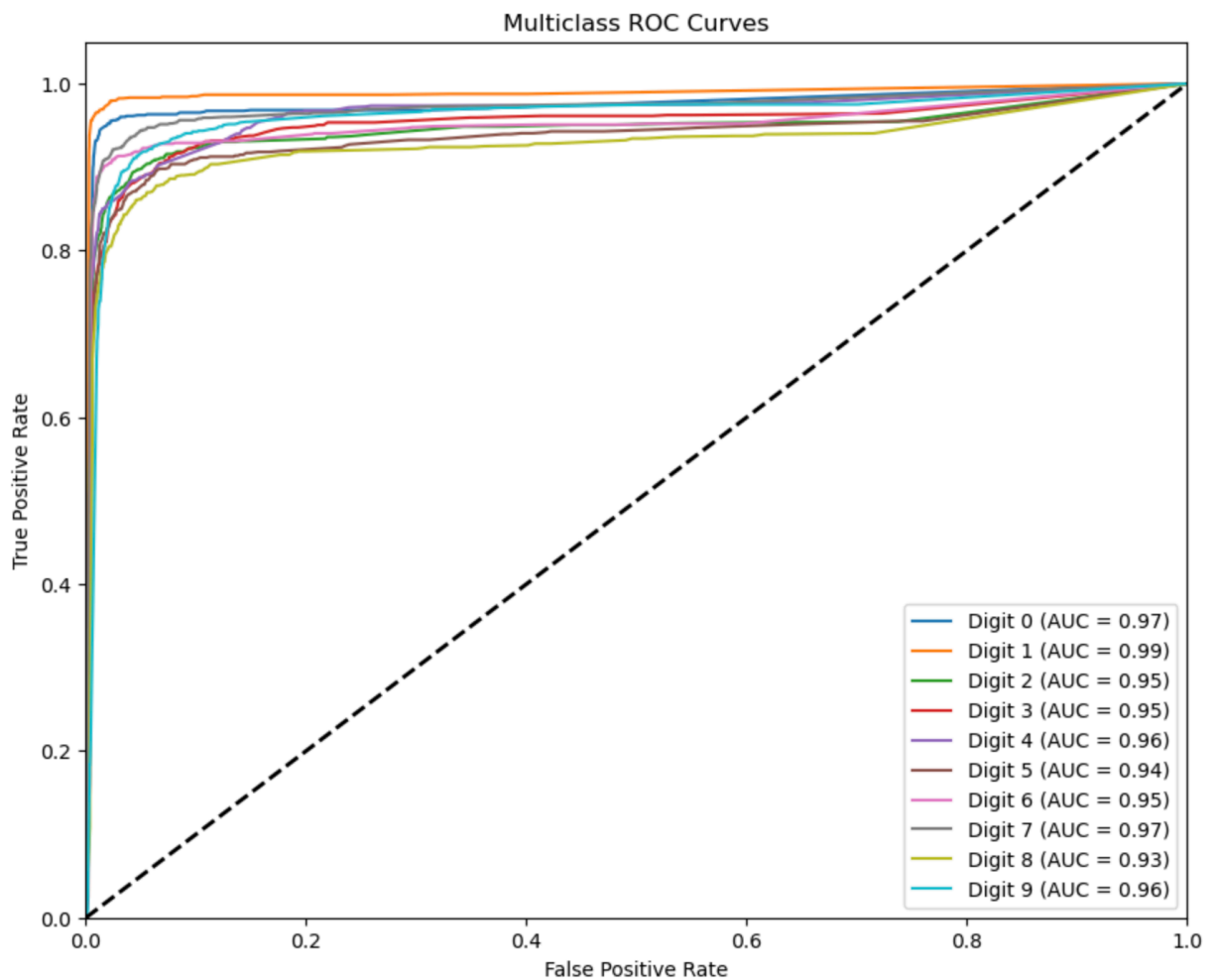
Test Accuracy: 0.8663

Test Precision: 0.8656

Test Recall: 0.8644

Test F1-score: 0.8648

The resulting confusion matrix provides insight into which digits are most frequently misclassified, while ROC curves for each digit class—displayed on a single graph with corresponding AUC scores—offer a comprehensive view of the model's discriminative capability. This rigorous evaluation demonstrates the decision tree's effectiveness on the handwritten digit recognition task and serves as a valuable benchmark against the k-NN classifier.



From this plot, it can be seen that the model demonstrates high overall performance, as indicated by the AUC (Area Under the Curve) values exceeding 0.90 for each digit. However, there is a slight performance variation across classes. For example, one digit might have an AUC of 0.99, showing almost perfect discrimination, while another with an AUC of around 0.93. This indicates that the model still performs well but may be more prone to misclassification. The ROC curves also give us insight into how changing the decision threshold affects the balance between correctly identified positives and false alarms. In general, the curves' overall proximity to each other and to the top-left corner suggests a consistent classification capability across all digits.