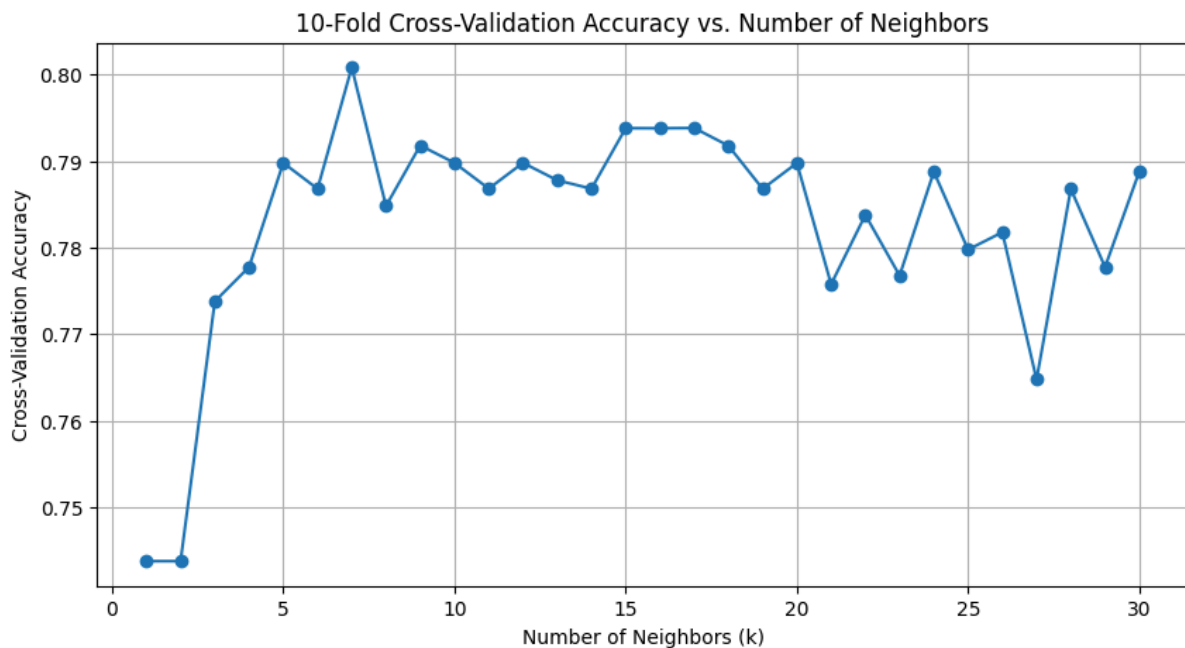


Digit Classifier

1. A graph that shows the average accuracy based on 10-fold cross validation when varying the number of neighbours from 1 to 30.



2. The best number of neighbours found by 10-fold cross validation and its cross-validation accuracy.

The K-Fold Cross-Validation Accuracy vs. k Graph highlights how the choice of K impacts KNN performance. The best K is 7, where the model achieves the highest accuracy, balancing complexity and generalization. At K=1, the model overfits, memorizing the training data and becoming too sensitive to noise, leading to lower accuracy. Conversely, at K=27, the model underfits by averaging too many neighbors, making it too simplistic and unable to capture meaningful patterns. The decline in accuracy at higher K-values confirms that choosing an optimal K is crucial for model performance.

3. The test accuracy based on the best number of neighbours.

The model was tested using K = 7, resulting in an accuracy of 0.7798. While test accuracy may not always match validation accuracy exactly, it typically remains close. Cross-validation provides a reliable estimate of the optimal K-value, helping to select a model that generalizes well and achieves the best average performance on unseen data.

Fake News Detection

1. Outputs of the 5 max depth values used to validate the model

The 5 depth values we decided to use was 3, 5, 10, 15, 30

The accuracies are:

max_depth = 3, Validation Accuracy = 0.6939

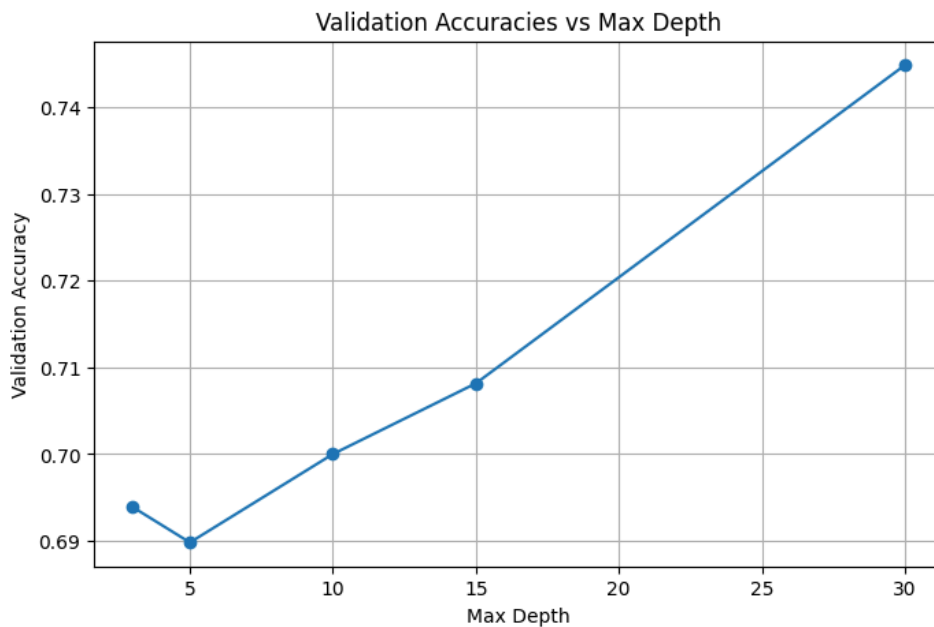
max_depth = 5, Validation Accuracy = 0.6898

max_depth = 10, Validation Accuracy = 0.7000

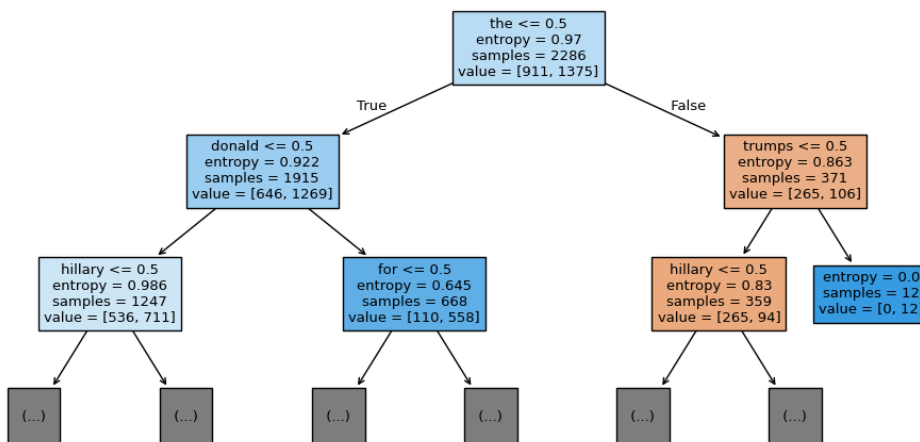
max_depth = 15, Validation Accuracy = 0.7082

max_depth = 30, Validation Accuracy = 0.7449

2. Plot of the validation accuracy vs max_depth



3. Visualize the first two layers of the tree



Here we used scikit-learn's tree-plot library to get a visual of the first 2 layers of the decision tree.