Introduction to ML – Decision Tree Coursework COMP70050

Output of the tree visualisation function

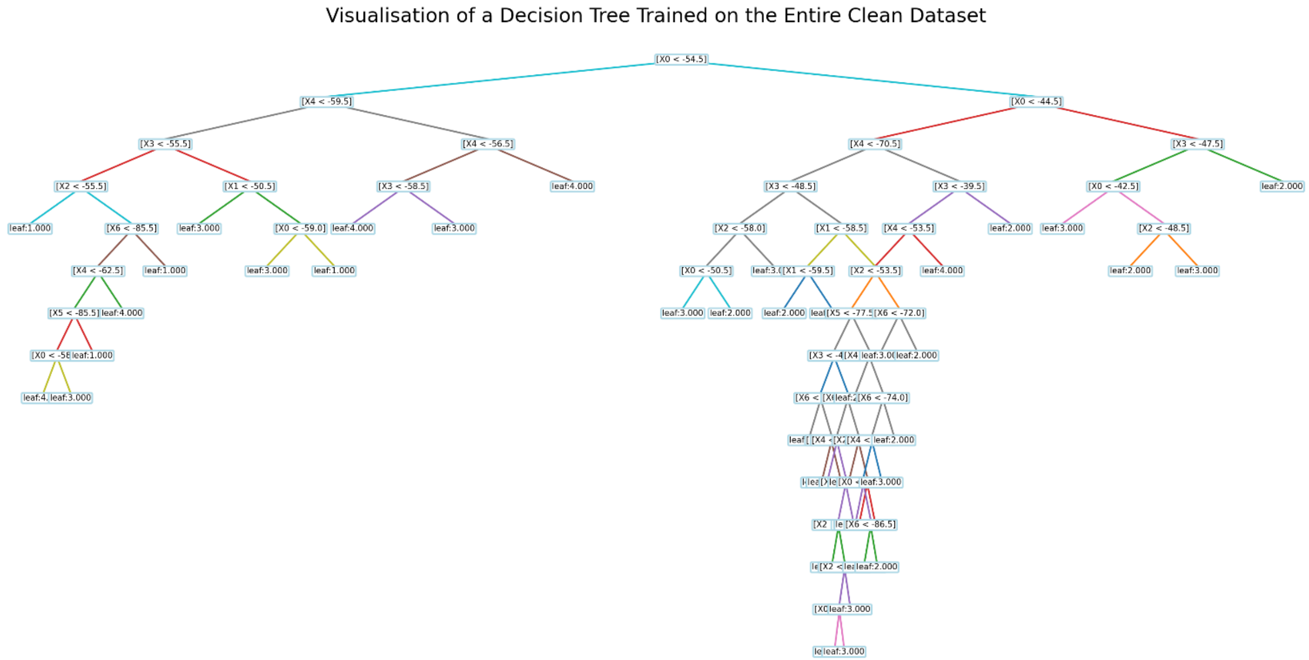

Figure 1: Visualisation of a decision tree trained on the entire clean dataset.

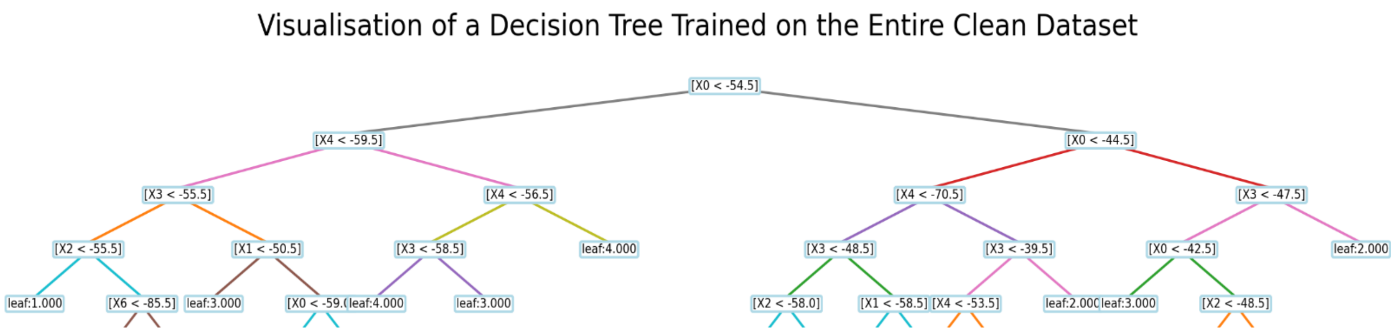

Figure 2: Visualisation of a decision tree, a fragment of the decision tree trained on the entire clean dataset is displayed.

Evaluation before pruning

Cross validation classification metrics

Confusion matrix values are presented as floating-point numbers with one decimal precision to reflect the minor differences in results. Rounding to integers would result in a loss of valuable details.

Clean dataset

Accuracy: 97.3%

Average depth: 12.2

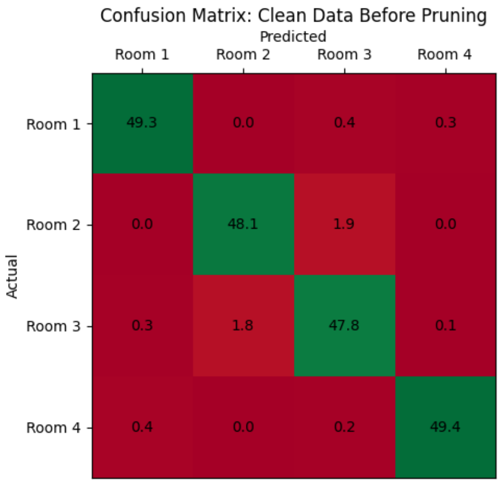

Figure 3: Measures per class for clean data before pruning.

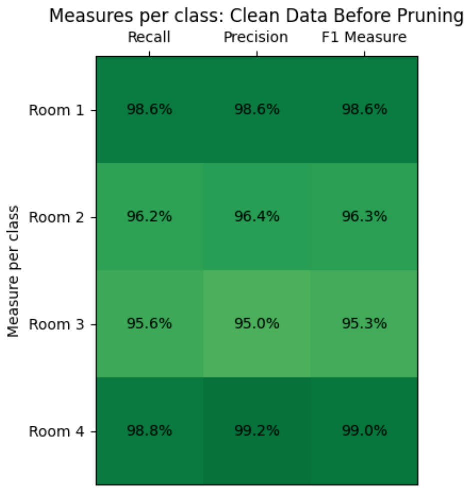

Figure 4: Confusion matrix for clean data before pruning.

Noisy Dataset

Accuracy: 80.5%

Average depth: 18.4

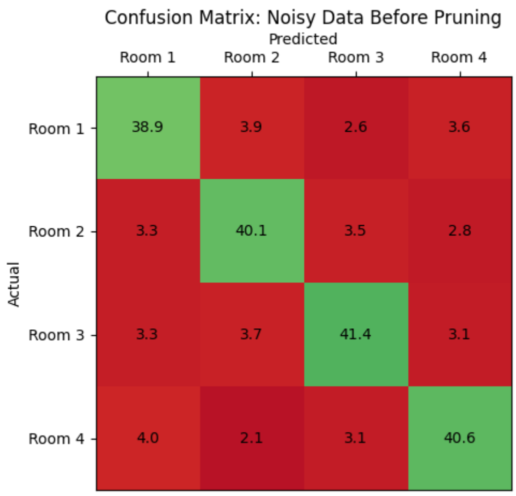

Figure 5: Confusion matrix for noisy data before pruning.

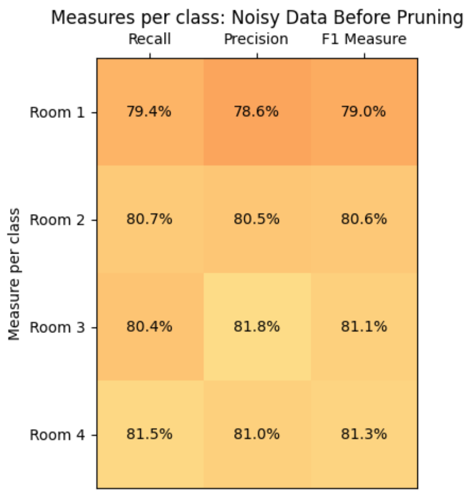

Figure 6: Measures per class for noisy data before pruning.

Result analysis

Clean dataset:

All cross-validation metrics (accuracy, precision, recall, and F1) exceed 95% for all rooms which is highly promising (Figure 4). Rooms 1 and 4 are recognised with very high precision (above 98%), while Rooms 2 and 3 show slightly lower scores, and appear to be minorly confused with each other (Figures 3-4).

Noisy dataset:

The same trend is less pronounced using noisy data. Room 3 achieves the highest recognition rate (81.8%), whereas Room 1 has the lowest performance, with precision of 78.6% (Figures 5-6).

Dataset differences

The noisy dataset shows a significantly lower performance, with all metrics below 82%, compared to over 95% for the clean dataset. This decline is due to noise causing misclassifications between classes, indicating the model's sensitivity to small changes which substantially influence the model's performance. Therefore, there is potential for improved generalisation.

Evaluation – after pruning

Cross validation classification metrics

Clean dataset

Accuracy: 96.7%

Average depth: 8.6

Figure 7: Confusion matrix for clean data after pruning.

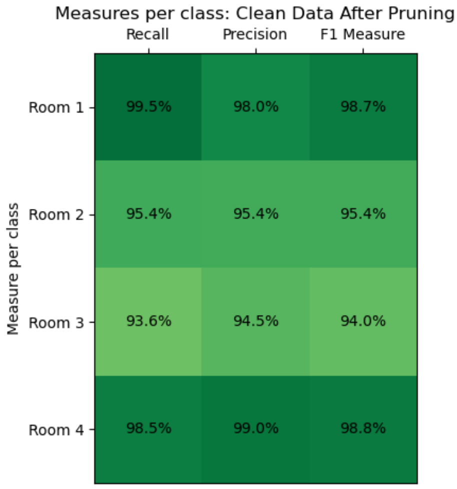

Figure 8: Measures per class for clean data after pruning.

Noisy Dataset

Accuracy: 87.1%

Average depth: 13.8

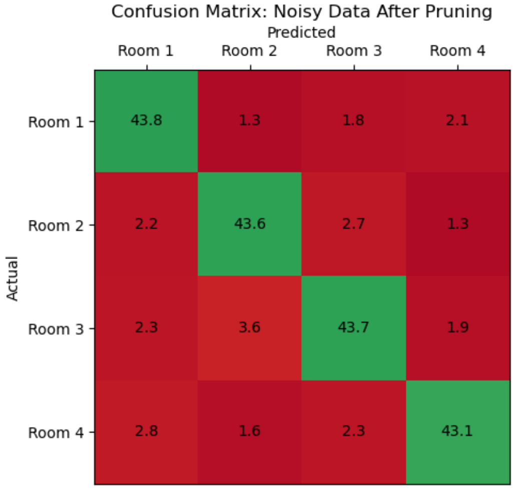

Figure 9: Confusion matrix for noisy data after pruning.

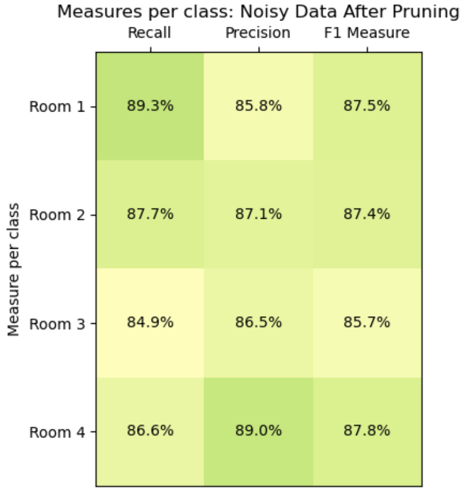

Figure 10: Measures per class for noisy data after pruning.

Result analysis after pruning

For the clean dataset, average accuracy reduced only slightly from 97.3% to 96.7% with an average depth of 8.6. As shown in Figures 7-8, all rooms still performed strongly, with F1-scores above 94% and with a similar pattern to the results for clean data before pruning. Room 1 achieved the highest recall (99.5%) and precision (98.0%). For the noisy dataset, pruning improved stability from 80.5% to 87.1% accuracy with an average depth of 13.8. Rooms 1-4 maintained balanced precision and recall (≈ 85-89%), reducing overfitting.

Depth analysis

Clean trees were shallower and more accurate than those trained on noisy data. After pruning, depth decreased from 12.2 to 8.6 for the clean dataset and from 18.4 to 13.8 for the noisy dataset. Accuracy remained high for the clean data but improved notably for the noisy data. Pruning simplified the trees, reducing depth and therefore overfitting. This made little difference to the clean dataset, but a significant difference in the noisy dataset where small changes substantially influenced the models performance. Here, greater depth correlated with lower accuracy (Figure 11).

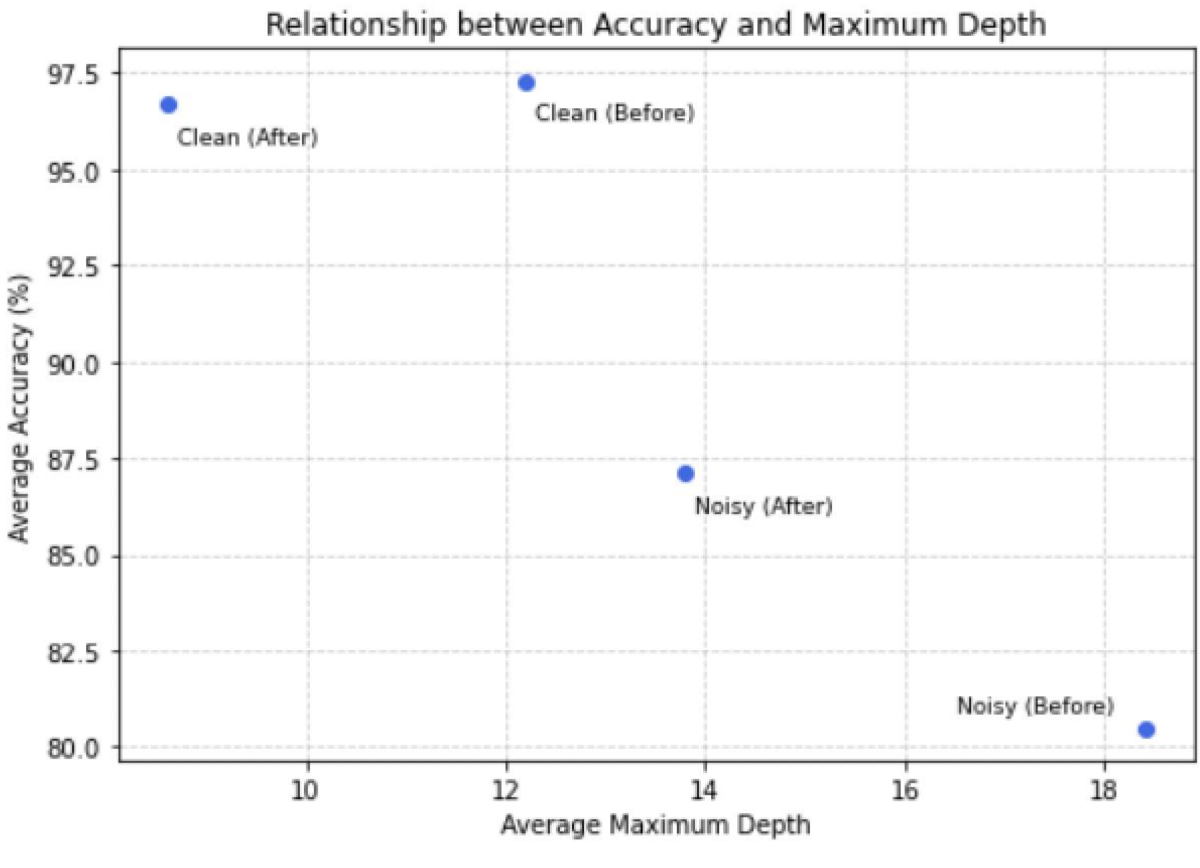

Figure 11: Relationship between average accuracy and average maximum depth.