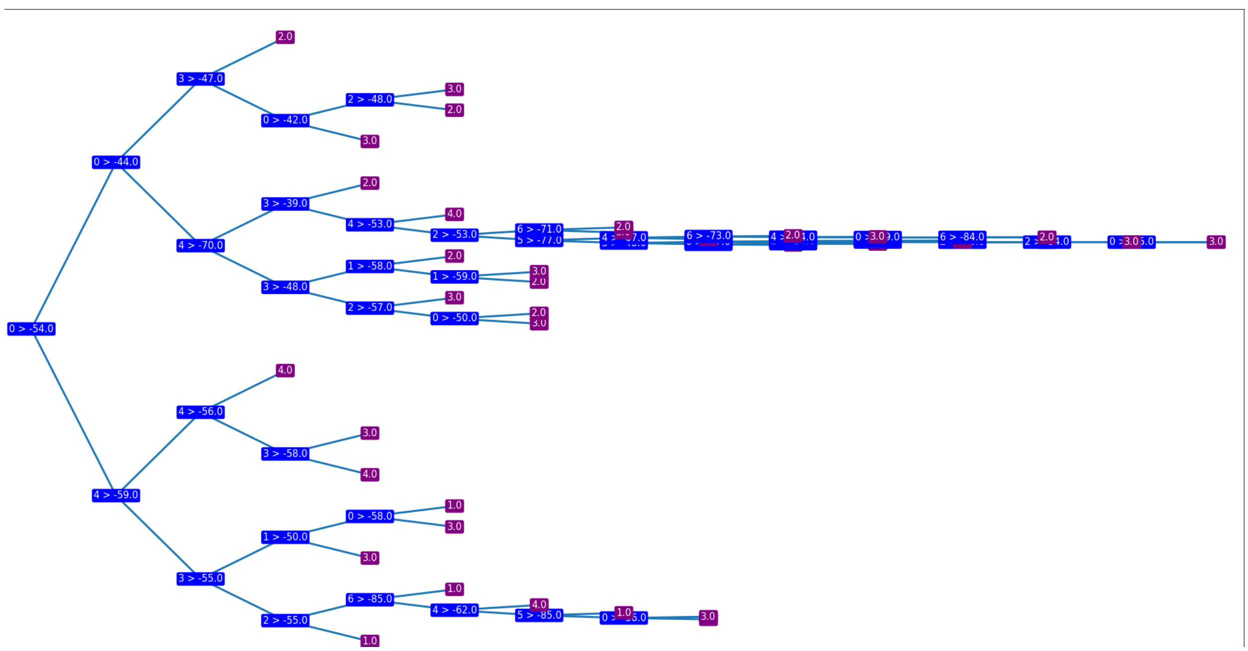
Intro to ML - Decision Tree CW

Output Visualisation



Evaluation

Clean Dataset Cross Validation Metrics (average):

- Confusion Matrix:

$$\begin{bmatrix} 49.7 & 0.0 & 0.1 & 0.2 \\ 0.0 & 48.0 & 2.0 & 0.0 \\ 0.3 & 2.1 & 47.6 & 0.0 \\ 0.4 & 0.0 & 0.3 & 49.3 \end{bmatrix}$$

- Accuracy: 0.973
- Precision:

$$\begin{bmatrix} 0.99 & 0.96 & 0.95 & 1.0 \end{bmatrix}$$

- Recall:

$$\begin{bmatrix} 0.99 & 0.96 & 0.95 & 0.99 \end{bmatrix}$$

- F1-Scores:

$$\begin{bmatrix} 0.99 & 0.96 & 0.95 & 0.99 \end{bmatrix}$$

Noisy Dataset Cross Validation Metrics (average):

- Confusion Matrix:

$$\begin{bmatrix} 38.5 & 2.8 & 3.3 & 4.4 \\ 3.4 & 39.3 & 4.2 & 2.8 \\ 2.8 & 3.4 & 40.8 & 4.5 \\ 5.6 & 2.2 & 3.1 & 38.9 \end{bmatrix}$$

- Accuracy: 0.7875
- Precision:
  [0.77 0.83 0.79 0.77]
- Recall:
  [0.79 0.79 0.79 0.78]
- F1-Scores:
  [0.77 0.81 0.79 0.77]

Results Analysis: From the confusion matrices, we see that the tree mixes up room 1 and room 4 in rare cases and it occasionally mixes up room 2 and room 3.

Dataset Differences: All the evaluation metrics are lower for the noisy dataset by about 0.2. The noisy dataset is less uniform, more random, and the generated trees are deeper. Hence, the decision tree algorithm has a harder job of recognizing patterns. The noisy dataset has high variance.

Pruning + Evaluation

Clean Dataset Cross Validation Metrics (average):

- Confusion Matrix:

$$\begin{bmatrix} 49.8 & 0.0 & 0.2 & 0.0 \\ 0.0 & 47.9 & 2.1 & 0.0 \\ 0.6 & 2.3 & 46.8 & 0.3 \\ 0.5 & 0.0 & 0.2 & 49.3 \end{bmatrix}$$

- Accuracy: 0.969
- Precision:

$$\begin{bmatrix} 0.98 & 0.95 & 0.95 & 0.99 \end{bmatrix}$$

- Recall:
  $\begin{bmatrix} 0.99 & 0.96 & 0.94 & 0.99 \end{bmatrix}$
- F1-Scores:
  $\begin{bmatrix} 0.99 & 0.96 & 0.94 & 0.99 \end{bmatrix}$

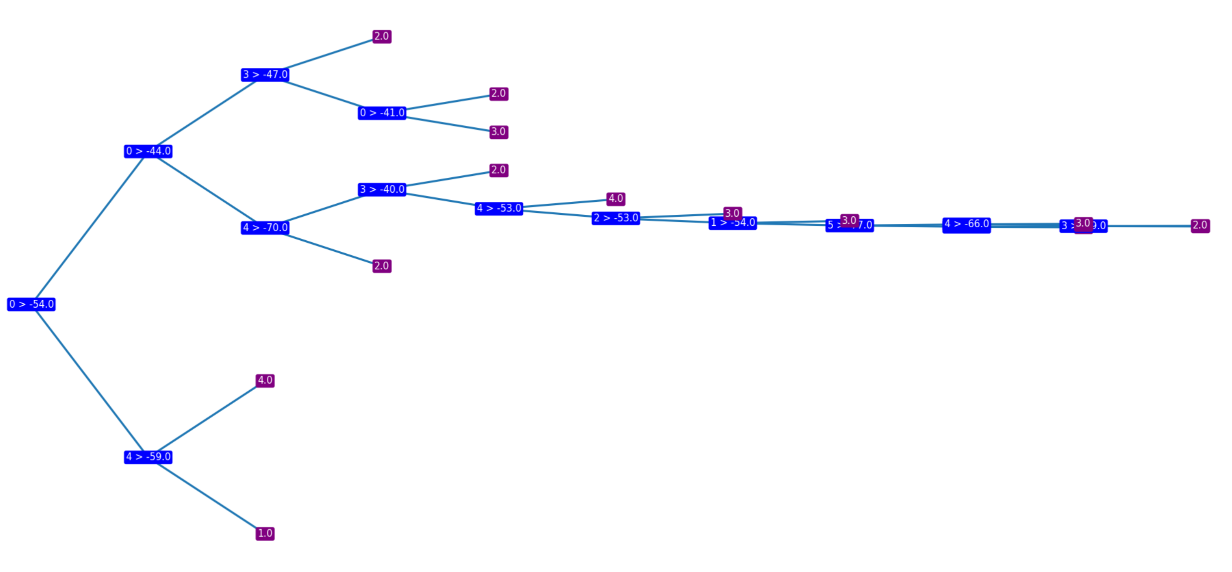
Noisy Dataset Cross Evaluation Metrics (average):

- Confusion Matrix:
  $$\begin{bmatrix} 43.4 & 1.6 & 1.6 & 2.4 \\ 2.4 & 43.7 & 2.5 & 1.1 \\ 2.4 & 3.1 & 44.2 & 1.8 \\ 1.9 & 1.6 & 2.3 & 44.0 \end{bmatrix}$$
- Accuracy: 0.8765
- Precision:
  [0.87 0.88 0.88 0.89]
- Recall:
  [0.89 0.88 0.86 0.88]
- F1-Scores:
  [0.87 0.88 0.86 0.89]

Results Analysis: As expected, the pruned dataset performs slightly worse in correctly identifying labels for the clean dataset; resulting in lower accuracy, recall, precision, and F1-scores across the board. The noisy dataset benefits from pruning. Hence, the unpruned noisy tree suffers from overfitting, which pruning solves.

Depth Analysis:

In general, the pruned trees were much more shallow than the unpruned ones - which is entirely within expectations. The depth of the decision trees was generally brought on by singular branches that went very deep in spite of not contributing much to the "width" of the tree. For clean datasets, as depth decreases, accuracy stays the same. For noisy ones, once depth decreases, accuracy increases.


2.0
3 > -47.0
0 > -41.0
2.0
3.0
0 > -44.0
2.0
3 > -40.0
4 > -53.0
4.0
2 > -53.0
3.0
4 > -70.0
3.0
4 > -66.0
3.0
2.0
2.0
0 > -54.0
4.0
4 > -59.0
1.0