

# Introduction to ML Decision Tree Coursework

Candidates:

Li, Yuyong  
Mohammed, Areeb Hamad  
Pandya, Rahil  
Jaddu, Koti

November 3, 2022

# Contents

<b>Step 2 - Creating Decision Trees</b>	<b>2</b>
(Bonus) Visualisation . . . . .	2
<b>Step 3 - Evaluation</b>	<b>3</b>
Cross Validation Classification Metrics . . . . .	3
Result Analysis . . . . .	5
Dataset Differences . . . . .	5
<b>Step 4 - Pruning (and Evaluation again)</b>	<b>6</b>
Cross Validation Classification Metrics After Pruning . . . . .	6
Result analysis after pruning . . . . .	8
Depth Analysis . . . . .	8

## Step 2 - Creating Decision Trees

### (Bonus) Visualisation

Please view the final page to see the visualisation of the decision tree trained on the entire clean dataset.

## Step 3 - Evaluation

### Cross Validation Classification Metrics

#### Clean Dataset

Average Confusion Matrix (2 d.p.)

	Room 1 Predicted	Room 2 Predicted	Room 3 Predicted	Room 4 Predicted
Room 1 Actual	49.50	0.00	0.40	0.10
Room 2 Actual	0.00	47.80	2.20	0.00
Room 3 Actual	0.00	1.80	48.00	0.20
Room 4 Actual	0.50	0.00	0.10	49.40

Derived Average Accuracy (3 s.f.)

0.974

Derived Average Precision per class (3 s.f.)

Room 1 Precision	0.990
Room 2 Precision	0.964
Room 3 Precision	0.947
Room 4 Precision	0.994

Derived Average Recall per class (3 s.f.)

Room 1 Recall	0.990
Room 2 Recall	0.956
Room 3 Recall	0.960
Room 4 Recall	0.988

Derived Average F1 Measure per class (3 s.f.)

Room 1 F1 Measure	0.990
Room 2 F1 Measure	0.960
Room 3 F1 Measure	0.953
Room 4 F1 Measure	0.991

Average Maximum Tree Depth (2 d.p.)

12.50

Figure 1: Cross Validation Classification Metrics for the Clean Dataset with Seed 123

## Noisy Dataset

Average Confusion Matrix (2 d.p.)

	Room 1 Predicted	Room 2 Predicted	Room 3 Predicted	Room 4 Predicted
Room 1 Actual	38.50	3.00	3.90	3.60
Room 2 Actual	2.90	38.70	5.20	2.90
Room 3 Actual	2.90	4.30	40.80	3.50
Room 4 Actual	3.50	3.10	3.50	39.70

Derived Average Accuracy (3 s.f.)

0.789

Derived Average Precision per class (3 s.f.)

Room 1 Precision	0.805
Room 2 Precision	0.788
Room 3 Precision	0.764
Room 4 Precision	0.799

Derived Average Recall per class (3 s.f.)

Room 1 Recall	0.786
Room 2 Recall	0.779
Room 3 Recall	0.792
Room 4 Recall	0.797

Derived Average F1 Measure per class (3 s.f.)

Room 1 F1 Measure	0.795
Room 2 F1 Measure	0.783
Room 3 F1 Measure	0.778
Room 4 F1 Measure	0.798

Average Maximum Tree Depth (2 d.p.)

18.80

Figure 2: Cross Validation Classification Metrics for the Noisy Dataset with Seed 123

## Result Analysis

We will use F1 scores to infer accuracy. For the clean dataset, rooms 1 and 4 are recognised with the highest accuracy. Room 3 is recognised with the lowest accuracy. For the noisy dataset, room 4 is recognised with the highest accuracy. Room 3 is recognised with the lowest accuracy. For both datasets, rooms 2 and 3 are the most confused as room 2 is mistaken as room 3 and vice versa. This is seen by relatively higher values in the confusion matrices for the corresponding entries.

## Dataset Differences

The derived average accuracy is much higher on the clean dataset (0.974) than the noisy dataset (0.789) as the model does not generalise as well for the noisy dataset. This is because decision trees are prone to overfit to the training data and any noise present. We can see this using the average maximum depth which is higher in the noisy dataset (18.8 vs 12.5), typically leading to more nodes and a complex tree which might incorrectly classify some unseen data.

## Step 4 - Pruning (and Evaluation again)

### Cross Validation Classification Metrics After Pruning

#### Clean Dataset

Average Confusion Matrix (2 d.p.)

	Room 1 Predicted	Room 2 Predicted	Room 3 Predicted	Room 4 Predicted
Room 1 Actual	49.66	0.00	0.344	0.00
Room 2 Actual	0.00	47.71	2.29	0.00
Room 3 Actual	0.58	2.44	46.56	0.42
Room 4 Actual	0.46	0.00	0.24	49.30

Derived Average Accuracy (3 s.f.)

0.966

Derived Average Precision per class (3 s.f.)

Room 1 Precision	0.980
Room 2 Precision	0.951
Room 3 Precision	0.942
Room 4 Precision	0.992

Derived Average Recall per class (3 s.f.)

Room 1 Recall	0.993
Room 2 Recall	0.954
Room 3 Recall	0.931
Room 4 Recall	0.986

Derived Average F1 Measure per class (3 s.f.)

Room 1 F1 Measure	0.986
Room 2 F1 Measure	0.953
Room 3 F1 Measure	0.936
Room 4 F1 Measure	0.989

Average Maximum Tree Depth (2 d.p.)

6.89

Figure 3: Nested Cross Validation Classification Metrics for the Clean Dataset with Seed 123

## Noisy Dataset

Average Confusion Matrix (2 d.p.)

	Room 1 Predicted	Room 2 Predicted	Room 3 Predicted	Room 4 Predicted
Room 1 Actual	44.38	1.08	1.43	2.11
Room 2 Actual	1.92	43.86	2.78	1.14
Room 3 Actual	2.17	3.62	43.90	1.81
Room 4 Actual	2.34	1.51	1.62	44.32

Derived Average Accuracy (3 s.f.)

0.882

Derived Average Precision per class (3 s.f.)

Room 1 Precision	0.873
Room 2 Precision	0.876
Room 3 Precision	0.883
Room 4 Precision	0.897

Derived Average Recall per class (3 s.f.)

Room 1 Recall	0.906
Room 2 Recall	0.882
Room 3 Recall	0.852
Room 4 Recall	0.890

Derived Average F1 Measure per class (3 s.f.)

Room 1 F1 Measure	0.889
Room 2 F1 Measure	0.879
Room 3 F1 Measure	0.867
Room 4 F1 Measure	0.894

Average Maximum Tree Depth (2 d.p.)

10.71

Figure 4: Nested Cross Validation Classification Metrics for the Noisy Dataset with Seed 123



## Result Analysis After Pruning

For the clean dataset, there is a 1% drop in accuracy. This is because the pruning will remove parts in the tree which has the potential to improve the model but is not traversed by the small validation data. For the noisy dataset, there is a 10% rise in accuracy. We are using the unseen validation data to remove the extra decision splits that are caused by the noise in the noisy dataset. This allows it to generalise better for the unseen test data leading to performance improvement.

## Depth Analysis

The average depth of the trees produced with the clean dataset before and after pruning is 12.5 and 6.9 whereas the corresponding average depths for the noisy dataset is 18.8 and 10.7. The average depth for both datasets dropped by around 45% after pruning. Generally, there is a negative correlation between the maximal depth and prediction accuracy but there will be diminishing returns if the tree is too small as shown by the clean dataset.

