A picture containing scale, device

Description automatically generatedDiagram

Description automatically generated1 - Output of Tree Visualisation Function

Figure 1: Overview of decision tree model

Figure 2: Close up view of decision tree root

2 - Step 3: Evaluation

**Chart

Description automatically generated**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Room 1** | **Room 2** | **Room 3** | **Room 4** |
| **Accuracy** | 0.972 | | | |
| **F1-score** | 0.987 | 0.959 | 0.949 | 0.989 |
| **Precision** | 0.985 | 0.954 | 0.957 | 0.989 |
| **Recall** | 0.990 | 0.964 | 0.942 | 0.988 |

**Figure 1: Clean dataset 10-fold cross validation metrics** (‘default\_rng’ seed set to 1)

Chart

Description automatically generated

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Room 1** | **Room 2** | **Room 3** | **Room 4** |
| **Accuracy** | 0.804 | | | |
| **F1-score** | 0.778 | 0.826 | 0.804 | 0.797 |
| **Precision** | 0.786 | 0.818 | 0.805 | 0.808 |
| **Recall** | 0.773 | 0.837 | 0.808 | 0.791 |

**Figure 2: Noisy dataset 10-fold cross validation metrics** (‘default\_rng’ seed set to 1)

**Result analysis:**

* **Clean dataset**: Room 3 is the least accurately predicted room with the largest number of false positives, Room 2 being the most mislabeled as Room 3. Room 4 was the most accurately predicted room.
* **Noisy dataset**: Room 1 is the least accurately predicted room with the largest number of false positives, Room 4 being the most mislabeled as Room 1. Room 2 was the most accurately predicted room.

**Dataset differences:** Yes, the noisy dataset resulted in the trained model making significantly more false positives and false negatives. These resulted in an accuracy decreased of 16.8% compared to clean dataset.

3 - Step 4: Pruning