Step3

1. Cross validation classification metrics
    - Clean dataset
        - cross_validation_final_accuracy: 0.9715
        - cross_validation_final_recalls:

          $[0.99266829 \quad 0.96280782 \quad 0.94940149 \quad 0.98337466]$
        - cross_validation_final_precisions:

          $[0.98789875 \quad 0.9595577 \quad 0.95050381 \quad 0.98823684]$
        - cross_validation_final_f1score:

          $[0.99023056 \quad 0.96079595 \quad 0.94952983 \quad 0.98564242]$
        - cross_validation_final_confusion_matrix:

          $$\begin{bmatrix} 49.6 & 0. & 0.1 & 0.3 \\ 0. & 47.9 & 2.1 & 0. \\ 0.1 & 2.2 & 47.5 & 0.2 \\ 0.3 & 0. & 0.3 & 49.4 \end{bmatrix}$$
    - Noisy dataset
        - cross_validation_final_accuracy: 0.8089999999999999
        - cross_validation_final_recalls:

          $[0.79170268 \quad 0.83078765 \quad 0.78520141 \quad 0.82999085]$
        - cross_validation_final_precisions:

          $[0.80395698 \quad 0.82873255 \quad 0.80690646 \quad 0.80255848]$
        - cross_validation_final_f1score:

          $[0.79594286 \quad 0.82744172 \quad 0.7936193 \quad 0.81426289]$
        - cross_validation_final_confusion_matrix:

          $$\begin{bmatrix} 38.9 & 3.2 & 2.7 & 4.2 \\ 2.1 & 41.3 & 4. & 2.3 \\ 4.2 & 3.1 & 40.4 & 3.8 \\ 3.1 & 2.4 & 3.1 & 41.2 \end{bmatrix}$$
2. Result Analysis
    - Clean dataset
        - Room1 & 4: both recall and precision are around 0.99, Room 1 has highest accuracy
        - Room2 & 3: both recall and precision are around 0.95, Room 3 has lowest accuracy
        - So Room 1 & 4 are well recognised and Room 2 & 3 are confused.
    - Noisy dataset
        - Room 2 & 4: both recall and precision are over 0.80, Room 2 has highest accuracy
        - Room 1 & 3: both recalls are below 0.80, Room 3 has lowest accuracy
        - So Room 2 & 4 are well recognised and Room 1 & 3 are confused.

3. Dataset Difference
    - Yes. Overall, performance when using clean dataset is better than when using noisy dataset because data in the noisy dataset has a higher variance and lower data correlation.

Step4

1. Cross validation classification metrics(after pruning)
    - Clean dataset
        - nested_cross_validation_final_accuracy: 0.976
        - nested_cross_validation_final_recalls:
          $[0.99011334 \quad 0.96810531 \quad 0.95812194 \quad 0.98741392]$
        - nested_cross_validation_final_precisions:
          $[0.987545 \quad 0.96879202 \quad 0.95844376 \quad 0.9897208]$
        - nested_cross_validation_final_f1score:
          $[0.98868944 \quad 0.96804624 \quad 0.95782215 \quad 0.98844923]$
        - nested_cross_validation_final_confusion_matrix:
          $$\begin{bmatrix} 49.51111111 & 0. & 0.25555556 & 0.23333333 \\ 0. & 48.4 & 1.6 & 0. \\ 0.24444444 & 1.55555556 & 47.93333333 & 0.26666667 \\ 0.4 & 0. & 0.24444444 & 49.35555556 \end{bmatrix}$$
    - noisy dataset
        - nested_cross_validation_final_accuracy: 0.8346666666666668
        - nested_cross_validation_final_recalls:
          $[0.83130037 \quad 0.84462976 \quad 0.83797499 \quad 0.82462574]$
        - nested_cross_validation_final_precisions:
          $[0.82689839 \quad 0.83976411 \quad 0.83695155 \quad 0.83682085]$
        - nested_cross_validation_final_f1score:
          $[0.82738953 \quad 0.8405816 \quad 0.83607172 \quad 0.82896481]$
        - nested_cross_validation_final_confusion_matrix:
          $$\begin{bmatrix} 40.7 & 2.41111111 & 2.55555556 & 3.33333333 \\ 2.6 & 41.96666667 & 2.93333333 & 2.2 \\ 2.53333333 & 3.28888889 & 43.15555556 & 2.52222222 \\ 3.44444444 & 2.33333333 & 2.91111111 & 41.11111111 \end{bmatrix}$$
2. Result Analysis
    - clean dataset
      the performance is improved slightly by 0.005
    - noisy dataset
      the performance is improved a little by 0.03
    - explain these performance differences

For both datasets, pruning improves the performance by reducing the overfitting during decision tree training. The data correlation is higher in clean dataset, which means pruning can do less on improving the performance for it.

3. Dataset Difference
    - before pruning
      the average depth of the clean dataset: 12.1
      the average depth of the noisy dataset: 19.4
    - after pruning
      the average depth of the clean dataset: 12.011111111111111
      the average depth of the noisy dataset: 18.144444444444446
    - the relationship
      The prediction accuracy is improved as the maximal depth decreases, this might be caused by the reduction of overfitting.