ML Assignment 2 Report
Omkar Bhandare (22CS30016)

	Noiseless Dataset	**Noisy Dataset**
Without Pruning	62.75%	45.12%
Post-Pruning	69.00%	60.33%

After the defaulted parameters and post-pruning were implemented, pre-pruning was tried with the min_sample_split hyperparameter; the max_depth hyperparameter was also tweaked to get better results in accuracy.
The default value of min_sample_split was set to 100, and the default value of the max_depth was 100.

For the noiseless dataset, the accuracies are mentioned in the below table: row tells the max_depth, and column tells the min_sample_split. The reported accuracy for the default parameters is **68.20%.**

Pre-Pruning	**5**	**10**	**15**	**20**	**25**	**50**
10	70.2%	69.35%	67.45%	65.6%	64.75%	64.25%
20	70.2%	69.45%	67.65%	66.3%	65.65%	65.3%
50	70.2%	70.05%	68.85%	67.85%	67.55%	67.55%
200	70.4%	70.95%	70.05%	69.95%	69.95%	69.95%
300	70.4%	71.15%	71.15%	71.15%	71.15%	71.15%
400	70.4%	70.85%	70.85%	70.85%	70.85%	70.85%
500	70.0%	**70.45%**	70.45%	70.45%	70.45%	70.45%

For the noisy dataset, the accuracies are mentioned in the table below: row tells the max_depth, and column tells the min_sample_split. The reported accuracy for the default parameters is **59.38%.**

Pre-Pruning	**5**	**10**	**15**	**20**	**25**	**50**
10	62.62%	60.66%	56.29%	52.83%	49.33%	46.87%
20	62.62%	61%	57.20%	54.17%	51.62%	49.79%
50	62.66%	61.91%	60.08%	58.29%	57%	56.75%
200	62.41%	63.04%	62.04%	61.79%	61.62%	61.67%

300	62.41%	63.04%	62.45%	62.38%	62.2%	62.25%
400	62.5%	62.83%	62.42%	62.29%	62.25%	62.25%
500	62.92%	**63.25%**	62.88%	62.87%	62.87%	62.88%

As we can see from the highlighted value, the trees for these hyperparameters minimizes the accuracy gap between the noisy and noiseless data. So, for the adopted optimisation technique (10, 500) are the optimal hyperparameters.