Machine Learning Assignment-2 Report

Classification using Decision Tree

Dataset Used: cardio.csv and cardio_noise.csv

Results:

Without Pre-Pruning

Noiseless Dataset	Before Reduced Error Pruning	After Reduced Error Pruning
Accuracy	0.6355	0.6955
Macro Precision	0.635351872667416	0.7013120692522685
Macro Recall	0.6356999555844751	0.6990266683644157

Noisy Dataset	Before Reduced Error Pruning	After Reduced Error Pruning
Accuracy	0.44625	0.6066666666666667
Macro Precision	0.4459869281045752	0.6067203344600605
Macro Recall	0.44619776721602006	0.6066447406798352

With Pre-Pruning (GINI index & Minimum Size)

Noiseless Dataset	Before Reduced Error Pruning	After Reduced Error Pruning
Accuracy	0.7075	0.7075
Macro Precision	0.7091154883733981	0.7092982456140351
Macro Recall	0.7092171738692339	0.7093194398630839

Noisy Dataset	Before Reduced Error Pruning	After Reduced Error Pruning
Accuracy	0.615	0.6279166666666667
Macro Precision	0.6150458460683523	0.6296081525562294
Macro Recall	0.615014208372801	0.6278191050530695

Pruning significantly improves the accuracy, precision, and recall of the Decision Tree model on both noiseless and noisy datasets, with a more pronounced effect on the noisy dataset.