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)

Noiseless Dataset	Before Reduced Error Pruning	After Reduced Error Pruning
Accuracy	0.7025	0.715
Macro Precision	0.7042857142857143	0.7201041482342829
Macro Recall	0.704306400948868	0.7182451556498453

Noisy Dataset	Before Reduced Error Pruning	After Reduced Error Pruning
Accuracy	0.5120833333333333	0.61125
Macro Precision	0.5121082131162233	0.6112875893421045
Macro Recall	0.5121021169503248	0.6112315589765527

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.