Assignment 2

Matthias
Bennett

February 15th, 2024

1 Experiment Results

index	Normalization	Max Depth	Splitter	Accuracy %
0	No Preprocessing	3	best	0.9416666666666668
1	No Preprocessing	3	random	0.9333333333333333
2	No Preprocessing	5	best	0.9333333333333332
3	No Preprocessing	5	random	0.95
4	No Preprocessing	7	best	0.9333333333333332
5	No Preprocessing	7	random	0.9416666666666668
6	No Preprocessing	9	best	0.9
7	No Preprocessing	9	random	0.9583333333333334
8	Z-Score	3	best	0.9333333333333332
9	Z-Score	3	random	0.925
10	Z-Score	5	best	0.9
11	Z-Score	5	random	0.9166666666666666
12	Z-Score	7	best	0.9333333333333332
13	Z-Score	7	random	0.9583333333333334
14	Z-Score	9	best	0.9166666666666667
15	Z-Score	9	random	0.9583333333333334
16	MinMax	3	best	0.9416666666666668
17	MinMax	3	random	0.95
18	MinMax	5	best	0.9333333333333332
19	MinMax	5	random	0.9333333333333332
20	MinMax	7	best	0.9333333333333332
21	MinMax	7	random	0.925
22	MinMax	9	best	0.9083333333333332
23	MinMax	9	random	0.9333333333333333

2 Discussion

Initially when the data was processed, before cross validation the average wasn't present, so running through all processes produced outliers and inconsistent results are various depths. In some cases MinMax random at depth 9 proved the best one run, then Zscore random at depth 9 proved to be the new best with variance from as low as 76% accuracy all the way to 99% accuracy.

The step that helped get better performance was cross validation, running through all the processes multiple times across folds and getting the average results to produce better accuracy percentages across each technique.

In our final test in the report here depth 7 seems to be the most accurate, as refining the decision tree, but with our streamlit this could be tested and compared to other depths however much is needed and those results could eventually be compiled to find the best depth.