Assignment 2

Alex Radecki

Mar 1^{st} 2023

1 Experiment Results

Normalization	Decision Tree(DT) parameter	Splitter	Accuracy %
No preprocessing	DT with max depth 5	Best	91%
Minmax	DT with max depth 5	Best	90%
Z-score	DT with max depth 5	Best	90%
No preprocessing	DT with max depth 5	Random	93%
Minmax	DT with max depth 5	Random	96%
Z-score	DT with max depth 5	Random	93%
No preprocessing	DT with max depth 6	Best	86%
Minmax	DT with max depth 6	Best	88%
Z-score	DT with max depth 6	Best	88%
No preprocessing	DT with max depth 6	Random	90%
Minmax	DT with max depth 6	Random	93%
Z-score	DT with max depth 6	Random	93%
No preprocessing	DT with max depth 7	Best	90%
Minmax	DT with max depth 7	Best	88%
Z-score	DT with max depth 7	Best	87%
No preprocessing	DT with max depth 7	Random	90%
Minmax	DT with max depth 7	Random	98%
Z-score	DT with max depth 7	Random	93%

2 Discussion

Looking at the table, while there isn't a noticeable correlation between depth and accuracy, we can see that on average, trees that used a random split were more accurate than trees that used the best split. This is likely due to the fact that the dataset contained a large number of features and noise, causing the models using the best split to overfit. Minmax appeared to consistently perform better than both z-score and no preprocessing. Increasing the max depth does not necessarily mean a more accurate model due to overfitting, which makes sense as a max depth of 5 was the most accurate in our experiment.