

Prediction #2

—

By Michael Ryvin

Recursive Partitioning

- Although ideally a simpler model would result in better accuracy on different samples of the training data, I instead found that controlling the tree to include more separations resulted in a more accurate prediction
 - cp to 0.001
 - minsplit to 10
 - minbucket was not used
- Despite the simplicity of these numbers, they were obtained through trial and error by changing the values of cp, minsplit, and minbucket, until obtaining the best prediction model I could find
 - I found that lowering cp helped improve the model, but eventually reached a point where lowering any more would worsen the model
 - The same idea was used for both minsplit and minbucket, though I found that using minbucket did not help the model and so I removed it from the control

Tree

- Unfortunately given my method, the tree is essentially illegible due to its immense size
- However, the score received on Kaggle is still good, currently being at 0.82765
- Presumably there is a much simpler tree that is a better prediction model to the dataset, but I could not find it

