PREDICTION CHALLENGE 2

KEVIN LARKIN

THE SETUP

- Setting up rpart and analyzing the decision tree helped me gain a way better understanding of what was going on with the data and how I could have predicted it better having this tool last week.

- Once I had rpart and my prediction set up, I tried different combinations of minsplit, minbucket, and cp to determine which had better accuracy than the default prediction of about 83-84% according to the cross validation.

- I found that both minsplit and minbucket actually decreased my initial accuracy almost every time they were applied, so I found it best to leave them alone for this prediction.

- Cp had the biggest impact on accuracy, and I found that a cp of 0.001 yielded the best accuracy at around 85-86% without overfitting, so this is all I ended up applying.

- With a cp that small, it made the decision tree virtually impossible to read, so I used the default options to analyze the tree first.