

Most models were between 69% and 76% accurate, and the most accurate models were those that had a minimum of 10 samples, and max depths varying between 10 and 12. Certain models work better because the limits set by the parameters happen to correspond somewhat well to the categorizing of the data. One parameter criterion, which was set at entropy, or randomness, indicating the way the order of categories were decided. Other parameters included the minimum number of samples allowed in a subset and the maximum depth, or the most number of layers allowed in the tree. The higher the minimum number of samples and the smaller the maximum depth, the smaller and shorter the decision tree was and the lower the accuracy was. This prevented overfitting because having a limit set ensured that the tree could only get so complex.