In this module, we discuss two applications of data analytics – Supreme Court (prediction of decision before getting to the court) and healthcare cost prediction.

To predict the outcomes of cases, Classification and Regression Trees (CART) method is used. The outcome of this method is binary, in this task – whether the Supreme Court accepts the case or rejects the case. Logistic regression models are generally not interpretable when compared to CART.

In CART, we build a tree by splitting on variables. To control the number of splits, one method is to create a lower bound for the number of points in each subset. In R, a parameter that does the above function is “minbucket”.

To improve accuracy of CART, we use a method called “Random Forests”. But this model reduces the interpretability, when compared to CART.

Optimal Classification Tree is the most optimised method to decision tree. While taking the input minbucket value, k cross validation helps to choose the best fit for the decision tree.