

The C5.0 classification model was used in this 4-class problem data with Ntrain=165, P=11, using caret R-package by running the code below. The winnowing option was tuned over in the model, which is a kind of feature selection approach. This excerpt I quote regarding winnowing from the companion book of caret, a must-have book in my opinion to realize hidden gems coded in the package: Kuhn M, Johnson K. Applied predictive modeling. 1st edition. New York: Springer. 2013.

C5.0 also has an option to winnow or remove predictors: an initial algorithm uncovers which predictors have a relationship with the outcome, and the final model is created from only the important predictors. To do this, the training set is randomly split in half and a tree is created for the purpose of evaluating the utility of the predictors (call this the winnowing tree). Two procedures characterize the importance of each predictor to the model: 1. Predictors are considered unimportant if they are not in any split in the winnowing tree. 2. The half of the training set samples not included to create the winnowing tree are used to estimate the error rate of the tree. The error rate is also estimated without each predictor and compared to the error rate when all the predictors are used. If the error rate improves without the predictor, it is deemed to be irrelevant and is provisionally removed.

```
c50Grid <- expand.grid(.trials = c(1:9, (1:10)*10),  
                      .model = c("tree", "rules"),  
                      .winnow = c(TRUE, FALSE))  
  
c50Grid  
  
set.seed(1) # important to have reproducible results  
c5Fitvac <- train(Class ~ .,  
                  data = training,  
                  method = "C5.0",  
                  tuneGrid = c50Grid,  
                  trControl = ctrl,  
                  metric = "Accuracy", # not needed it is so by default
```

```

importance=TRUE, # not needed
preProc = c("center", "scale"))
> c5Fitvac$finalModel$tuneValue
      trials model winnow
16      70  tree  FALSE
CV tuning output:
enter image description here

Excerpt from the C5.0 tree output:

> c5Fitvac$finalModel$tree
[1] "id=\"See5/C5.0 2.07 GPL Edition 2014-01-22\"\\nentries=\"70\"\\nntype=\"2\" class
Now importance of predictors:

> predictors(c5Fitvac )
[1] "IL23R"    "IL12RB2" "IL8"      "IL23A"    "IL6ST"    "IL12A"    "IL12RB1"
[8] "IL27RA"   "IL12B"   "IL17A"    "EBI3"

```

Why is it in the plot, the accuracy levels of No-winnowing about two times that of the winnowing? Can you please help interpreting this output when it says winnow = FALSE? How to visualize the tree output, instead of the computed junk text that appeared in my case? is there any way to behold a tree instead of crowded symbols?