

# TM\_\_DecisionTree

*Jashon Newlun*

*2/11/2020*

Below is the code and tree output for the tidymodels decision tree on the Iris data

```
set.seed(100)
```

```
iris_split <- iris %>% initial_split(prop = .7)
iris_split %>% training() %>% glimpse()
```

```
## Observations: 105
## Variables: 5
## $ Sepal.Length <dbl> 5.1, 4.9, 4.7, 4.6, 5.0, 5.4, 4.6, 5.0, 4.9, 5.4,...
## $ Sepal.Width <dbl> 3.5, 3.0, 3.2, 3.1, 3.6, 3.9, 3.4, 3.4, 3.1, 3.7,...
## $ Petal.Length <dbl> 1.4, 1.4, 1.3, 1.5, 1.4, 1.7, 1.4, 1.5, 1.5, 1.5,...
## $ Petal.Width <dbl> 0.2, 0.2, 0.2, 0.2, 0.2, 0.4, 0.3, 0.2, 0.1, 0.2,...
## $ Species <fct> setosa, setosa, setosa, setosa, setosa, setosa, s...
```

```
iris_recipe <- training(iris_split) %>%
  recipe(Species ~.) %>%
  step_corr(all_predictors()) %>%
  step_center(all_predictors(), -all_outcomes()) %>%
  step_scale(all_predictors(), -all_outcomes()) %>%
  prep()
iris_recipe
```

```
## Data Recipe
##
## Inputs:
##
##      role #variables
## outcome      1
## predictor      4
##
## Training data contained 105 data points and no missing data.
##
## Operations:
##
## Correlation filter removed Petal.Length [trained]
## Centering for Sepal.Length, Sepal.Width, Petal.Width [trained]
## Scaling for Sepal.Length, Sepal.Width, Petal.Width [trained]
```

```
iris_testing <- iris_recipe %>%
  bake(testing(iris_split))

glimpse(iris_testing)
```

```
## Observations: 45
```

```
## Variables: 4
## $ Sepal.Length <dbl> -1.65316231, -0.01005981, -0.47951767, -0.8316110...
## $ Sepal.Width <dbl> -0.458715825, 2.100785517, 0.704693876, 1.4027396...
## $ Petal.Width <dbl> -1.24015416, -1.24015416, -1.24015416, -0.9863214...
## $ Species <fct> setosa, setosa, setosa, setosa, setosa, setosa, s...
```

```
iris_training <- juice(iris_recipe)

glimpse(iris_training)
```

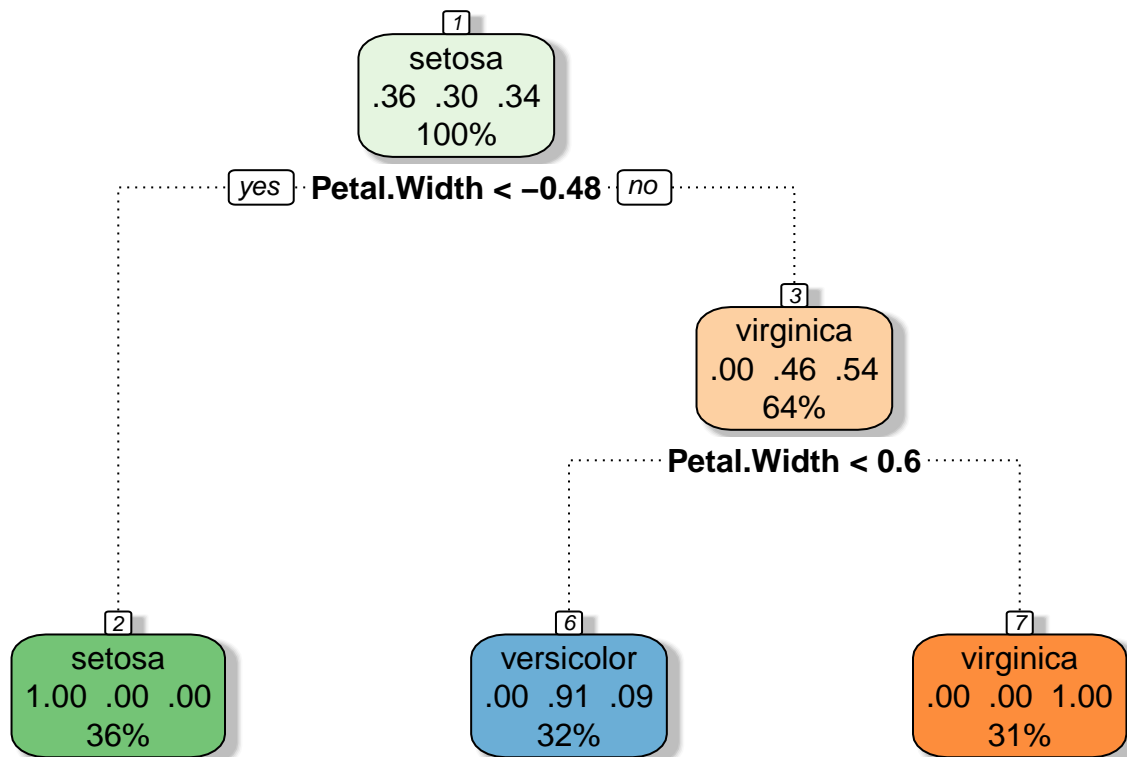
```
## Observations: 105
## Variables: 4
## $ Sepal.Length <dbl> -0.8316111, -1.0663400, -1.3010689, -1.4184334, -...
## $ Sepal.Width <dbl> 0.937375816, -0.226033885, 0.239329996, 0.0066480...
## $ Petal.Width <dbl> -1.2401542, -1.2401542, -1.2401542, -1.2401542, -...
## $ Species <fct> setosa, setosa, setosa, setosa, setosa, setosa, s...
```

```
iris_tree <- decision_tree(tree_depth = 4, mode = "classification") %>%
  set_engine("rpart") %>%
  fit(Species ~ ., data = iris_training)

iris_tree %>%
  predict(iris_testing) %>%
  bind_cols(iris_testing) %>%
  metrics(truth = Species, estimate = .pred_class)
```

```
## # A tibble: 2 x 3
##   .metric .estimator .estimate
##   <chr>    <chr>      <dbl>
## 1 accuracy multiclass    0.933
## 2 kap      multiclass    0.898
```

```
rattle::fancyRpartPlot(iris_tree$fit)
```



Rattle 2020-Feb-11 10:29:27 jasho

```
iris_tree$fit
```

```
## n= 105
##
## node), split, n, loss, yval, (yprob)
##      * denotes terminal node
##
## 1) root 105 67 setosa (0.36190476 0.29523810 0.34285714)
##    2) Petal.Width< -0.478656 38 0 setosa (1.00000000 0.00000000 0.00000000) *
##    3) Petal.Width>=-0.478656 67 31 virginica (0.00000000 0.46268657 0.53731343)
##      6) Petal.Width< 0.6001331 34 3 versicolor (0.00000000 0.91176471 0.08823529) *
##      7) Petal.Width>=0.6001331 33 0 virginica (0.00000000 0.00000000 1.00000000) *
```