

# STA 663: Graduate Project

Brian N. White

4/7/2020

```
library(tidyverse)
library(tidymodels)
library(ranger)
```

## Research Question

Can one use the combat statistics of a pokemon to predict whether or not that pokemon is legendary?

```
pokemon=read.csv("data/Pokemon.csv")
```

```
summary(pokemon$Legendary)
```

```
## False  True
```

```
##    735    65
```

```
65/735
```

```
## [1] 0.08843537
```

```
set.seed(7)
```

```
poke_split=initial_split(pokemon, prop=0.5)
```

```
poke_train=training(poke_split)
```

```
poke_test=testing(poke_split)
```

```
poke_cv=vfold_cv(poke_train, v=5)
```

```
rf_spec=rand_forest(mode="classification",
                     mtry=floor(sqrt(6)),
                     trees=tune()) %>%
  set_engine("ranger")
```

```
grid=expand.grid(trees=c(10, 25, 50, 100, 200, 300))
```

```
poke_model=tune_grid(rf_spec,
                      Legendary~Total + HP + Attack + Defense + Sp..Atk + Sp..Def + Speed,
                      grid=grid,
                      resamples=poke_cv,
                      metrics=metric_set(gain_capture, accuracy))
```

```
best <- poke_model %>%
```

```

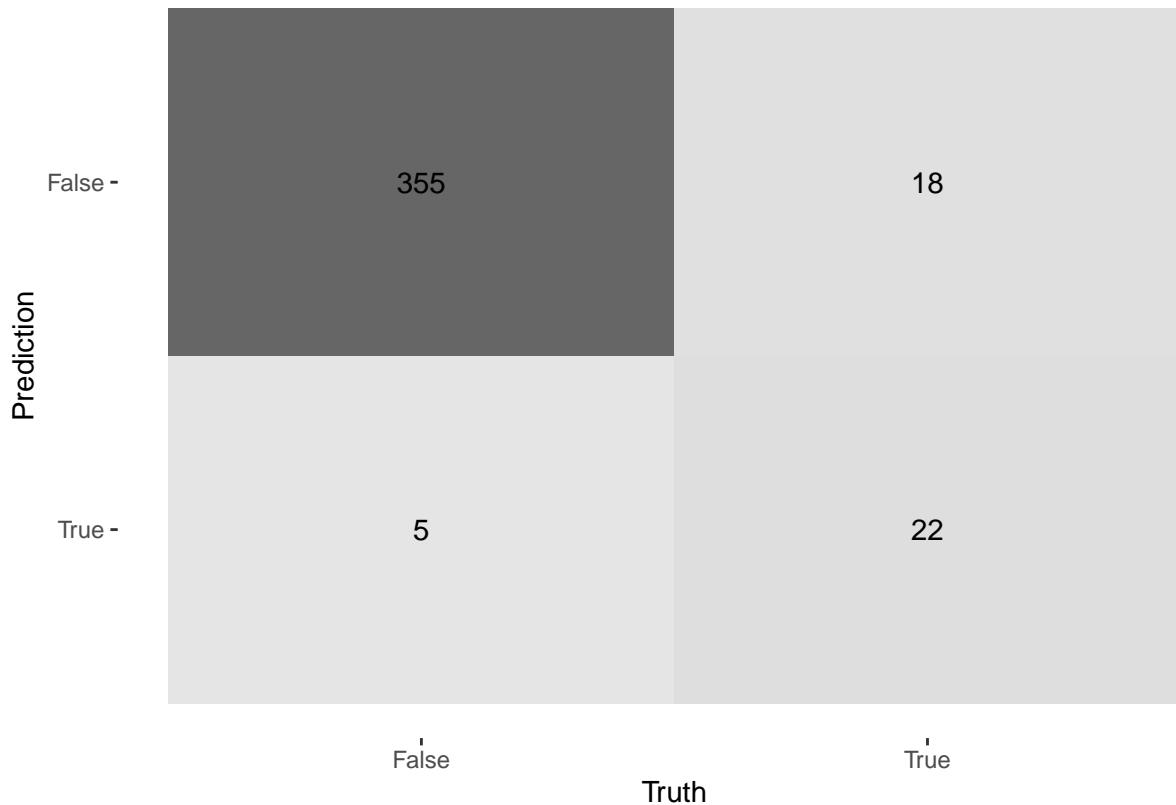
select_best(metric = "gain_capture") %>%
pull()

rf_final=rand_forest(mode="classification",
                      mtry=floor(sqrt(6)),
                      trees=best) %>%
set_engine("ranger")

poke_final=fit(rf_final,
               Legendary~Total + HP + Attack + Defense + Sp..Atk + Sp..Def + Speed,
               data=poke_train)

poke_final %>%
  predict(new_data = poke_test) %>%
  bind_cols(poke_test) %>%
  conf_mat(truth = Legendary, estimate = .pred_class) %>%
  autoplot(type = "heatmap")

```



```

Legendary_Accuracy=26/(14+26)
Not_Legendary_Accuracy=353/(353+7)

```

```
Legendary_Accuracy
```

```
## [1] 0.65
```

```
Not_Legendary_Accuracy
```

```
## [1] 0.9805556
```