Random Forest Model

Author: Rajat Jain

Last Updated: 2018-05-29

Contents

Setting Training Control Params

Using 10-fold Cross Validation with 10 repeatitions.

```
ctrl <- trainControl(method="repeatedcv",
                     number=10,
                     repeats=10,
                     classProbs=TRUE,
                     savePredictions=TRUE,
                     allowParallel=TRUE)
set.seed(123)
```

Training - Random Forest Model

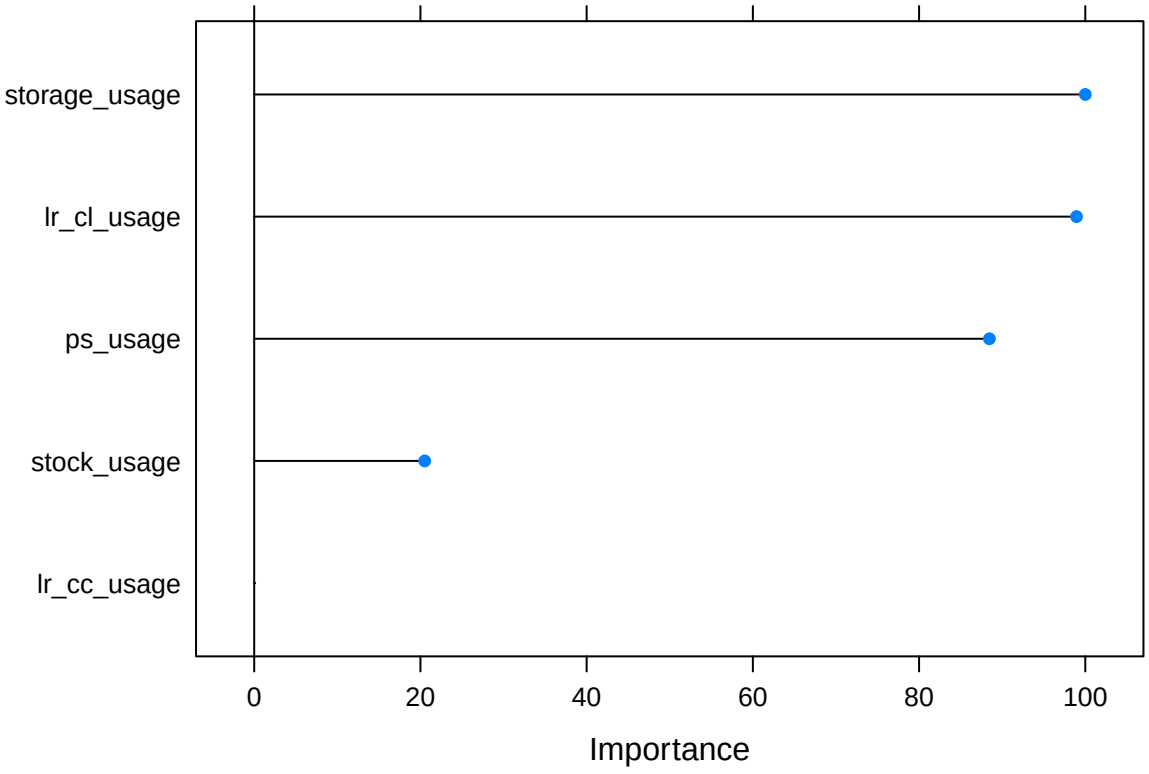
Caret is Awesome! So using caret to train the model.

```
fit <- train(class ~ lr_cc_usage + lr_cl_usage + storage_usage + ps_usage + stock_usage,
             data=usage.data, method = "rf", trControl = ctrl)
```

Summary of the trained Model

```
## Random Forest
##
## 2650 samples
##    5 predictor
##    2 classes: 'OTHER', 'PHOTOGRAPHER'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 10 times)
## Summary of sample sizes: 2385, 2385, 2385, 2385, 2385, 2385, ...
## Resampling results across tuning parameters:
##
##   mtry  Accuracy   Kappa
##   2     0.5914717  0.13909581
##   3     0.5767547  0.10725457
##   5     0.5672075  0.08732077
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 2.
```

Variable Importance



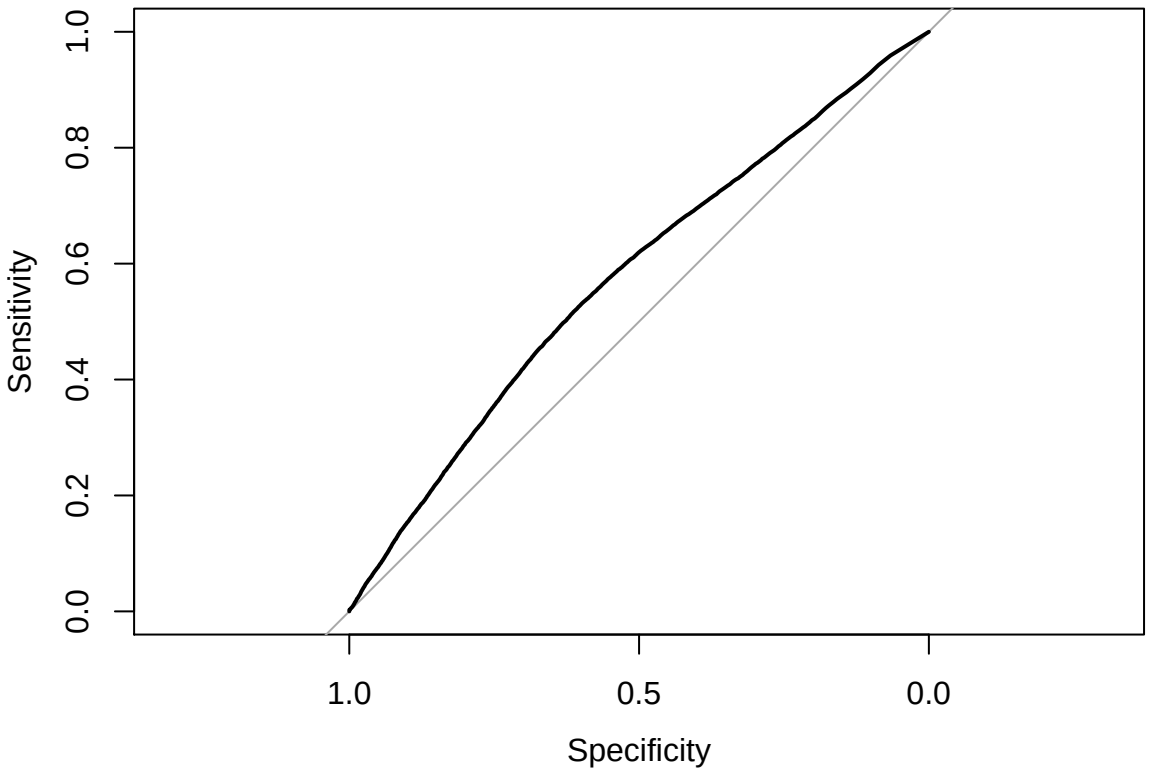
Performance

Based on the measure defined in the FPS, we will use classification accuracy as our performance measure.

Confusion Matrix

```
## Cross-Validated (10 fold, repeated 10 times) Confusion Matrix
##
## (entries are percentual average cell counts across resamples)
##
##                Reference
## Prediction      OTHER PHOTOGRAPHER
##   OTHER          42.5         26.8
##   PHOTOGRAPHER  14.1         16.6
##
##  Accuracy (average) : 0.5915
```

ROC Curve




```
##
## Call:
## plot.roc.default(x = fit$pred$obs, predictor = fit$pred$PHOTOGRAPHER)
##
## Data: fit$pred$PHOTOGRAPHER in 45000 controls (fit$pred$obs OTHER) < 34500 cases (fit$pred$obs PHOTO
## Area under the curve: 0.5763
```

Accuracy

- Kohen's Kappa: 0.14
- Observed Accuracy : 59.15%
- Desired accuracy : 70%
- Performance is Not Satisfactory.