**Customer Brand Preferences Report.**

**Hunter Wu. 13Nov. 2019**

**This task requires you to prepare one deliverable for Danielle Sherman:**

* ***Customer Brand Preferences Report.*** A report in a Zip file that includes:
  + A brief summary in Word or Excel of your methods and results that includes:
    - The classifiers you tried.
    - The classifier you selected to make the predictions, including a rationale for selecting the method you did and the level of confidence in the predictions.
    - The predicted answers to the brand preference question for the instances of survey results that are missing that answer.
    - A chart that displays the customer preference for each brand based on the combination of the actual answers and the predicted answers to the brand preference survey question.
  + The results of each classifier you ran exported from R

***Customer Brand Preferences Report***

* For each method, did you experiment with adjusting the input values until you did not see an improvement in the error metrics?
* Did you include a brief explanation of your rationale for selecting the training model you did?
* Did you provide a simple chart that displays which brand Blackwell customers prefer?
* Did you include the logs of the results of each classifier run?

**Business goal: to find which brand does customer prefer based on**

**Use current data to predict a model > fill out the missing data from the test set > final prediction**

**Find relationships between age/ salary to Brand (1/2)**

**Ggplot2 deepdive**

**GOAL 12nov> at least one model**

**Whats the business question?**

**How to pre process ? do you bin or**

**Metric:**

**Kappa**

**Accuracy**

**ROC**

**0>0**

**0>1**

In train.default(x, y, weights = w, ...) :

The metric "ROC" was not in the result set. Accuracy will be used instead.

**Randomforest with all**

|  |
| --- |
| Confusion Matrix and Statistics  Reference  Prediction 0 1  0 868 115  1 73 1418    Accuracy : 0.924  95% CI : (0.9129, 0.9341)  No Information Rate : 0.6196  P-Value [Acc > NIR] : < 2.2e-16    Kappa : 0.8402    Mcnemar's Test P-Value : 0.002788    Sensitivity : 0.9224  Specificity : 0.9250  Pos Pred Value : 0.8830  Neg Pred Value : 0.9510  Prevalence : 0.3804  Detection Rate : 0.3508  Detection Prevalence : 0.3973  Balanced Accuracy : 0.9237    'Positive' Class : 0    > varImp(randomFit)  rf variable importance  Overall  salary 100.000  age 46.836  credit 12.314  car 4.653  zipcode 1.729  elevel 0.000 |
|  |
| |  | | --- | |  | |

**# Subsetting data = exclude variables v1, v2**

**rfvars <- names(Brand) %in% c("elevel", "zipcode", "car")**

> confusionMatrix(data = randomPredict1, Rtesting1$brand)

Confusion Matrix and Statistics

Reference

Prediction 0 1

0 863 115

1 78 1418

Accuracy : 0.922

95% CI : (0.9107, 0.9323)

No Information Rate : 0.6196

P-Value [Acc > NIR] : < 2e-16

Kappa : 0.8357

Mcnemar's Test P-Value : 0.00956

Sensitivity : 0.9171

Specificity : 0.9250

Pos Pred Value : 0.8824

Neg Pred Value : 0.9479

Prevalence : 0.3804

Detection Rate : 0.3488

Detection Prevalence : 0.3953

Balanced Accuracy : 0.9210

'Positive' Class : 0

> varImp(randomFit1)

rf variable importance

Overall

salary 100.0

age 30.2

credit 0.0

**################## Only Salary, age, Credit, C5.0**

> #confusionMatrix

> confusionMatrix(data = modelPredict1, testing1$brand)

Confusion Matrix and Statistics

Reference

Prediction 0 1

0 829 90

1 112 1443

Accuracy : 0.9184

95% CI : (0.9069, 0.9288)

No Information Rate : 0.6196

P-Value [Acc > NIR] : <2e-16

Kappa : 0.826

Mcnemar's Test P-Value : 0.1395

Sensitivity : 0.8810

Specificity : 0.9413

Pos Pred Value : 0.9021

Neg Pred Value : 0.9280

Prevalence : 0.3804

Detection Rate : 0.3351

Detection Prevalence : 0.3715

Balanced Accuracy : 0.9111

'Positive' Class : 0

> varImp(modelFit)

C5.0 variable importance

Overall

salary 100.00

age 85.51

car 80.97

credit 37.43

zipcode 0.00

elevel 0.00

**C5.0 all**

#confusionMatrix

> confusionMatrix(data = modelPredict, testing$brand)

Confusion Matrix and Statistics

Reference

Prediction 0 1

0 858 102

1 78 1436

Accuracy : 0.9272

95% CI : (0.9163, 0.9372)

No Information Rate : 0.6217

P-Value [Acc > NIR] : < 2e-16

Kappa : 0.8461

Mcnemar's Test P-Value : 0.08647

Sensitivity : 0.9167

Specificity : 0.9337

Pos Pred Value : 0.8938

Neg Pred Value : 0.9485

Prevalence : 0.3783

Detection Rate : 0.3468

Detection Prevalence : 0.3880

Balanced Accuracy : 0.9252

'Positive' Class : 0

> varImp(modelFit)

C5.0 variable importance

Overall

age 100.00

salary 100.00

credit 91.84

car 64.95

elevel 0.00

zipcode 0.00

**5 bins salary**

onfusion Matrix and Statistics

Reference

Prediction Acer Sony

Acer 863 149

Sony 78 1384

Accuracy : 0.9082

95% CI : (0.8962, 0.9193)

No Information Rate : 0.6196

P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.8081