**Brand Preference Report**

List of models tried :

1. Stochastic GBM
2. Random Forest
3. C5.0

Classifiers chosen for Predictions on complete dataset :

1. Random Forest – Used crossvalidation with 10 sets and a tunelength of 5.

Result : Accuracy Kappa

0.9203074 0.8314316

Variable importance :

# Overall

# salary 100.0000

# age 51.6659

# credit 13.9493

# elevel3 0.8993

# elevel4 0.8240

# elevel1 0.8213

# elevel2 0.7658

# zipcode3 0.4909

# zipcode4 0.4897

# zipcode6 0.4876

# zipcode7 0.4655

# zipcode2 0.4634

# zipcode1 0.4451

# zipcode5 0.4411

# zipcode8 0.3464

# car15 0.3000

# car8 0.2092

# car12 0.1529

# car10 0.1418

# car5 0.1413

1. C5.0 Model - Used crossvalidation with 10 sets without any grid control variables since this yielded more accuracy.

Result : Accuracy Kappa

0.9215210 0.8326841

Variable Importance :

# salary 100.00

# age 85.21

# car2 16.62

# zipcode4 14.96

# elevel1 13.61

# zipcode2 10.91

# car20 9.52

Survey results on Incomplete dataset :

Using the above 2 models listed above predicted on Incomplete dataset. The accuracy score was very low as expected since the predicted values did not match with the ground truth since brand data had missing data (‘0’ values which could also mean ‘Acer’ Brand)

Attached documents :

1. 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.
2. The plot below shows the brand preference to be more towards ‘Sony’ . Also the plot shows the accuracy of brand preference which matches brand preference with complete dataset.

Diagram

Description automatically generated