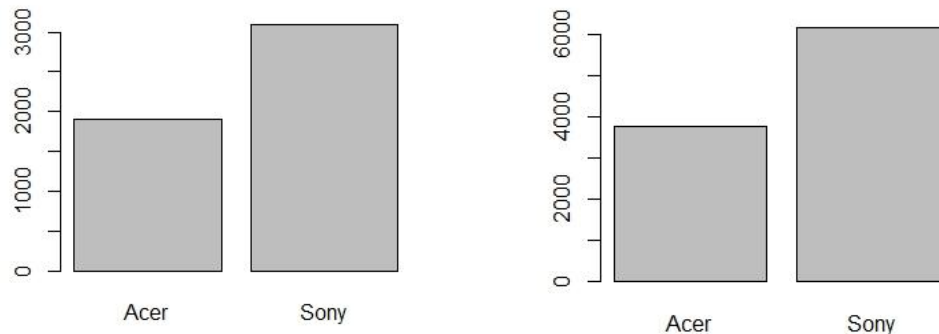# Customer Brand Preferences Report

Sony vs. Acer

Hunter Wu.
14Nov. 2019

This research aim at the brand preference survey does customers prefer between Sony and Acer, predicting from the demographics we have collected in the survey of 15000 respondents.

Predictions(left) shows the same distribution as survey data(right).

Concluded from "completeresponse" dataset we have trained a model (tested with 25% of sample with accuracy= 0.92, and Kappa=0.83) to predict the missing data on brand preference in "incompleteresponse" set. In the missing data set of 5000 examples, we have predicted 1909(38%) customers prefer Acer and 3091(62%) prefer Sony. Adding up prediction and response, out of 14898 response we have at hand, Acer got 5653**(38%)** and Sony got 9245 **(62%)** preference.

The final model implemented on prediction uses 2 variables, namely salary and age, while the salary is discretize into 5 income groups in our population data. Below can be seen as a simple log on model training with classifiers defining how the model was trained before predicting brand preference. The training set we take 75% of sample and verified its accuracy with 25% and get the following result.

Algorithms- Two classification algorithms employed to ensure we have the best model. Both C5.0 and RandomForest (rf) shows remarkable accuracy of 90 % and more.

Classifiers- Comparing the result on variable importance between C5.0 and RF; C5.0 differentiate the importance between classifiers more clearly. Thus it is decided to use C5.0 to test further with the assumption salary and age to be key classifiers.

Discretization- The need of discretization was also take into account, 5 salary groups is the best way (tested 3 salary groups lessen accuracy to below 0.90) to identify and differentiate groups.

RandomForest	C5.0
0.92 accuracy	0.92 accuracy
0.84 Kappa	0.82 Kappa
Overall	Overall
salary 100.000	salary 100.00
age 46.836	age 85.51
credit 12.314	car 80.97
car 4.653	credit 37.43
zipcode 1.729	zipcode 0.00
elevel 0.000	elevel 0.00

We can conclude the highest salary group with clear preference on Sony,

and different age with different salary group can also see clear patterns.

After discretization, C5.0 shows 0.92 accuracy and 0.83 Kappa performance.

Age, Salary towards Brand preference
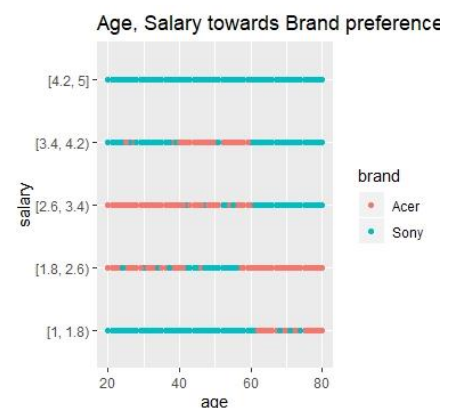
As a result, my optimized model would be C5.0 algorithm with 2 independent variables, Age and Salary. Salary discretize to 5 groups to predict my dependent variable (Brand preference).

```
> summary(focusBrand)
     salary                age               brand
[20000, 46000)  :2024  Min.   :20.00    Acer:3744
[46000, 72000)  :1912  1st Qu.:35.00    Sony:6154
[72000, 98000)  :2029  Median :50.00
[98000, 124000) :1933  Mean   :49.78
[124000, 150000]:2000  3rd Qu.:65.00
                       Max.   :80.00
```

Performance metric		
	Reference	
Prediction	Acer	Sony
Acer	829	90
Sony	112	1443
Accuracy : 0.9184		
P-Value: <2e-16		
Kappa : 0.826		

Applying this prescriptive model on the missing "incompleteresponse" set, we have concluded final prediction with 1909 on Acer and 3091 on Sony which shows the same distribution pattern of the complete data set. In conclusion, 62% of customers prefer Sony while 38% customer prefer Acer.

In regards of strategic relationship with these two brands. Taking into account the general brand image and price structure, Sony is considered a premium brand while Acer to be more economical and student based. It is logical that the highest income group only purchase Sony. It will definitely interesting to compare our own sales report with the report from this two brands and find out why Sony also sales better in low income, young group?

This is only an prediction on brand preference, to understand and further research on marketing campaigns could also take educational level(Q3) and see if there are regional behavior(Q5) like we have found out before.