Lessons Learned

# iPhone

The classifier selected was the C5.0 decision tree with 20 features. The 20 features are the following:

1. Iphone
2. Samsunggalaxy
3. Htcphone
4. Iphonedisunc
5. Googleandroid
6. Iphoneperpos
7. Iphonedispos
8. Sonyxperia
9. Iphonecampos
10. Ios
11. Htccampos
12. Iphoneperunc
13. Iphonedisneg
14. Htcdispos
15. Iphonecamunc
16. Iphoneperneg
17. Sonyperpos
18. Iphonecamneg
19. Htcperpos
20. Samsungperpos

The C5.0 classifier was selected because it had the best accuracy and kappa.

Accuracy Kappa

0.7733231 0.5611647

The features eliminated were selected through a recursive feature elimination which took around 3 hours to complete with a sample of 1000 observations.

Chart, bar chart

Description automatically generated

Altering the dependent variable worked well because it increased both metrics significantly. PCA didn’t work as well as altering the dependent variable because it hardly increased both metrics.

# Galaxy

The classifier selected was the C5.0 decision tree with 19 features. The 20 features are the following:

1. Iphone
2. Googleandroid
3. Htcphone
4. Samsunggalaxy
5. Iphonedisunc
6. Iphonedispos
7. Htccampos
8. Iphoneperpos
9. Iphonecampos
10. Iphonedisneg
11. Sonyxperia
12. Iphonecamneg
13. Iphoneperunc
14. Ios
15. Htcdispos
16. Iphoneperneg
17. Iphonecamunc
18. Sonyperpos
19. Htcperpos

The C5.0 classifier was selected because it had the best accuracy and kappa

Accuracy Kappa

0.7618802 0.5170860

The features eliminated were selected through a recursive feature elimination which took around 3 hours to complete with a sample of 1000 observations.

Chart, bar chart

Description automatically generated

Altering the dependent variable worked well because it increased both metrics significantly. PCA didn’t work as well as altering the dependent variable because it hardly increased both metrics.

The process for similar projects should execute the RFE for feature selection, a quick model to verify everything is working fine, and other 4 or 5 models of the analysts choosing to compare and select the best model.