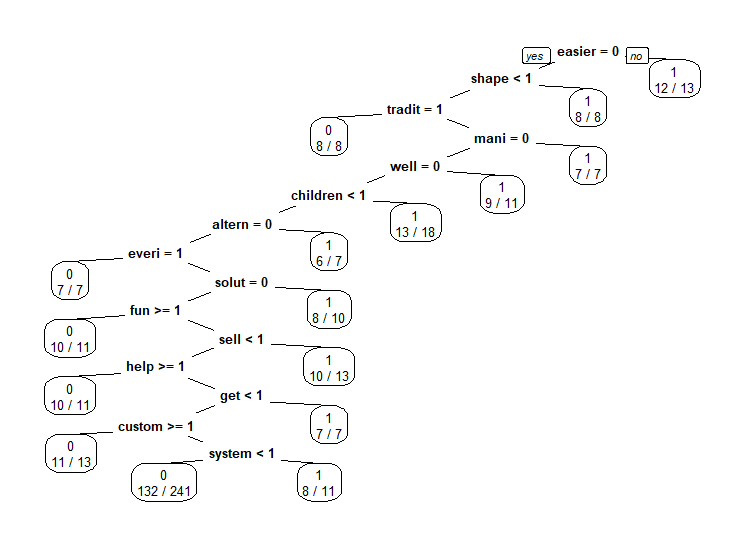
**Business Intelligence Using Text Mining**

**Problem Description:** A dataset of Shark Tank episodes is made available. It contains 495 entrepreneurs making their pitch to the VC sharks. we will be using classification techniques such as CART, Random Forest and logistic regression to predict the final outcome Deal or No-deal (1,0) based on 3 attributes which are Description of the product, Deal and Ratio (asked for valued).

**CART Before and After taking Ratio:**

**Before Tree:**



Confusion Matrix and Statistics:

Reference

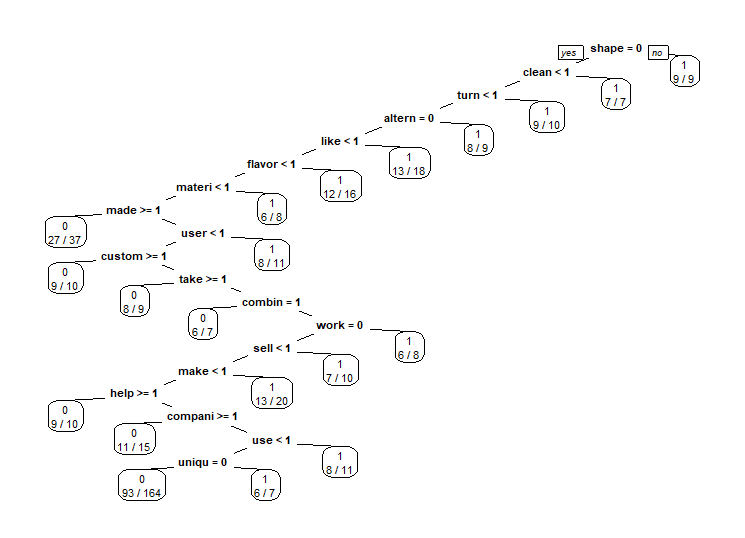
Prediction 0 1

0 40 39

1 9 11

Accuracy : 0.5152

After Tree:



Confusion Matrix and Statistics:

Reference

Prediction 0 1

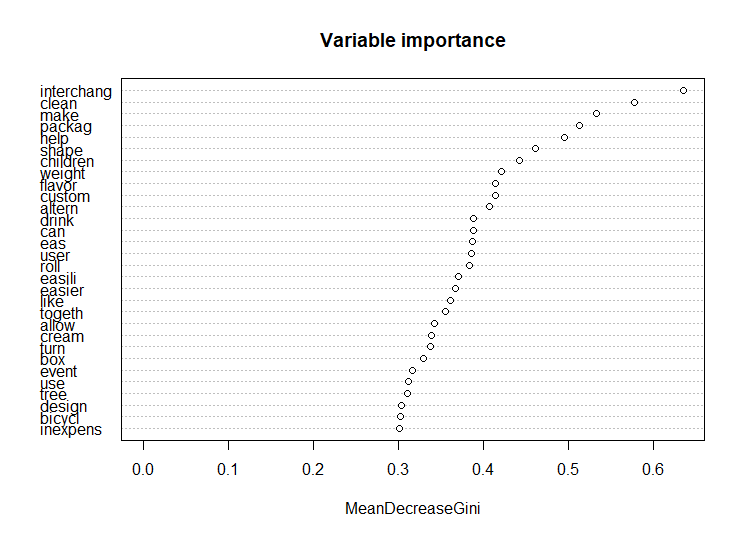
0 32 27

1 17 23

Accuracy : 0.5556

Random Forest Before and After Ratio:

Before Plot:



Confusion Matrix and Statistics:

Reference

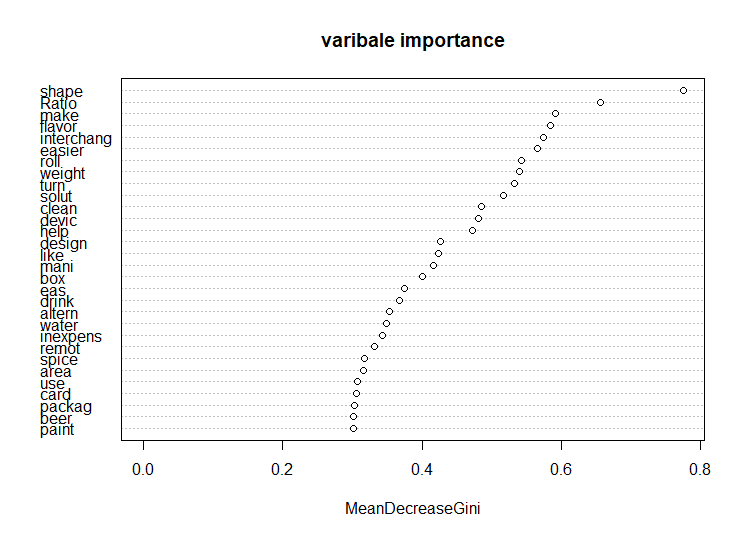
Prediction 0 1

0 42 37

1 7 13

Accuracy : 0.5556

After Plot:



Confusion Matrix and Statistics:

Reference

Prediction 0 1

0 43 32

1 6 18

Accuracy : 0.6162

Logistic Regression Before and After Ratio:

Before Confusion Matrix:

Prediction

Reference FALSE TRUE

0 24 25

1 21 29

Accuracy : 0.5353

After Confusion Matrix:

Prediction

Reference FALSE TRUE

0 17 32

1 18 32

Accuracy : 0.4949

Interpretation: Based on the models used to predict the outcome variable “Deal”, we can conclude that Random Forest with the Ratio as an Independent variable along with words such as shape, Ratio, make, being the top 3 and we can predict this with 62% accuracy as compared with other models built which had a range of 50-55% in accuracy.