Suppose we are interested in the factors that influence whether a political candidate wins an election.

The outcome (response) variable is binary (0/1); win or lose.

The predictor variables of interest are the amount of money spent on the campaign,

the amount of time spent campaigning negatively and whether or not the candidate is an incumbent.

Business Problem:

To predict whether the political candidate wins an election.

Data Cleaning:

Remove the ‘Na’ values from the dataset as the first row of the dataset consists of all the ‘Na’ values.

Remove the first column of the dataset as it is the id.

Build Model:

model <- glm('Result~Year+Amount.Spent+Popularity.Rank',data = election,family = "binomial")

Summary:

Null deviance: 1.3460e+01 on 9 degrees of freedom

Residual deviance: 6.5897e-10 on 6 degrees of freedom

AIC: 8

we observe difference between null and residual dev to compare with different models.

Confusion Matrix:

Classification and regression training(caret)

caret contain the process to streamline the process for complex regression and classification problem.

By using the caret package we can know all the required information about the confusion matrix.

#Confusion Matrix and Statistics

0 1

0 4 0

1 0 6

Accuracy: 1

95% CI: (0.6915, 1)

No Information Rate: 0.6

P-Value [Acc > NIR] : 0.006047

Kappa: 1

Mcnemar's Test P-Value : NA

Sensitivity: 1.0

Specificity: 1.0

Pos Pred Value: 1.0

Neg Pred Value: 1.0

Prevalence: 0.4

Detection Rate: 0.4

Detection Prevalence: 0.4

Balanced Accuracy: 1.0

'Positive' Class: 0

ROC Curve:

roc curve is used to evaluate the betterness of the logistic model.

more the area under the curve better the model is.

we will use roc curve for any classification technique not only for logistic model.

