

GOA ELECTIONS 2017

Data Source:

- Election Commission website
- My Neta website

The data set arrived is obviously an unbalanced data set with respect to the Lose (0) - Win (1) ratio.

```
      0      1
0.84 0.16
```

Models tried out to analyse the election results:

- Linear Regression
- Logistic Regression
- Random Forest
- CART

Results:

1. Linear Regression:

A Linear Regression model run keeping Percentage of votes polled as dependant variable shows that the significant variables on which a person's win depended are

- Assets
- Liabilities

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -7.661e+00  5.219e+00  -1.468 0.143388
Candidate.Age  3.918e-01  1.081e-01   3.624 0.000353 ***
Criminal.Case  3.142e+00  1.757e+00   1.789 0.074897 .
Total.Assets   5.919e-08  1.396e-08   4.240 3.17e-05 ***
Liabilities    1.444e-07  5.178e-08   2.789 0.005709 **
---
```

2. Logistic Regression:

A Logistic Regression model run with Win (A person's win) being a dependant variable shows that the significant variables on which a person's win depended are

- Assets
- Liabilities
- Total Assets & Liabilities combined
- Candidate age & Liabilities combined

Coefficients:

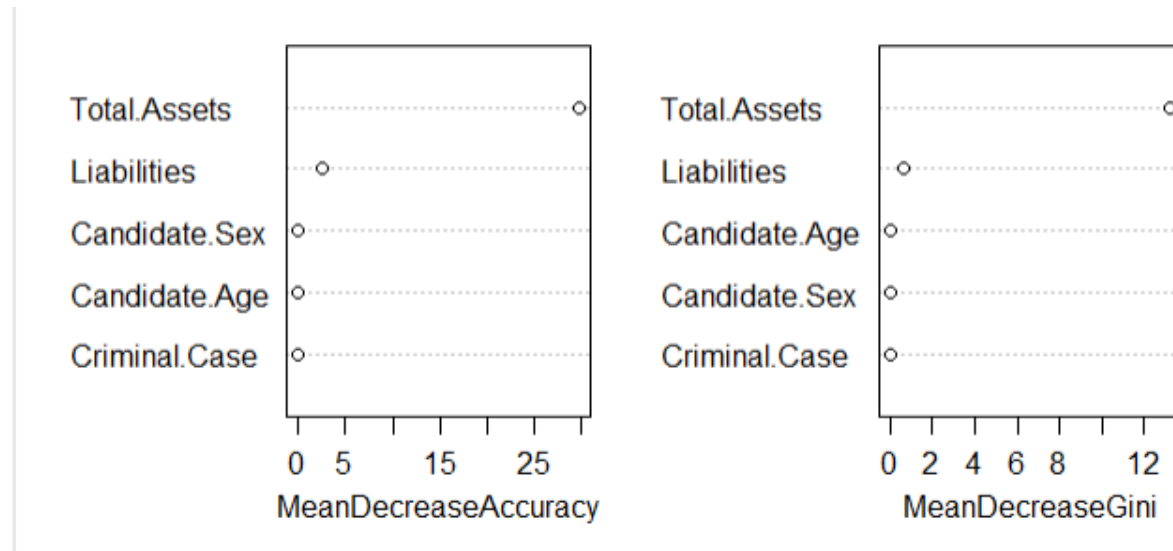
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-6.465e+00	1.922e+00	-3.364	0.000769	***
Candidate.SexM	2.787e-02	1.012e+00	0.028	0.978038	
Candidate.Age	7.381e-02	3.078e-02	2.398	0.016495	*
Criminal.Case	-1.845e+00	1.244e+00	-1.483	0.137961	
Total.Assets	1.618e-08	5.233e-09	3.091	0.001993	**
Liabilities	2.314e-07	7.894e-08	2.931	0.003374	**
Total.Assets:Liabilities	-4.369e-16	1.624e-16	-2.689	0.007157	**
Candidate.Age:Liabilities	-3.147e-09	1.533e-09	-2.052	0.040124	*
Criminal.Case:Total.Assets	1.401e-08	9.081e-09	1.543	0.122887	

The Mc Fadden Value for the above model is

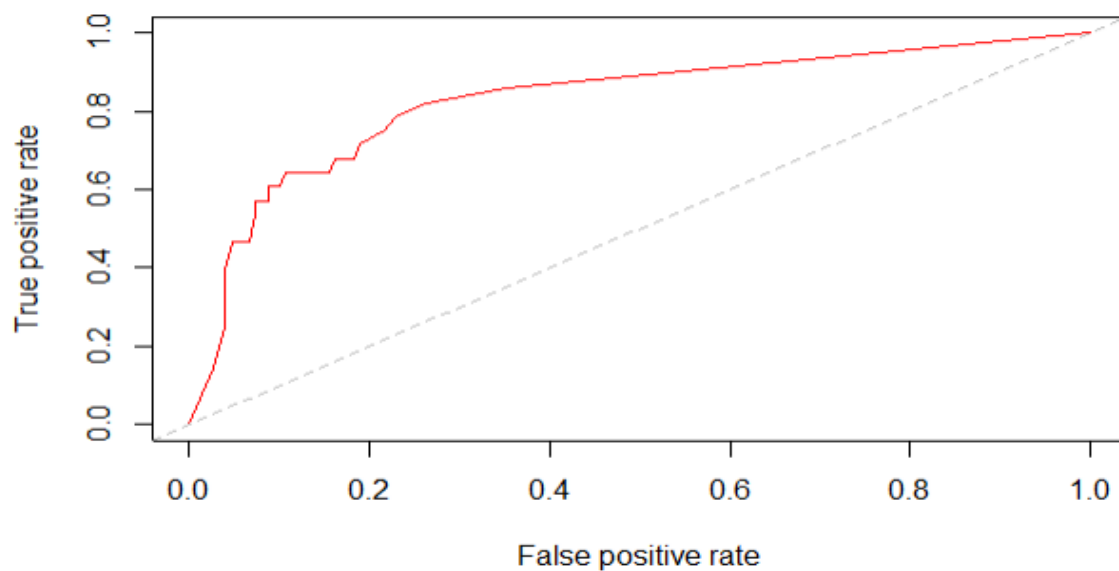
McFadden
0.2646843

3. Random Forest:

This model was particularly chosen to find out the Variable Importance and ROC plot.



Variable Importance



ROC Plot

The KF and AUC values respectively for train data:

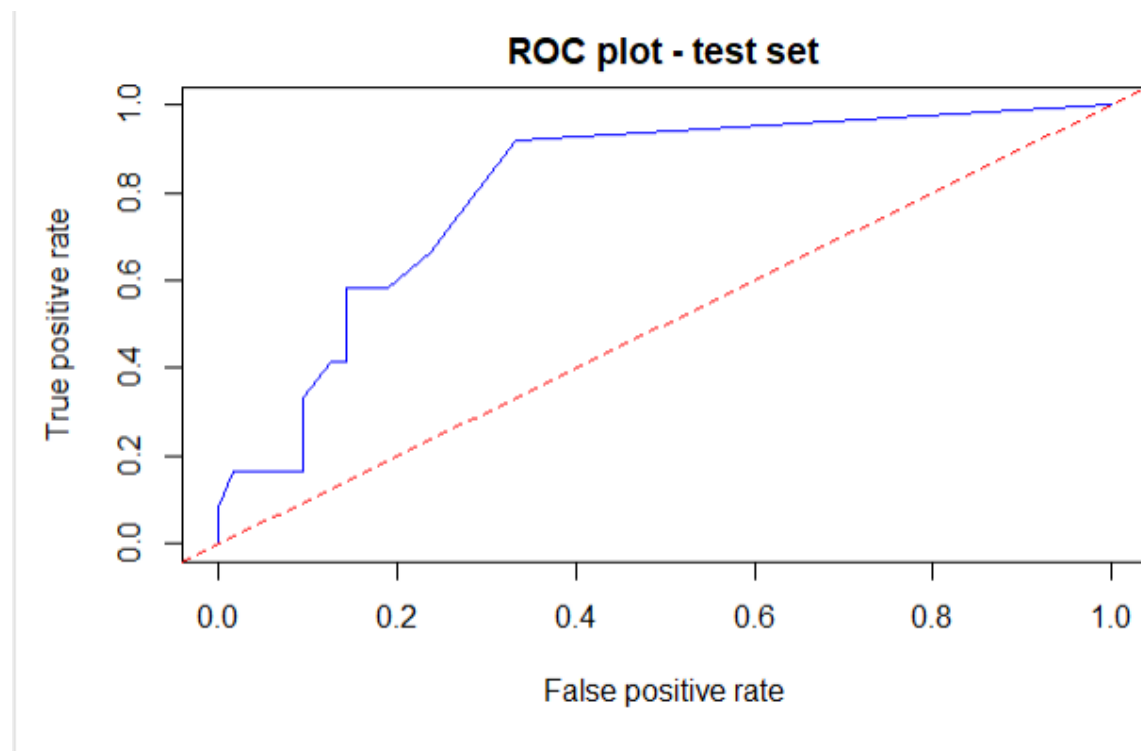
```
[1] 0.5579151  
[1] 0.8270994
```

The KF and AUC values respectively for test data:

```
[1] 0.5833333  
[1] 0.8055556
```

CART Model:

The ROC plot for the test data is as given



ROC Plot

The KS and AUC values for train data respectively are

```
[1] 0.519305  
[1] 0.7596525
```

The KS and AUC values for Test data respectively are

```
[1] 0.5833333  
[1] 0.8055556
```

Conclusion and insights from the data analysed:

- Taking into consideration the results obtained from the models, each model is useful in finding a particular kind of a result viz: one model shows the variable importance, one model shows that criminal records of a person are significant.
- But, all of the models predicted a person's winning ability on the total assets and total liabilities of a person as the most significant factor in his/her winning ability.
- According to the analysis from the above models, a person's gender, and age are not significantly good predictors of his/her win.

