10 Dec,2017 **Report**  **Sai kiran**

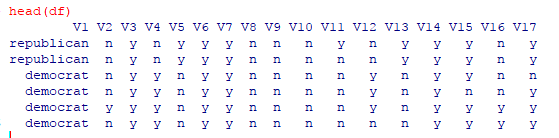
In this project, the algorithms Naïve Bayes and Decision Tree (C5.0) have been implemented.

The few steps performed are:

**Step 1. Read the data and impute missing values with mostly repeated value.**

Read the data into data frame by setting attributes of read.table method as header = False and sep = “,”. Install necessary packages for imputing data. By the end of this step we have clean data in the data frame df and that data is available for analyses in further steps.

Output:



**Step 2. Implement two algorithms and analyze, compare their performances.**

Once the data was prepared by imputing the missing values, the two algorithms i.e. Naïve Bayes

and Decision Tree was applied.

For both the algorithms the data was divided in test and train sets as 135 for test data and 300 for train data.

The Decision Tree algorithm has accuracy of 98.9%.

It predicted 87 out of 88 democrats and 46 out of 47 republicans correctly.



On the other hand, The Naïve Bayes algorithm has accuracy of about 94%.

It predicted 73 out of 75 democrats and 53 out of 61 republicans correctly.



However, the performance of the Decision Tree C5.0 was much better than the Naïve Bayes.

**Step 3. Implementing 10-fold Cross-Validation to estimate how well the two algorithms perform.**

To implement the Cross-Validation we need caret package, and the number of folds was set to 10.

After implementation of the cross-validation method,

the Decision Tree algorithm gave an accuracy of 0.9068906 and



the Naïve Bayes algorithm gave an accuracy of 0.7984631.



However, the performance of the Decision Tree C5.0 was much better than the Naïve Bayes.

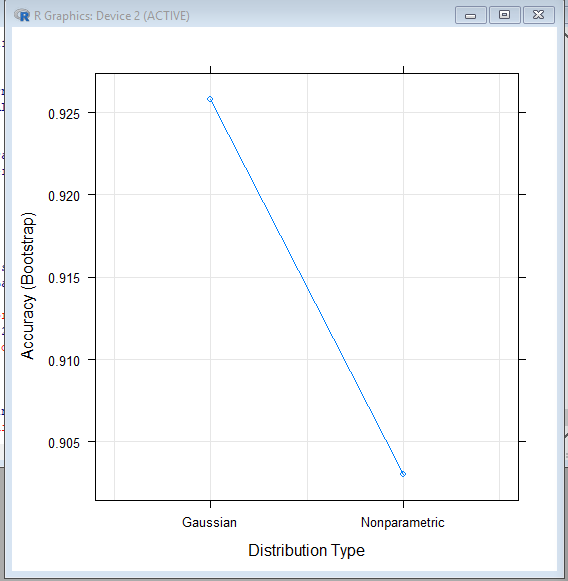
**Step 4. Perform Automated Parameter Tuning for both models using the caret package.**

To perform the automated tuning, we need train () function and the method parameters are set to “nb” and” C5.0” for the Naïve Bayes and Decision Tree algorithm.

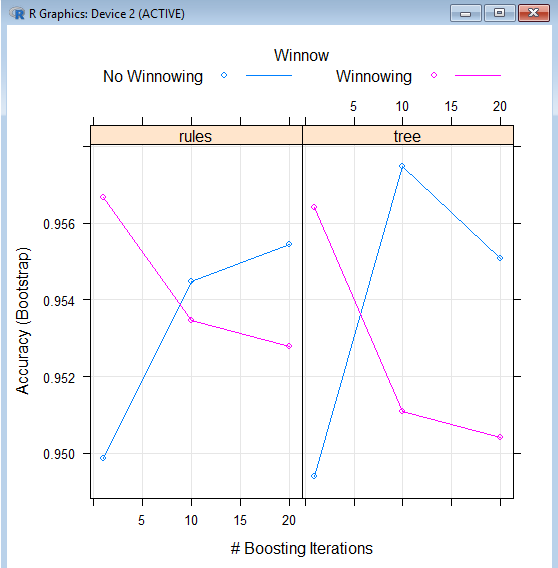
After the automated parameter tuning was performed

The plots for both the algorithms are mentioned below:

the Decision Tree produced an accuracy of almost .95 and kappa value of .89.



the Naïve Bayes algorithm produced an accuracy of 0.90 and the kappa value was 0.79.



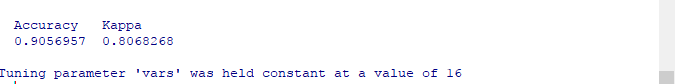
**Step 5. Improve the performance of each algorithm by ‘bagging” (ensemble learning) and the caret package.**

For the ensemble learning we need to use Bagging method.

However, for the Decision tree algorithm when Bagging was applied, it produced an accuracy of .958 and kappa value of .912.



When Bagging was implemented on the Naïve Bayes algorithm an accuracy of 0.905 and the kappa value of 0.806 was produced.



The performance of the Decision Tree algorithm C5.0 was better than the Naïve Bayes algorithm.

Finally, we can conclude that Decision Tree algorithm always performed better than the Naïve

Bayes algorithm in all the cases.