Note:

Note that the best parameters procuded an accuracy of 91% which is not significantly deterred from the accuracy of the original random forest, which is pegged around 92%

Conclusion :

The best model to predict the churn is observed to be Random Forest based on the accuracy as a performance measure.

The incoming calls (with local same operator mobile/other operators mobile/fixed lines, STD or Special) plays a vital role in understanding the possibility of churn. Hence, the operator should focus on incoming calls data and has to provide some kind of special offers to the customers whose incoming calls turning lower.

Details:

After cleaning the data, we broadly employed three models as mentioned below including some variations within these models in order to arrive at the best model in each of the cases.

Logistic Regression :

Logistic Regression with RFE Logistic regression with PCA Random Forest For each of these models, the summary of performance measures are as follows:

Logistic Regression

Train Accuracy : ~79% . Test Accuracy : ~80%

Logistic regression with PCA

Train Accuracy : ~91% . Test Accuracy : ~92%

Decision Tree with PCA:

Train Accuracy : ~93% . Test Accuracy : ~92%

Random Forest with PCA:

Train Accuracy :~ 91% . Test Accuracy :~ 92%