TELECOM CHURN CASE STUDY

M KRITHIKA

PROBLEM STATEMENT

In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another. In this highly competitive market, the telecommunications industry experiences an average of 15-25% annual churn rate. Given the fact that it costs 5-10 times more to acquire a new customer than to retain an existing one, **customer retention** has now become even more important than customer acquisition.

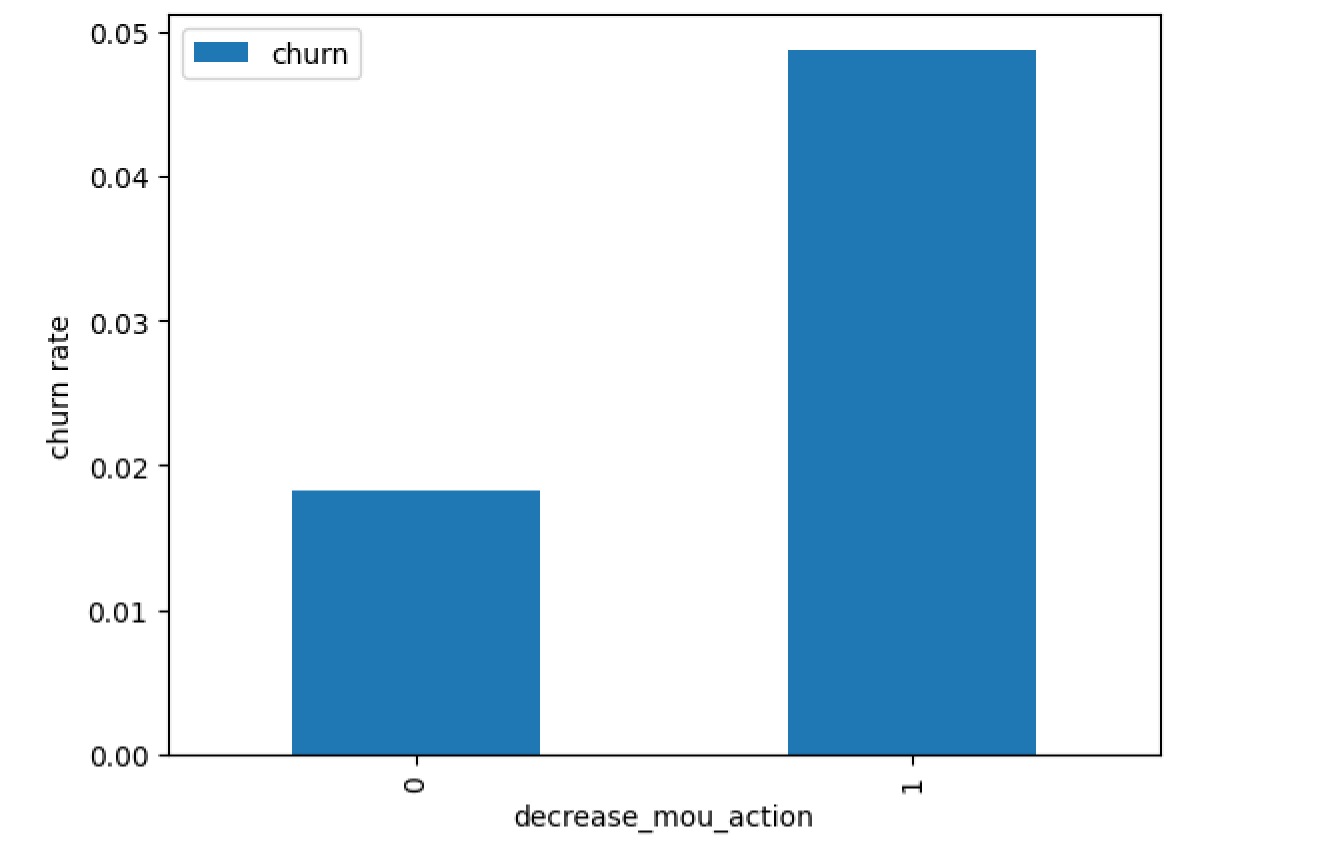
For many incumbent operators, *retaining high profitable customers is the number one business goal.*

To reduce customer churn, telecom companies need to **predict which customers are at high risk of churn.**

In this project, we will analyze customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn.

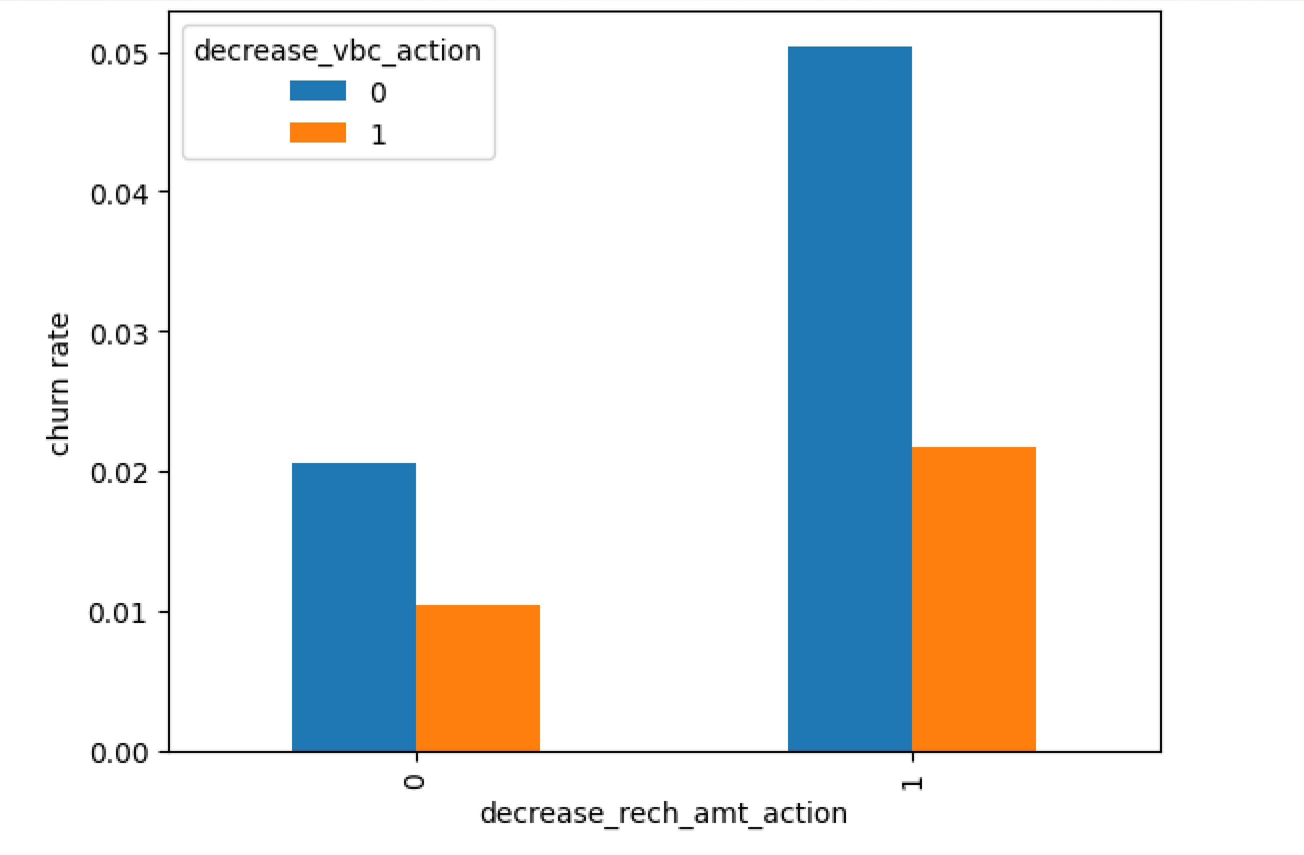
EDA

UNIVARIATE ANALYSIS



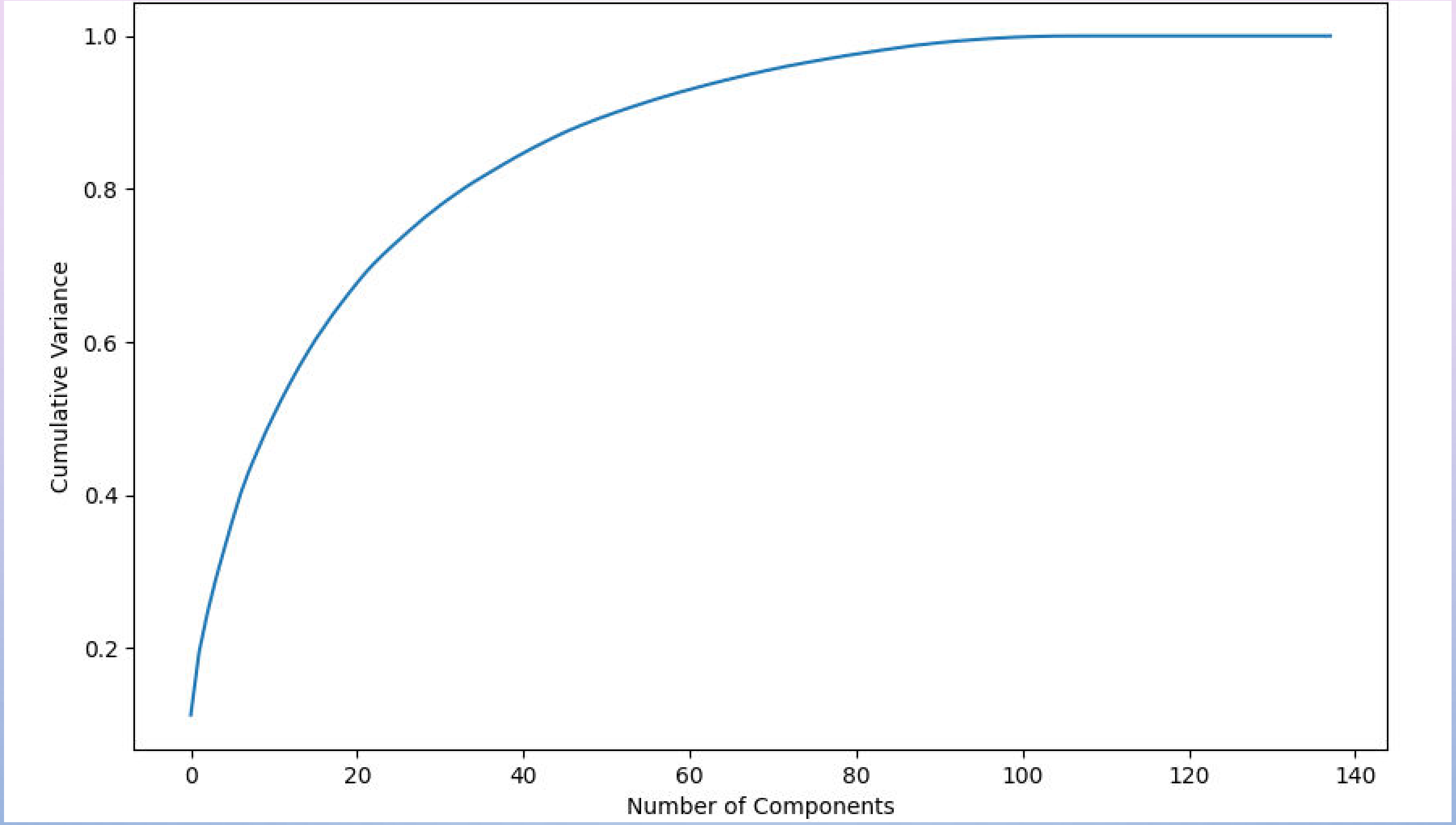
Analysis : The churn rate is more for the customers, whose minutes of usage(mou) decreased in the action phase than the good phase.

BIVARIATE ANALYSIS

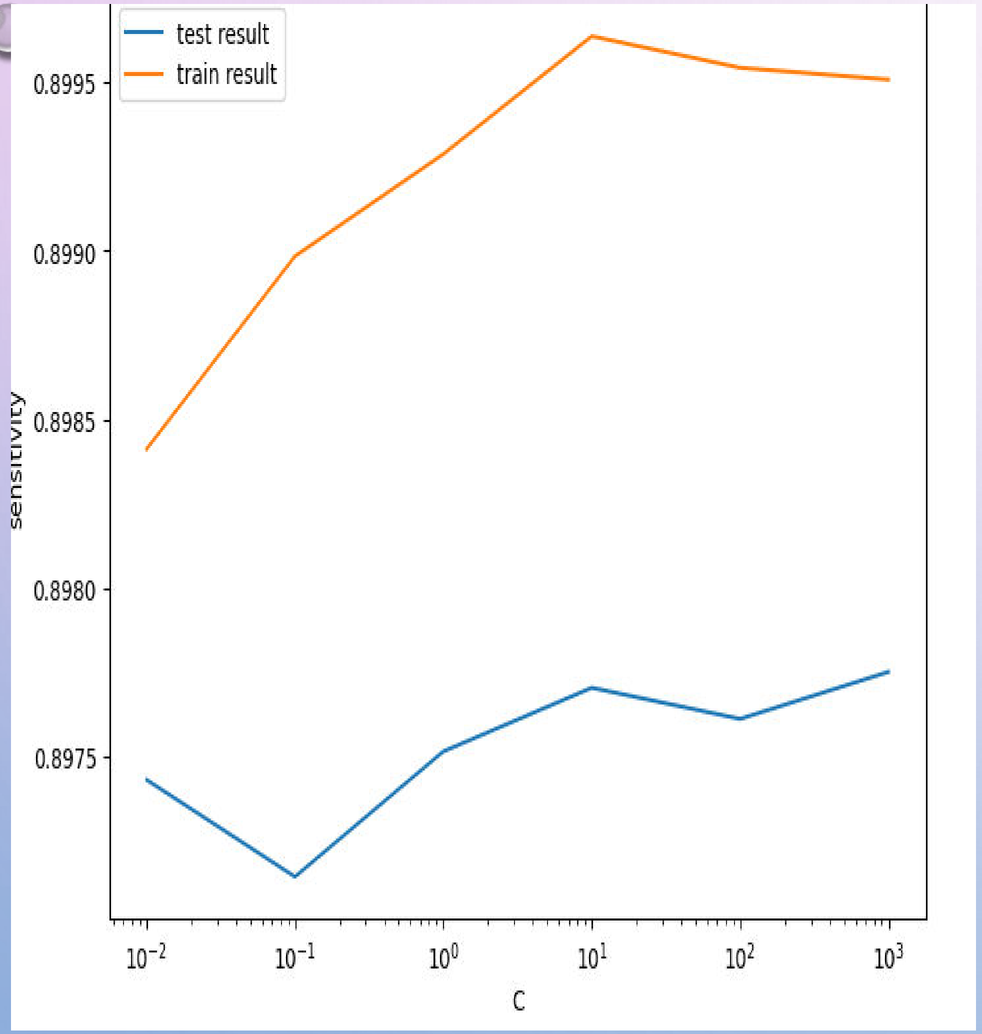


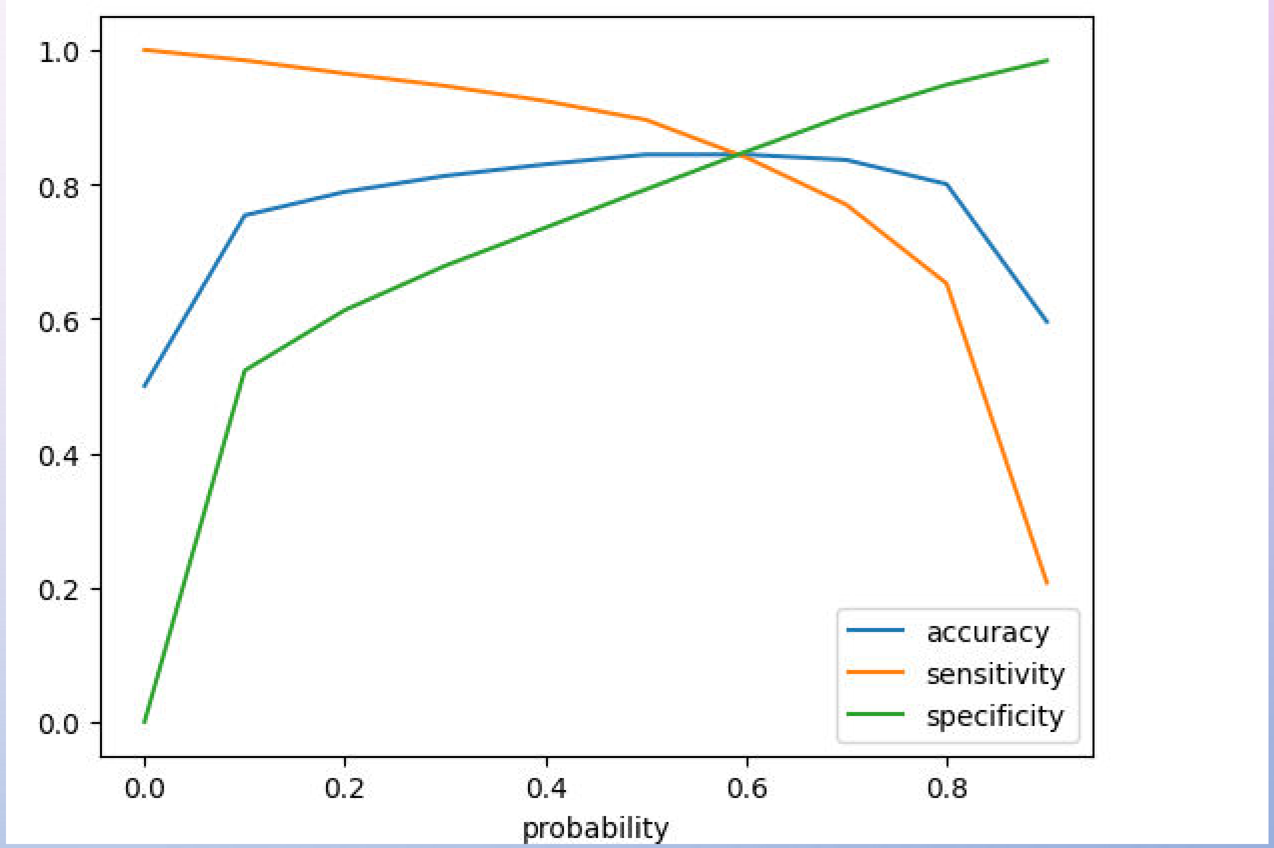
Analysis : The churn rate is more for the customers, whose recharge amount is decreased along with the volume based cost is increased in the action month.

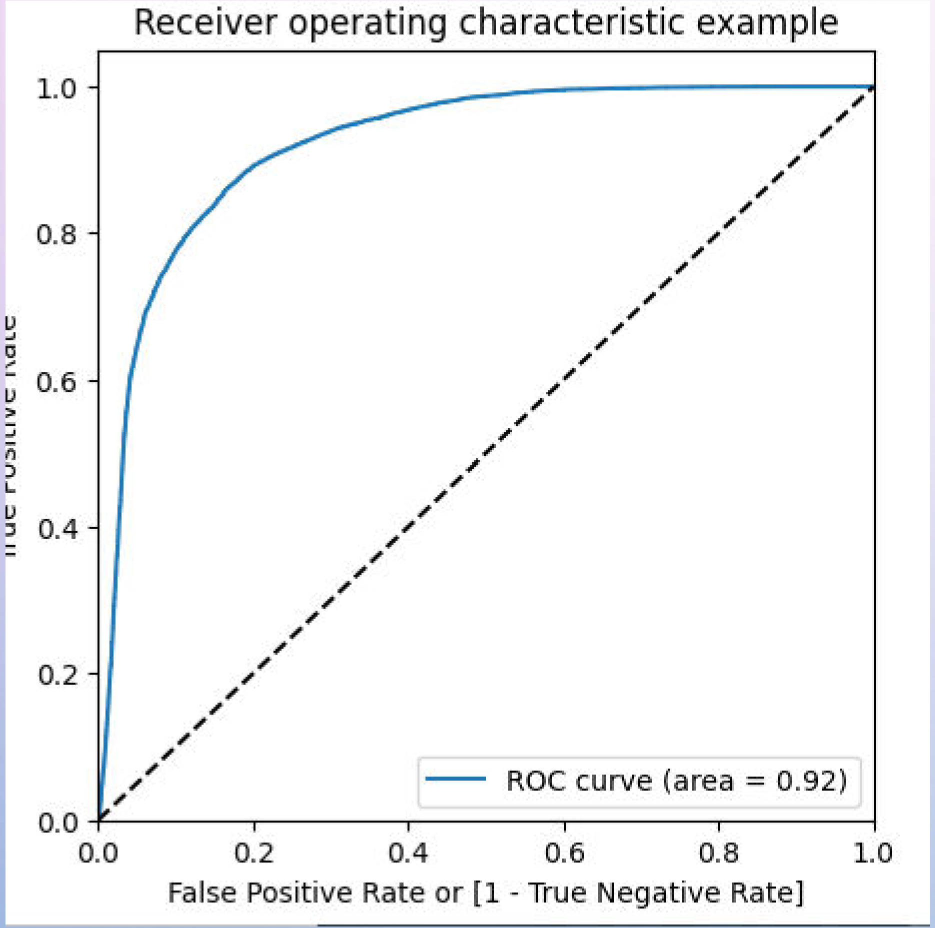
MODEL WITH PCA – CUMULATIVE VARIANCE



plot of C versus train and validation scores






Receiver operating characteristic example
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate or [1 - True Negative Rate]
ROC curve (area = 0.92)

RECOMMENDATIONS

Target the customers, whose minutes of usage of the incoming local calls and outgoing ISD calls are less in the action phase (mostly in the month of August).
Target the customers, whose outgoing others charge in July and incoming others on August are less.Also, the customers having value based cost in the action phase increased are more likely to churn than the other customers.
Hence, these customers may be a good target to provide offer.Cutomers, whose monthly 3G recharge in August is more, are likely to be churned.
Customers having decreasing STD incoming minutes of usage for operators T to fixed lines of T for the month of August are more likely to churn.
Cutomers decreasing monthly 2g usage for August are most probable to churn.
Customers having decreasing incoming minutes of usage for operators T to fixed lines of T for August are more likely to churn.roam_og_mou_8 variables have positive coefficients (0.7135).
That means for the customers, whose roaming outgoing minutes of usage is increasing are more likely to churn.

CONCLUSIONS

The logistic model with no PCA has good sensitivity and accuracy, which are comparable to the models with PCA. So, we can go for the more simplistic model such as logistic regression with PCA as it expliains the important predictor variables as well as the significance of each variable. The model also hels us to identify the variables which should be act upon for making the decision of the to be churned customers. Hence, the model is more relevant in terms of explaining to the business.

For acheiving the best sensitivity, which was our ultimate goal, the classic Logistic regression performs well.For both the models the sensitivity was approx 81%. Also we have good accuracy of apporx 85%.