TELECOM CHURN CASE STUDY

Presented By :
Tejaswini Gude

OBLEM ATEMENT

The Indian and SEA market telecom market is highly competitive, where the industry's customer churn rate is on an average of 15%-25% annually. As the new customer acquisition is 5-10 times more costlier than customer retention, it is imperative that the organisation is focusing more upon the customer retention.

In this industry, retaining the high profitable customers is the primary business goal. To retain customer and reduce the churn, the companies need to predict which customers are at higher risk.

Here the organisation wants to identify the customers who are high value customers and have stopped using the services

BUSINESS GOAL

- The company requires a model to be built for identifying the customers who are high value and have stopped using the services
- To identify the high value customers
- To predict the churn in the last month which is the 9th month.

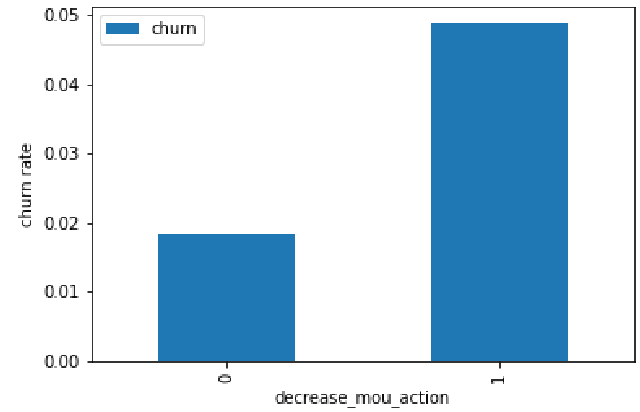
STRATEGY

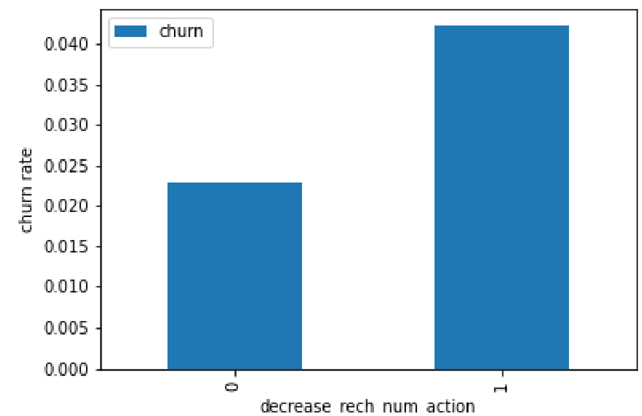
- Import Data.
- Clean and prepare the acquired data for further analysis.
- Filtering out the High Value Customers
- Missing value Treatment
- Exploratory data analysis for figuring out most helpful attributes for conversion.
- Scaling features.
- Handling data imbalance
- Using PCA to reduce dimensionality.
- Building Logistic Regression Model with PCA
- Test the model on train set.
- Evaluate model by different measures and metrics.
- Test the model on test set.
- Measure the accuracy of the model and other metrics for evaluation.
- Checking the performance metrics by using Decision Tree, Random Forest with PCA and Logistic Regression without PCA

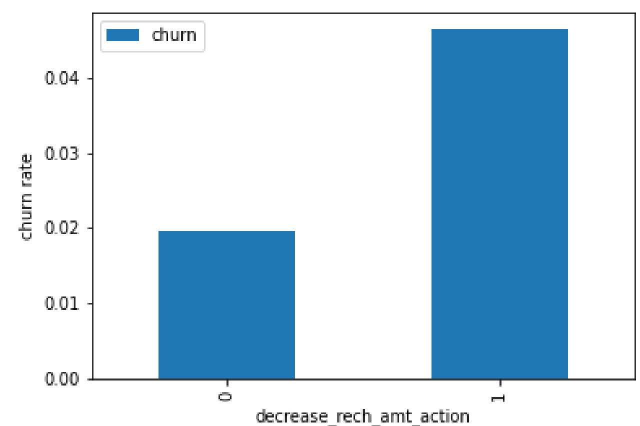
PLORATORY DATA ANALYSIS

viour of Churn Customer in Action Phase

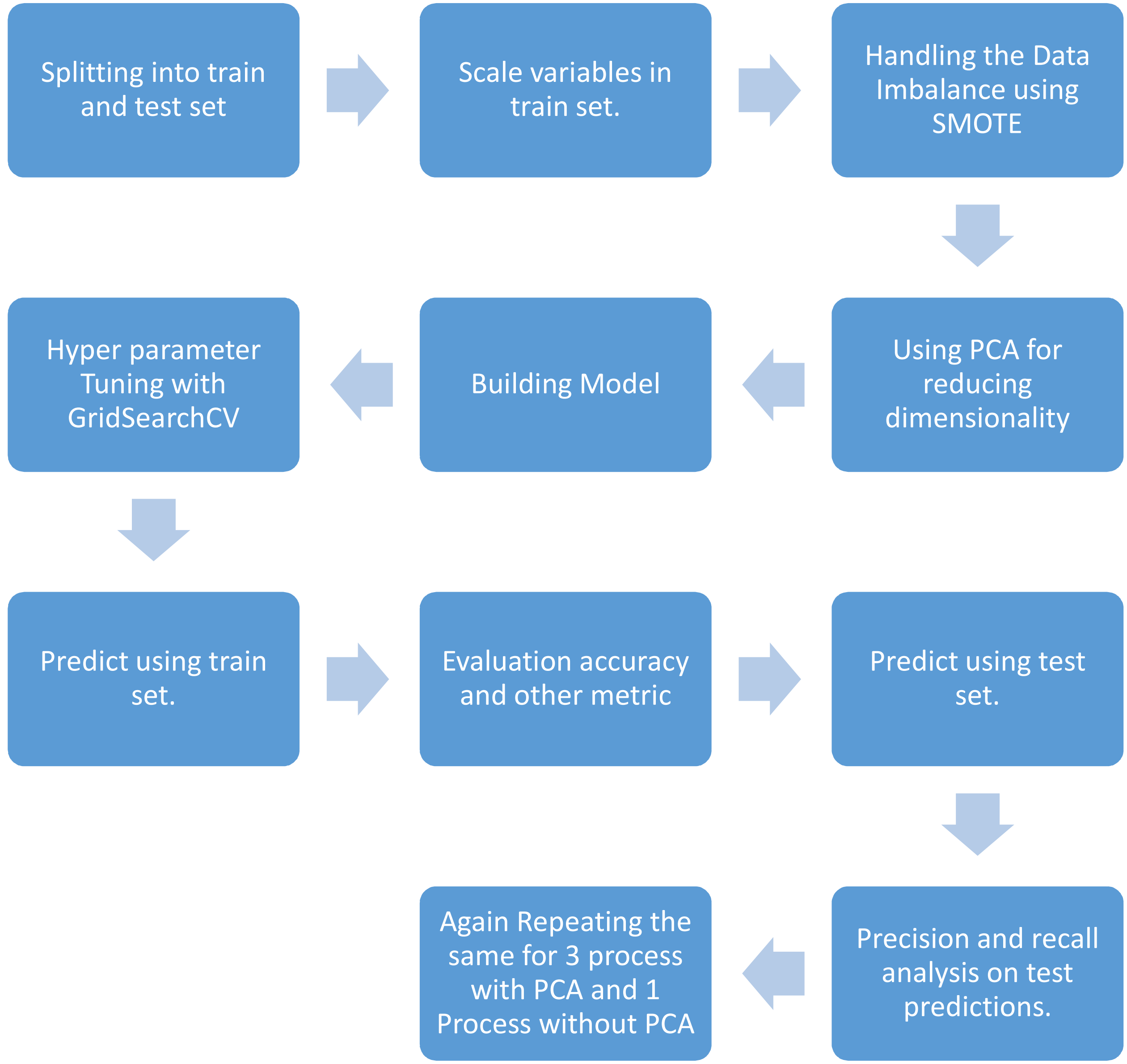
rly depicting that the churn customers have
their average usage time, count of recharge,
of recharge in the third month



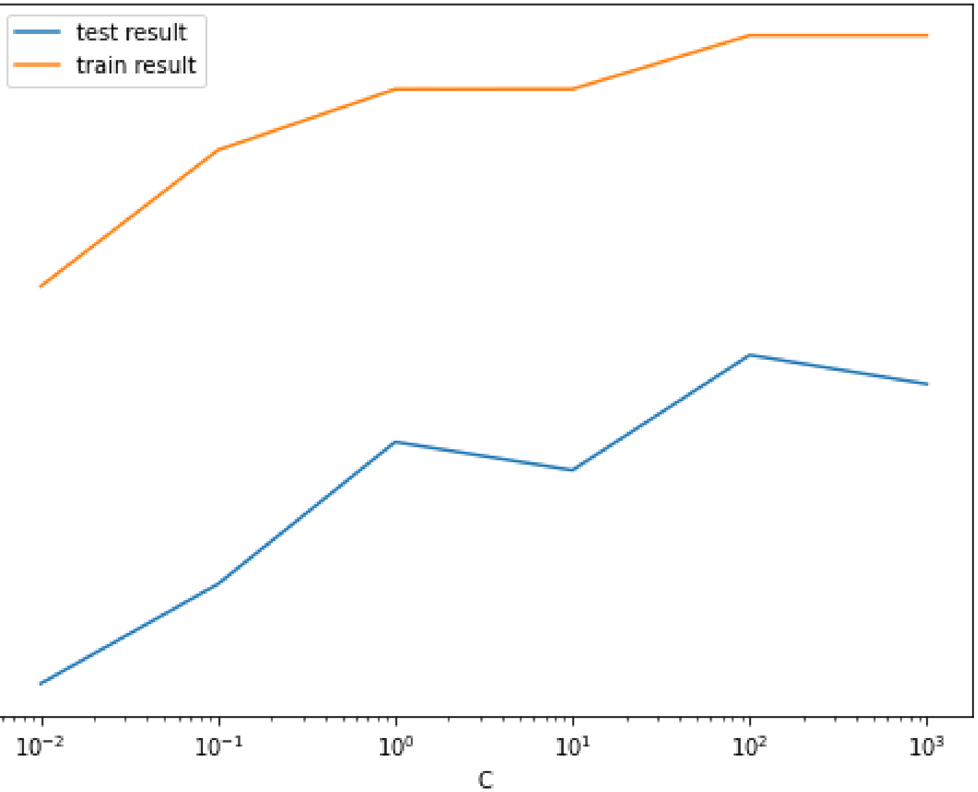




ODEL
UILDING



MODEL EVALUATION
TRAIN with Optimal C



est test sensitivity is 0.8974883072810858 at C = 100

Confusion Matrix

15580	3170
1902	16848

- Accuracy : 86.47%
- Sensitivity : 89.85%
- Specificity : 83.09%

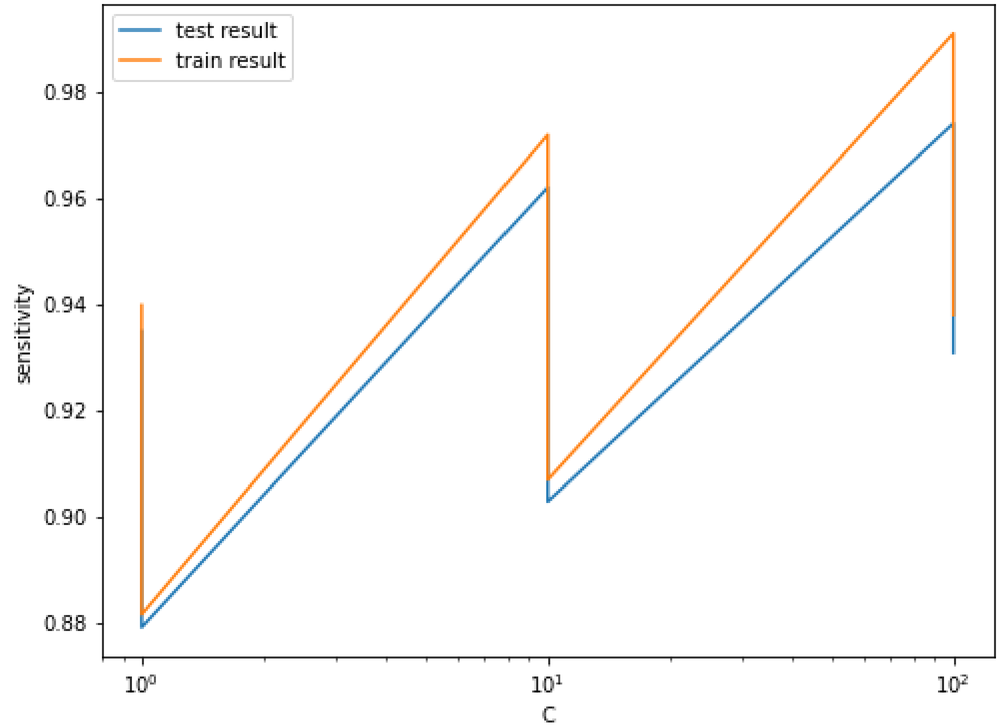
DEL
LUATION-TEST
Optimal C

- Confusion Matrix

6634	1389
54	235

•Accuracy : 82.63%
•Sensitivity : 81.31%
•Specificity : 82.68%

DEL
LUATION-TEST
with PCA
imal C and
Score)



•The highest test sensitivity is 0.9740533333333334 at C = 100

DEL EVALUATION-TRAIN by Decision
e

fusion Matrix

649	2101
324	17426

ccuracy : 90.86%

ensitivity : 92.93%

ecificity : 88.79%

DEL
LUATION-TEST
ecision Tree

- Confusion Matrix

6977	1046
94	195

•Accuracy : 86.28%
•Sensitivity : 67.47%
•Specificity : 86.96%

DEL EVALUATION-TRAIN by Random Forest

fusion Matrix

649	2101
324	17426

ccuracy : 90.86%

ensitivity : 92.93%

pecificity : 88.79%

DEL LUATION-TEST andom Forest

- Confusion Matrix

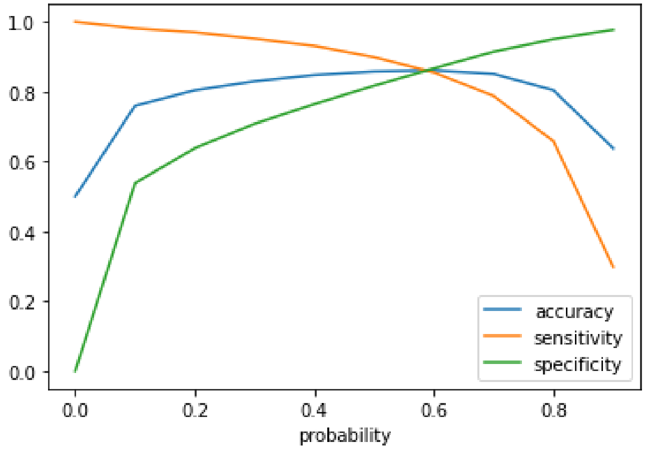
6977	1046
94	195

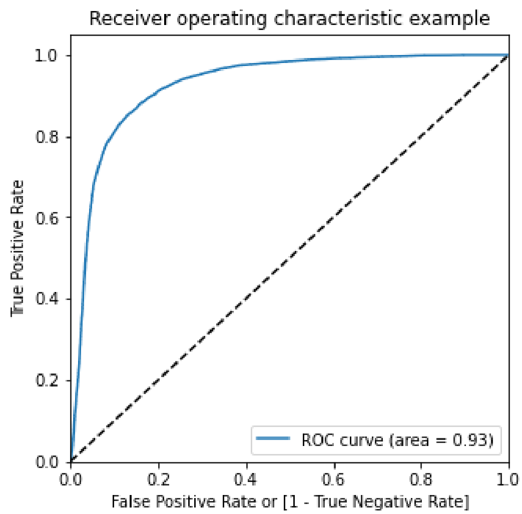
- Accuracy : 86.28%
- Sensitivity : 67.47%
- Specificity : 86.96%

CONCLUSION-Model Building with PCA

After trying several models we can see that for achieving the bes ensitivity, which was our ultimate goal, the classic Logistic regressio or the SVM models preforms well. For both the models the sensitivit vas approx. 85%. Also we have good accuracy of approx. 80%.

MODEL EVALUATION- TRAIN with Logistic Regression





Confusion Matrix

15314	3436
1903	16847

- Accuracy : 85.76%
- Sensitivity : 89.85%
- Specificity : 81.67%

st test sensitivity is 0.8974883072810858 at C = 100

DEL
LUATION-TEST
g Logistic
ession

- Confusion Matrix

6549	1474
47	242

- Accuracy : 81.70%
- Sensitivity : 83.73%
- Specificity : 81.62%

CONCLUSION-Model Building without PCA

Ve can see that the logistic model with no PCA has good sensitivity anc
ccuracy, which are comparable to the models with PCA. So, we can go
or the more simplistic model such as logistic regression with PCA as it
xplains the important predictor variables as well as the significance of
ach variable. The model also hels us to identify the variables which
hould be act upon for making the decision of the to be churned
ustomers. Hence, the model is more relevant in terms of explaining to
he business.

CONCLUSION- Business Recommendation

The customers, whose MoU of the incoming local calls and outgoing ISD calls are less in the action phase (mostly in the month of August) are to be targeted.

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.

So, these customers may be a good target to provide offers.

Customers, whose monthly 3G recharge in August is more, are likely to be churned.

Customers having decreasing STD incoming MoU for operators T to fixed lines of T for th month of August are more likely to churn. Customers 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.

onnet_mou_8 variables have positive coefficients (4.59). That means for the customers, whose all kinds of calls within the same operator are likely to churn.

THANK YOU