TELECOM CHURN CASE STUDY

NINAD V DEOGHARKAR - DSC 58 JULY 2023 BATCH

BUSINESS PROBLEM

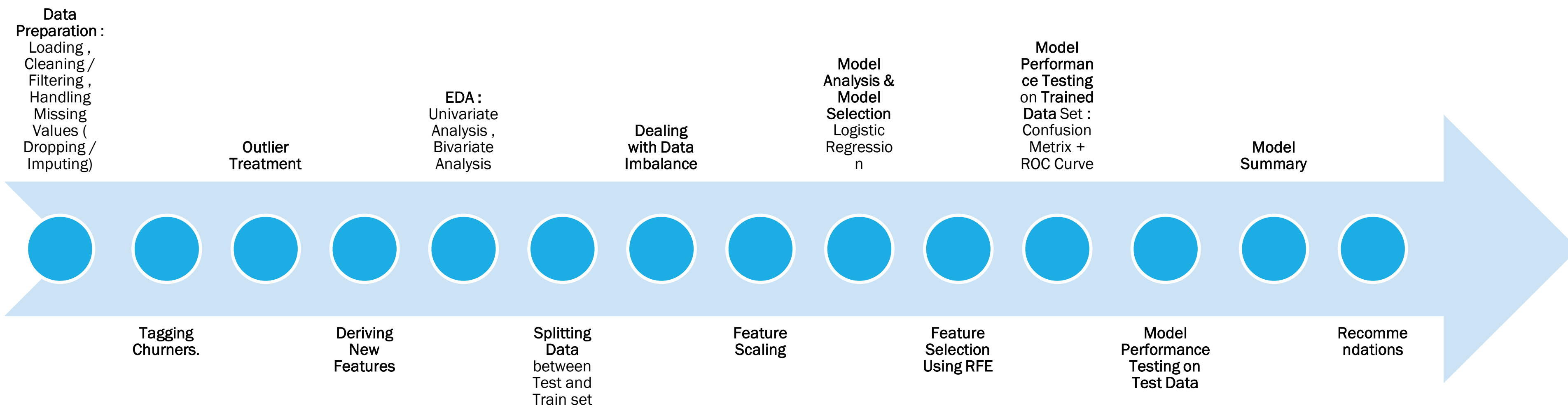
- In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another. In this highly competitive market, the telecommunications industry experiences an average of 15-25% annual churn rate. Given the fact that it costs 5-10 times more 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, you 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.

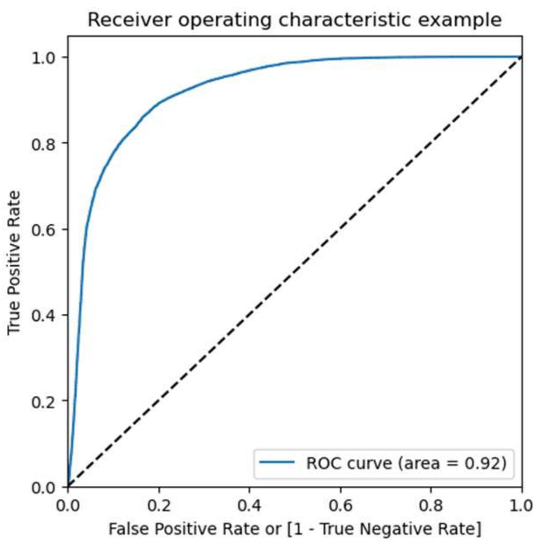
APPROACH

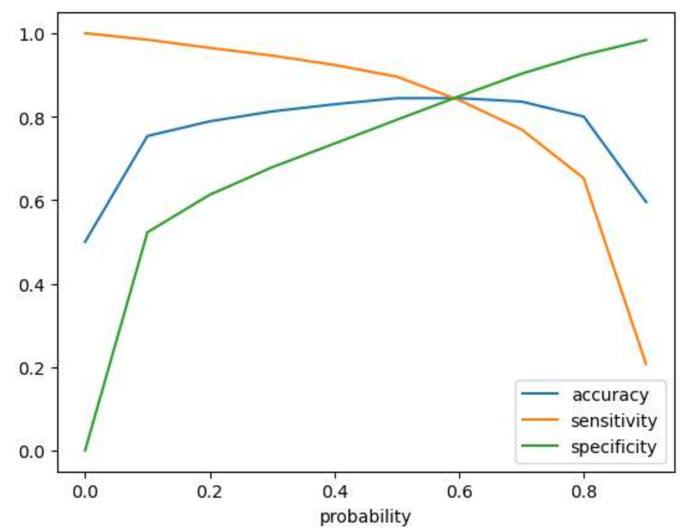


Data Preparation : Loading , Cleaning / Filtering , Handling Missing Values (Dropping / Imputing)
Tagging Churners.
Outlier Treatment
Deriving New Features
EDA : Univariate Analysis , Bivariate Analysis
Splitting Data between Test and Train set
Dealing with Data Imbalance
Feature Scaling
Model Analysis & Model Selection Logistic Regressio n
Feature Selection Using RFE
Model Performan ce Testing on Trained Data Set : Confusion Metrix + ROC Curve
Model Performance Testing on Test Data
Model Summary
Recomme ndations

OUR ANALYSIS

- Using RFE and based on p-value and VIF and after 3 modules we arrived at our final model
- We got 92% as Area Under Curve for ROC Curve
- Optimal Cutoff Point came at 0.6





MODEL SUMMARY

- Train set
  - Accuracy = 0.84
  - Sensitivity = 0.81
  - Specificity = 0.83
- Test set
  - Accuracy = 0.78
  - Sensitivity = 0.82
  - Specificity = 0.78

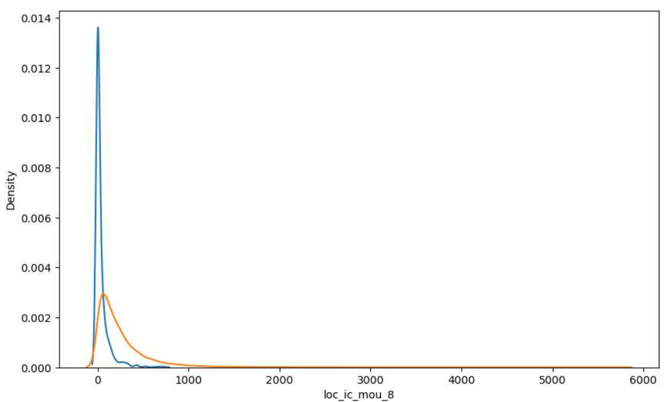
TOP PREDICTORS

- Top Predictors being Negative shows the variables are inversely correlated with the churn probability.

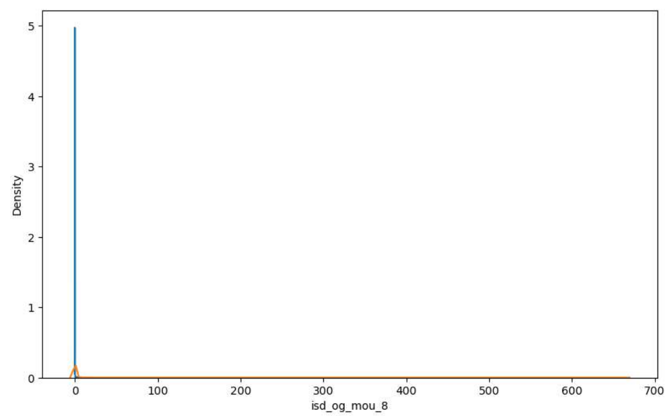
Variables	Coefficients
loc_ic_mou_8	-3.3287
og_others_7	-2.4711
ic_others_8	-1.5131
isd_og_mou_8	-1.3811
decrease_vbc_action	-1.3293
monthly_3g_8	-1.0943
std_ic_t2f_mou_8	-0.9503
monthly_2g_8	-0.9279
loc_ic_t2f_mou_8	-0.7102
roam_og_mou_8	0.7135

PLOTTING CHURN VS NON CHURN

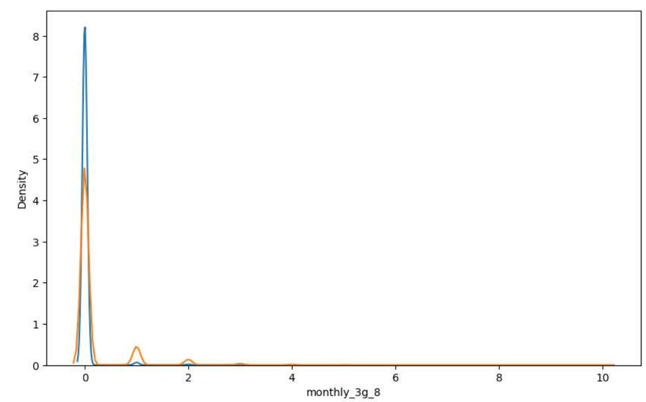
- Top Predictors being Negative shows the variables are inversely correlated with the churn probability.



for the churn customers the minutes of usage for the month of August is mostly populated on the lower side than the non churn customers.



the ISD outgoing minutes of usage for the month of August for churn customers are populated near zero. On the other hand for the non churn customers it is little more than the churn customers.



The number of monthly 3g data for August for the churn customers are populated around 1, whereas of non churn customers it spread across various numbers.

RECOMMENDATIONS

- We must 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).
- Also customers, whose outgoing others charge in July and incoming others on August are less.
- moreover, 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.
- Customers, 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.
- 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.
- 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.