Telecom Churn Case Study

Done by Pratibha Sangam

Problem Statement

As a Business Analyst, we need to **predict which customers are at high risk of churn** so that the Telecom companies can reduce the customer churn.

In this project, we will analyse 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.

Lastly , we will give our **business recommendations**

Business problem overview

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 analyse 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

Steps Involved

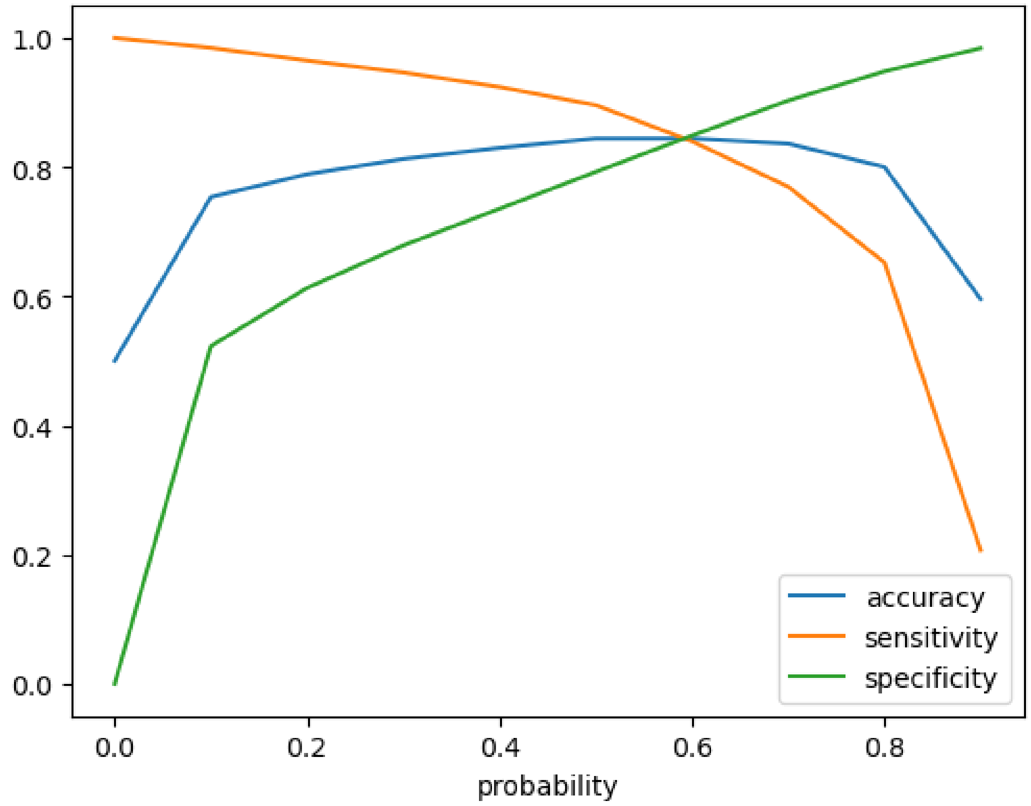
1. Reading, and understanding
2. Visualizing the data
3. Preparing the data for modelling
4. Building the model (Logistic Regression)
5. Evaluating the model

Analysis

- From the model summary and the VIF list we can see that all the variables are significant and there is no multicollinearity among the variables.
- Hence, we can conclude that ***Model-3 lg_3*** *will be the final model*.

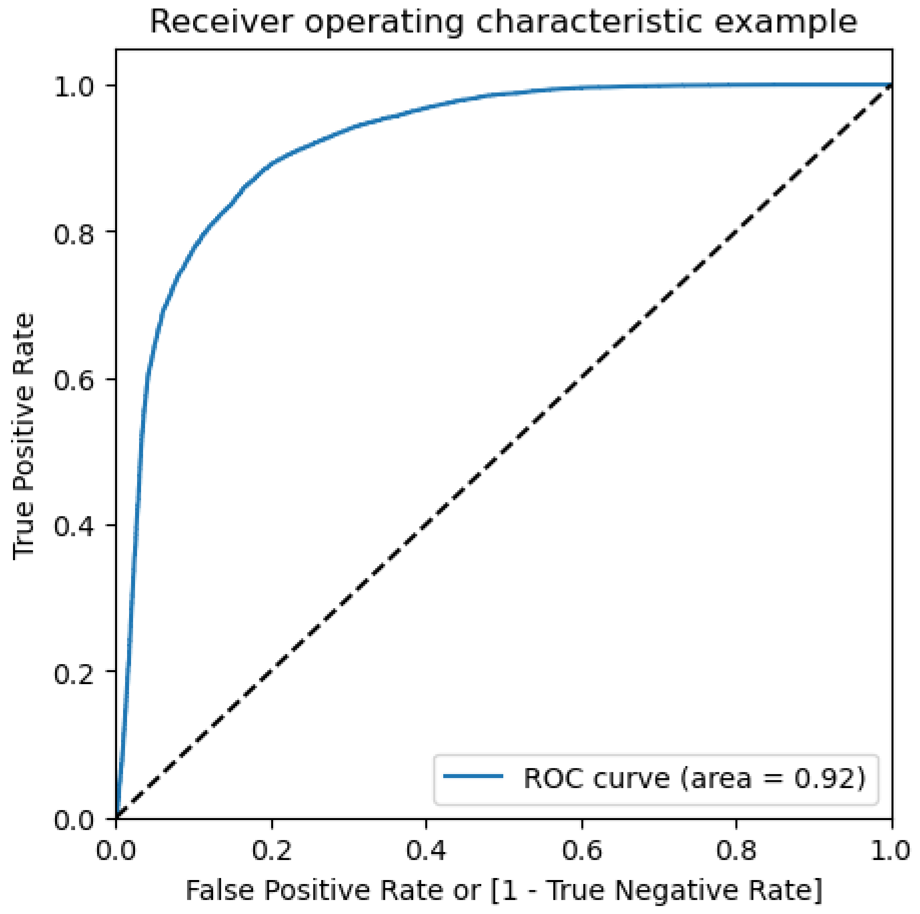
Accuracy, Sensitivity and Specificity curve

- Accuracy - Becomes stable around 0.6
- Sensitivity - Decreases with the increased probability.
- Specificity - Increases with the increasing probability.
- At point 0.6 where the three parameters cut each other, we can see that there is a balance between sensitivity and specificity with a good accuracy.
- Here we are intended to achieve better sensitivity than accuracy and specificity. Though as per the above curve, we should take 0.6 as the optimum probability cutoff, we are taking *0.5* for achieving higher sensitivity, which is our main goal.



ROC Curve (Trade off between sensitivity & specificity)

- We can see the area of the ROC curve is closer to 1, which is the Gini of the model.



Model summary

- Train set
  - Accuracy = 0.84
  - Sensitivity = 0.81
  - Specificity = 0.83
- Test set
  - Accuracy = 0.78
  - Sensitivity = 0.82
  - Specificity = 0.78
- Overall, the model is performing well in the test set, what it had learnt from the train set.

Business Recommendations

- **Top predictors**

Figure shows few top variables selected in the logistic regression model.

- We can see most of the top variables have negative coefficients. That means, the variables are inversely correlated with the churn probablity.
- E.g.:- If the local incoming minutes of usage (loc_ic_mou_8) is lesser in the month of August than any other month, then there is a higher chance that the customer is likely to churn.

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

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.