


Case Study

Assignment -4 (Telecom Churn)

Problem Statement



- Telecom Industry Dynamics: Highly competitive market with multiple service providers and frequent customer switching.
 - Challenge: Annual churn rate of 15-25%, costing 5-10 times more to acquire new customers than retaining existing ones.
 - Importance: Shift from customer acquisition to retention for profitability.
 - Focus: Retaining high-profit customers as a primary business goal.
- 

Data Preparation

1. Imported the library
2. Checking the Data set- *This telecom dataset has 99999 rows and 226 columns*
3. Data Correction - As we can see that the columns with datetime values represented as object, they can be converted into datetime format
4. Handling Missing Value- Handling the missing value across multiple attribute for example- arpu_3g_*,arpu_2g_* for month 6,7,8 and 9 etc

Churn and High Value Customer- 1/2

High Value Customer from Good Phase

1. The 70th quantile value to determine the High Value Customer is: 478.0
2. The total number of customers is now limited to ~30k who lies under the High Value customer criteria basen upon which the model is built.

Defining Churn Value

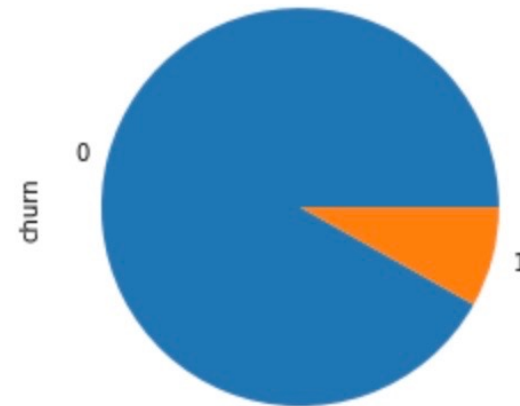
0 - 91.863605

1- 8.136395

As we can see that 91% of the customers do not churn, there is a possibility of class imbalance

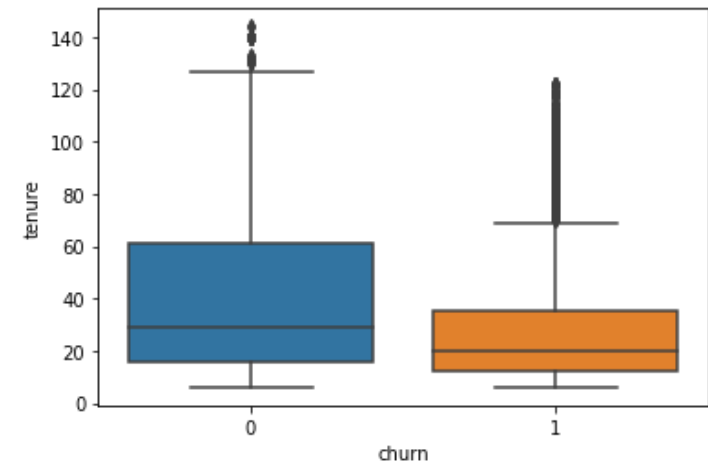
the current dimension of the dataset after dropping the churn related columns (30001, 141)

It can be seen that the maximum churn rate happens within 0-6 month, but it gradually decreases as the customer retains in the network.



Churn and High Value Customer- 2/2

1. Avg Outgoing Calls & calls on roaming for 6 & 7th months are positively correlated with churn.
2. Avg Revenue, No. Of Recharge for 8th month has negative correlation with churn.
3. From the above plot , its clear tenured customers do no churn and they keep availing telecom services
4. As the number of recharge rate increases, the churn rate decreases clearly.



Data Imbalance handling and Logistical Regression Model

Data Imbalance handling

1. Using SMOTE method, we can balance the data w.r.t. churn variable and proceed further

Logistical Regression Model

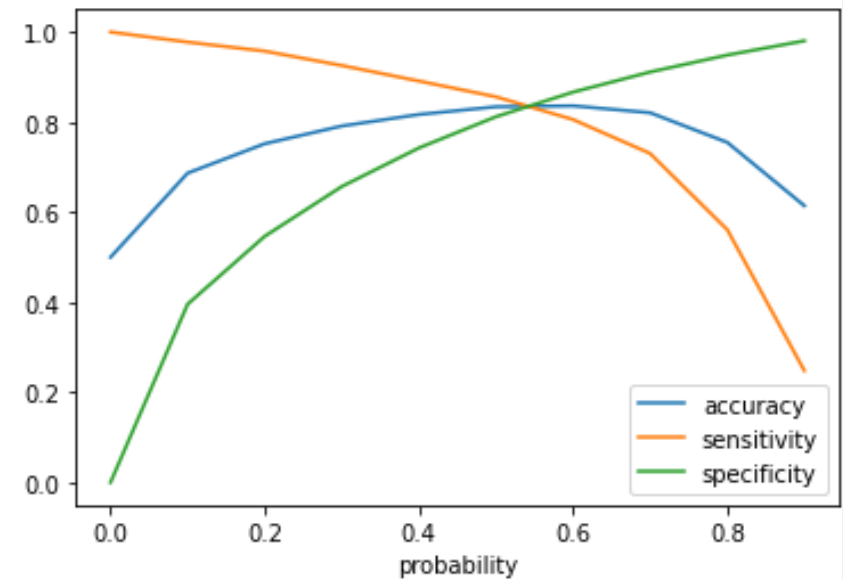
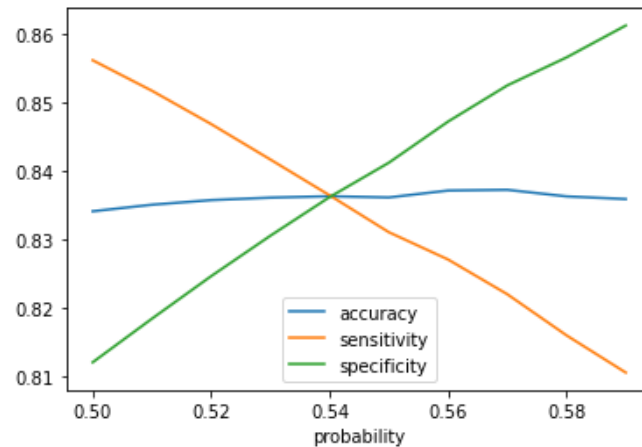
1. Logistic Regression using Feature Selection (RFE method)
2. Assessing the model with StatsModels
3. The overall accuracy of the model is: 0.8340418913313977

Optimum Point of Classification

1. Initially we selected the optimum point of classification as 0.5.

From the above graph, we can see the optimum cutoff is slightly higher than 0.5 but lies lower than 0.6. So let's tweak a little more within this range.

2. From the above graph we can conclude, the optimal cutoff point in the probability to define the predicted churn variable converges at 0.54



Explainign the Results

1. The accuracy of the predicted model is: 83.0 % The sensitivity of the predicted model is: 80.0 % As the model created is based on a sentivity model, i.e. the True positive rate is given more importance as the actual and prediction of churn by a customer
2. The AUC score for train dataset is 0.90 and the test dataset is 0.87.
This model can be considered as a good model.
3. Confusion Matirx for y_test & y_pred [[6250 2022] [185 544]] Accuracy of the logistic regression model with PCA: 0.7548050216642596

Explainign the Results

From above it is clear that the factors affecting the churn are

- total_ic_mou_8 (Total incoming call: Minutes of usage in the action phase),
- total_rech_amt_diff (Total recharge amount difference),
- total_og_mou_8 (Total outgoing call: Minutes of usage in the action phase),
- arpu (Average revenue per user),
- roam_ic_mou_8 (Roaming incoming call: Minutes of usage in the action phase),
- roam_og_mou_8 (Roaming outgoing call: Minutes of usage in the action phase),
- std_ic_mou_8 (STD incoming call: Minutes of usage in the action phase),
- std_og_mou_8 (STD outgoing call: Minutes of usage in the action phase),
- av_rech_amt_data_8 (average recharge amount in the action phase).

Explainign the Results

Steps to help reduce churn

1. Give special; discounts to customers according to their usage
2. Provide additional internet services on recharge.
3. Speak to customers to fulfil their desires.
4. Lower tariffs on data usage,a better 2G area coverage where 3G is not available.
5. Expansion of 3G network where 3G is currently not available.