TELECOM CHURN CASE STUDY

AGENDA

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.

DATA PREPARATION

The following data preparation steps are crucial for this problem:

1. Filter high-value customers: High value costumers are those who have recharged with an amount more than or equal to X, where X is the 70th percentile of the average recharge amount in the first two months (the good phase).

2. Tag churners and remove attributes of the churn phase. We have tagged the churned customers (churn=1, else 0) based on the fourth month as follows: Those who have not made any calls (either incoming or outgoing) AND have not used mobile internet even once in the churn phase.

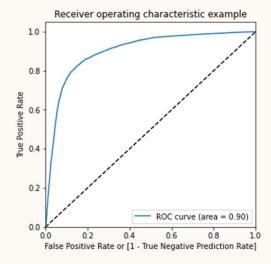3. To be able to start the modelling we had to deal with the data imbalance using Smote method.

MODELLING

We used the logistic regression and after the last iteration of the logistic regression the following results were obtained:

Generalized Linear Model Regression Results

Dep. Variable:			churn	No. Observations:		38576
Model:			GLM	Df Residuals:		38557
Model Family:			Binomial	Df Model:		18
Link Function:			logit	Scale:		1.0000
Method:			IRLS	Log-Likelihood:		-15852.
Date:			Mon, 01 Mar 2021	Deviance:		31704.
Time:			15:41:29	Pearson chi2:		8.51e+10
No. Iterations:			7			
Covariance Type:			nonrobust			

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.5682	0.071	8.055	0.000	0.430	0.706
arpu_8	-7.3871	1.294	-5.709	0.000	-9.923	-4.851
roam_ic_mou_7	8.2919	0.930	8.919	0.000	6.470	10.114
roam_og_mou_8	4.3369	0.602	7.208	0.000	3.158	5.516
loc_og_t2m_mou_8	-3.9987	0.650	-6.156	0.000	-5.272	-2.725
std_og_mou_7	7.7052	0.511	15.082	0.000	6.704	8.707
total_og_mou_8	-20.1259	0.952	-21.139	0.000	-21.992	-18.260
loc_ic_mou_6	9.1605	0.724	12.652	0.000	7.741	10.580
loc_ic_mou_8	-31.5914	1.068	-29.592	0.000	-33.684	-29.499
std_ic_mou_8	-11.9423	1.359	-8.790	0.000	-14.605	-9.280
spl_ic_mou_8	-19.8518	1.375	-14.440	0.000	-22.546	-17.157
total_rech_num_8	-7.1243	0.531	-13.408	0.000	-8.166	-6.083
last_day_rch_amt_8	-18.3312	0.810	-22.622	0.000	-19.919	-16.743
total_rech_data_8	-8.9197	0.612	-14.580	0.000	-10.119	-7.721
av_rech_amt_data_8	-5.2486	0.644	-8.155	0.000	-6.510	-3.987
vol_2g_mb_8	-10.5014	0.934	-11.242	0.000	-12.332	-8.671
monthly_2g_8	-5.7637	0.360	-16.025	0.000	-6.469	-5.059
aug_vbc_3g	-6.8479	0.588	-11.651	0.000	-8.000	-5.696
avg_arpu_6_7	18.3112	1.088	16.823	0.000	16.178	20.445

MODELLING

- We created a column 'churn_pred' with 1 if Churn_Prob > 0.5 else 0.
- The overall accuracy of the model is 0.8340418913313977.
- The following table shows the VIF of the features used for the logistic regression:

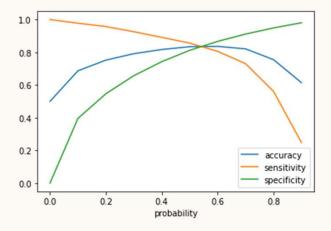|	Features	VIF
8	spl_ic_mou_8	83.90
7	loc_ic_mou_8	42.86
0	arpu_8	18.96
6	loc_ic_mou_6	18.68
5	total_og_mou_8	5.46
12	total_rech_data_8	3.58
4	std_og_mou_7	3.27
8	std_ic_mou_8	2.88
15	monthly_2g_8	2.76
3	loc_og_t2m_mou_8	2.54
14	vol_2g_mb_8	2.06
13	av_rech_amt_data_8	1.76
2	roam_og_mou_8	1.56
16	aug_vbc_3g	1.35
17	avg_arpu_6_7	1.33
1	roam_ic_mou_7	1.30
10	total_rech_num_8	1.15
11	last_day_rch_amt_8	1.05

MODELLING

- To further evaluate the model used we calculated the following:
  - Sensitivity = 0.8561281625881377
  - Specificity = 0.8119556200746578
  - False Positive Rate = 0.18804437992534218
  - Precision = 0.8199106256206554
  - True Negative Prediction Rate = 0.8494792796702104
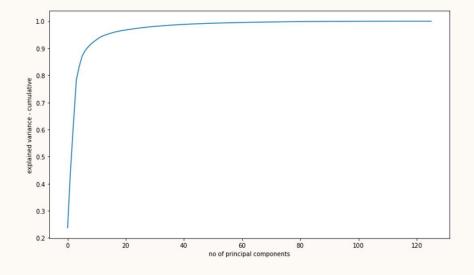
- ROC Curve:

MODELLING

- As a next we calculated the optimal cut-off point by calculating the accuracy and sensitivity of different probabilities and plotted it.

- The graph shows that the optimal cutoff is slightly higher than 0.5 but lies lower than 0.6. After re-working the model it was found that the optimal probability to define the predicted churn should be 0.54.

MODELLING

- We further evaluated the logistic regression model and used PCA. The results showed that accuracy of the logistic regression model with PCA: 0.818131318742362.

- We found out that 90% of the variance of the model could be explained with 8 components as illustrated on the below graph.

BUSINESS RECOMMENDATIONS TO REDUCE CHURN

- We observed that customers that churned usually did it during the first 6 months. Therefore we could recommend the company to design a marketing program focus on creating incentives for costumers to recharge and reduce the churn rate.

- Also, there could be a new marketing program to provide an incentive to make a recharge to those costumers who have not made any recharge in the last 17 days or more.

- We observed that the costumers that decided to churn had an average of 800 days of usage of less. Therefore we know the longer the tenure the lower probability of churn, the company should create a tenure program so that the costumers are motivated to stay longer.

BUSINESS RECOMMENDATIONS TO REDUCE CHURN

- We also observed that once a costumer reduced the recharges and reduced the income calls he was more likely to churn. Therefore we could recommend that once the company has identify the costumers that are recharging less often and reducing the incoming calls they should create an incentive to make them stay. For example offer an additional recharge amount that what the costumers are doing (offer 150 balance for a recharge of 100).

- Another important recommendation is to analyze the strength of the network signal in the coverage areas. A reduction of incoming calls and fewer recharges by costumers could be because they are facing network issues and that's why they decided to churn.