TELECOM CUSTOMER CHURN PREDICTION

Rimjhim Kakkar
Suganya Balasubramanian
Prithviraj Pillai

International Institute of Information Technology, Bangalore
&
UPGRAD

TELECOM CUSTOMER CHURN PREDICTION

BACKGROUND

- 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 become even more important than customer acquisition.
- Identifying these potential customers early on who may voluntarily churn and providing them retention incentives in form of discounts & combo offers will help the organization to retain those customers and reduce revenue loss.
- The company can also internally study any possible operational causes and improve its product offerings.
- Proactive actions will prevent the loss of revenue for the company and will improve / retain the market share among the industry peers in terms of the number of active subscribers.

OBJECTIVE

- For many incumbent operators, *retaining high profitable customers is the number one business goal*.
- To 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
- The business objective is to predict the churn in the last (i.e. the ninth) month using the data (features) from the first three months.

DATA DESCRIPTION

- Here we are given with 4 months of data related to customer usage. In this case study, we 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. Churn is predicted using two approaches. Usage based churn and Revenue based churn. Usage based churn: Customers who have zero usage, either incoming or outgoing - in terms of calls, internet etc. over a period of time. This case study only considers usage based churn. In the Indian and the southeast Asian market, approximately 80% of revenue comes from the top 20% customers (called high-value customers). Thus, if we can reduce churn of the high-value customers, we will be able to reduce significant revenue leakage. Hence, this case study focuses on high value customers only. The dataset contains customer-level information for a span of four consecutive months - June, July, August and September. The months are encoded as 6, 7, 8 and 9, respectively.
- The business objective is to predict the churn in the last (i.e. the ninth) month using the data (features) from the first three months.

TELECOM CUSTOMER CHURN PREDICTION

ANALYSIS RESULTS

The telecom company has many users with negative average revenues in both phases. These users are likely to churn. Most customers prefer the plans of '0' category. The customers with lesser 'aon' are more likely to Churn when compared to the Customers with higer 'aon'. Revenue generated by the Customers who are about to churn is very unstable. The Customers whose arpu decreases in 7th month are more likely to churn when compared to ones with increase in arpu. The Customers with high total_og_mou in 6th month and lower total_og_mou in 7th month are more likely to churn compared to the rest. The Customers with decrease in rate of total_ic_mou in 7th month are more likely to churn, compared to the rest. Customers with stable usage of 2g volume throughout 6 and 7 months are less likely to churn. Customers with fall in usage of 2g volume in 7th month are more likely to Churn. Customers with stable usage of 3g volume throughout 6 and 7 months are less likely to churn. Customers with fall in consumption of 3g volume in 7th month are more likely to Churn. The customers with lower total_og_mou in 6th and 8th months are more likely to Churn compared to the ones with higher total_og_mou. The customers with lesser total_og_mou_8 and aon are more likely to churn compared to the one with higher total_og_mou_8 and aon. The customers with less total_ic_mou_8 are more likely to churn irrespective of aon. The customers with total_ic_mou_8 > 2000 are very less likely to churn.

The following are the strongest indicators of churn

Customers who churn show lower average monthly local incoming calls from fixed line in the action period by 1.27 standard deviations , compared to users who don't churn , when all other factors are held constant. This is the strongest indicator of churn. Customers who churn show lower number of recharges done in action period by 1.20 standard deviations, when all other factors are held constant. This is the second strongest indicator of churn. Further customers who churn have done 0.6 standard deviations higher recharge than non-churn customers. This factor when coupled with above factors is a good indicator of churn. Customers who churn are more likely to be users of 'monthly 2g package-0 / monthly 3g package-0' in action period (approximately 0.3 std deviations higher than other packages), when all other factors are held constant.

TELECOM CUSTOMER CHURN PREDICTION

MODELS SUMMARY

Model 1 : Logistic Regression with RFE & Manual Elimination (Interpretable Model)

Most important predictors of Churn , in order of importance and their coefficients are as follows :
loc_ic_t2f_mou_8 -1.2736 total_rech_num_8 -1.2033 total_rech_num_6 0.6053 monthly_3g_8_0 0.3994 monthly_2g_8_0 0.3666 std_ic_t2f_mou_8 -0.3363 std_og_t2f_mou_8 -0.2474 const -0.2336 monthly_3g_7_0 -0.2099 std_ic_t2f_mou_7 0.1532 sachet_2g_6_0 -0.1108 sachet_2g_7_0 -0.0987 sachet_2g_8_0 0.0488 sachet_3g_6_0 -0.0399
PCA: PCA : 95% of variance in the train set can be explained by first 16 principal components and 100% of variance is explained by the first 45 principal components.

Model 2 : PCA + Logistic Regression

Train Performance : Accuracy : 0.627 Sensitivity / True Positive Rate / Recall : 0.918 Specificity / True Negative Rate : 0.599 Precision / Positive Predictive Value : 0.179 F1-score : 0.3 Test Performance : Accuracy : 0.086 Sensitivity / True Positive Rate / Recall : 1.0 Specificity / True Negative Rate : 0.0 Precision / Positive Predictive Value : 0.086 F1-score : 0.158

Model 3 : PCA + Random Forest Classifier

Train Performance : Accuracy : 0.882 Sensitivity / True Positive Rate / Recall : 0.816 Specificity / True Negative Rate : 0.888 Precision / Positive Predictive Value : 0.408 F1-score : 0.544 Test Performance : Accuracy : 0.86 Sensitivity / True Positive Rate / Recall : 0.80 Specificity / True Negative Rate : 0.78 Precision / Positive Predictive Value :0.37 F1-score :0.51

Model 4 : PCA + XGBoost

Train Performance : Accuracy : 0.873 Sensitivity / True Positive Rate / Recall : 0.887 Specificity / True Negative Rate : 0.872 Precision / Positive Predictive Value : 0.396 F1-score : 0.548 Test Performance : Accuracy : 0.086 Sensitivity / True Positive Rate / Recall : 1.0 Specificity / True Negative Rate : 0.0 Precision / Positive Predictive Value : 0.086 F1-score : 0.158

TELECOM CUSTOMER CHURN PREDICTION

RECOMMENDATIONS

- Following are the strongest indicators of churn
- Customers who churn show lower average monthly local incoming calls from fixed line in the action period by 1.27 standard deviations , compared to users who don't churn , when all other factors are held constant. This is the strongest indicator of churn. Customers who churn show lower number of recharges done in action period by 1.20 standard deviations, when all other factors are held constant. This is the second strongest indicator of churn. Further customers who churn have done 0.6 standard deviations higher recharge than non-churn customers. This factor when coupled with above factors is a good indicator of churn. Customers who churn are more likely to be users of 'monthly 2g package-0 / monthly 3g package-0' in action period (approximately 0.3 std deviations higher than other packages), when all other factors are held constant.
- Based on the above indicators the recommendations to the telecom company are :
- Concentrate on users with 1.27 std devations lower than average incoming calls from fixed line. They are most likely to churn. Concentrate on users who recharge less number of times (less than 1.2 std deviations compared to avg) in the 8th month. They are second most likely to churn. Models with high sensitivity are the best for predicting churn. Use the PCA + Logistic Regression model to predict churn. It has an ROC score of 0.87, test sensitivity of 100%.

CONCLUSION

- Concentrate on users with 1.27 std devations lower than average incoming calls from fixed line. They are most likely to churn. Concentrate on users who recharge less number of times (less than 1.2 std deviations compared to avg) in the 8th month. They are second most likely to churn. Models with high sensitivity are the best for predicting churn. Use the PCA + Logistic Regression model to predict churn. It has an ROC score of 0.87, test sensitivity of 100%

Thank you