TELECOM CHURN CASE STUDY

Himanshu Shukla

AGENDA

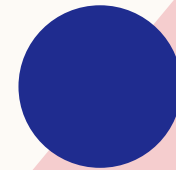Introduction

Primary goals

Understanding the business objectives and

the data

Factors affecting Churn

How we got there

Area of Focus

Summary

INTRODUCTION

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.

PRIMARY GOALS

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.

UNDERSTANDING THE BUSINESS OBJECTIVE AND THE DATA

- 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. To do this task well, understanding the typical customer behavior during churn will be helpful.

FACTORS AFFECTING CHURN

- Based on the Logistic Regression Classification Model following are the important variables affecting Churn:

1. onnet_mou_8 (Coefficient:0.2440)
2. roam_og_mou_8(Coefficient:1.4531)
3. std_og_t2m_mou_8 (Coefficient:0.3093)
4. isd_og_mou_8 (Coefficient:-0.5546)
5. loc_ic_t2f_mou_8 (Coefficient:-1.2894)
6. total_ic_mou_8 (Coefficient:-1.7335)
7. ic_others_8 (Coefficient:-1.6243)
8. monthly_2g_7 (Coefficient:-0.5587)
9. monthly_2g_8 (Coefficient:-0.6329)
10. monthly_3g_8 (Coefficient:-0.8767)
11. decrease_arpu_action (Coefficient:-0.5447)
12. decrease_vbc_3g_action (Coefficient:-1.2592)
13. arpu_8 (Coefficient:-1.3413)

HOW WE GOT THERE
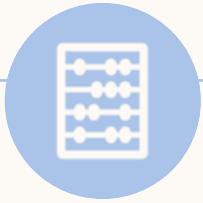
UNDERSTANDING DATA

- Load the Dataset
- Removed the Missing Values from Dataset
- Filtered the High Value Customer based on the recharge amount during Action Phase(Month of June and July)
- Tagging of Churn done by using the Minute of use variable and Internet use
- Outliers Removed from the Dataset for Numeric Variable

MODEL BUILDING

- Feature Engineering done for deriving New Variables for Analysis
- Data is divided into Train and Test Data Set
- Data Set is imbalanced do SMOTE technique is used for removing imbalance
- Logistic Regression Model building done using Recursive Feature Selection
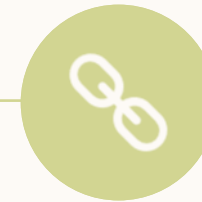
MODEL EVALUATION

- Final Model after low p-value and low VIF is tested on the Train Data Set.
- Accuracy and Recall Values obtained were good for the Test Dataset.
- Model was tested on the Test Dataset. Obtained Accuracy and Recall are good. So Model is performing well on the Test Dataset as well.

AREAS OF FOCUS

RECOMMENDATIONS

- Target the Customer whose total incoming call minutes and outgoing ISD call minutes are decreasing.

- Customers whose roaming outside call are increasing more likely to Churn.

- Customer for whom average revenue per user is decreased in the month of August are more likely to churn.

- Customers with decreasing monthly 2g and 3g recharge are more likely to Churn.

- 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.

- Customer having decrease incoming call minute from T to F Lines are more likey to Churn.

SUMMARY

We Needed Interpretability so we choose the Logistic Regression Model. Logistic Regression Model Provided the Variables with the Coefficients which are affecting the Churn Probability.

Accuracy and Recall Values are good for the Train and Test Data Set. So Model Performing well on Test and Train Dataset. Factors affecting the Churn are identified.

THANK YOU