TELECOM CHURN CASE STUDY

ALLAMPRABHU HIREMATH

ITIKA KHANDELWAL

STANY SUMANTH

TABLE OF CONTENTS

1. Problem Statement

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.

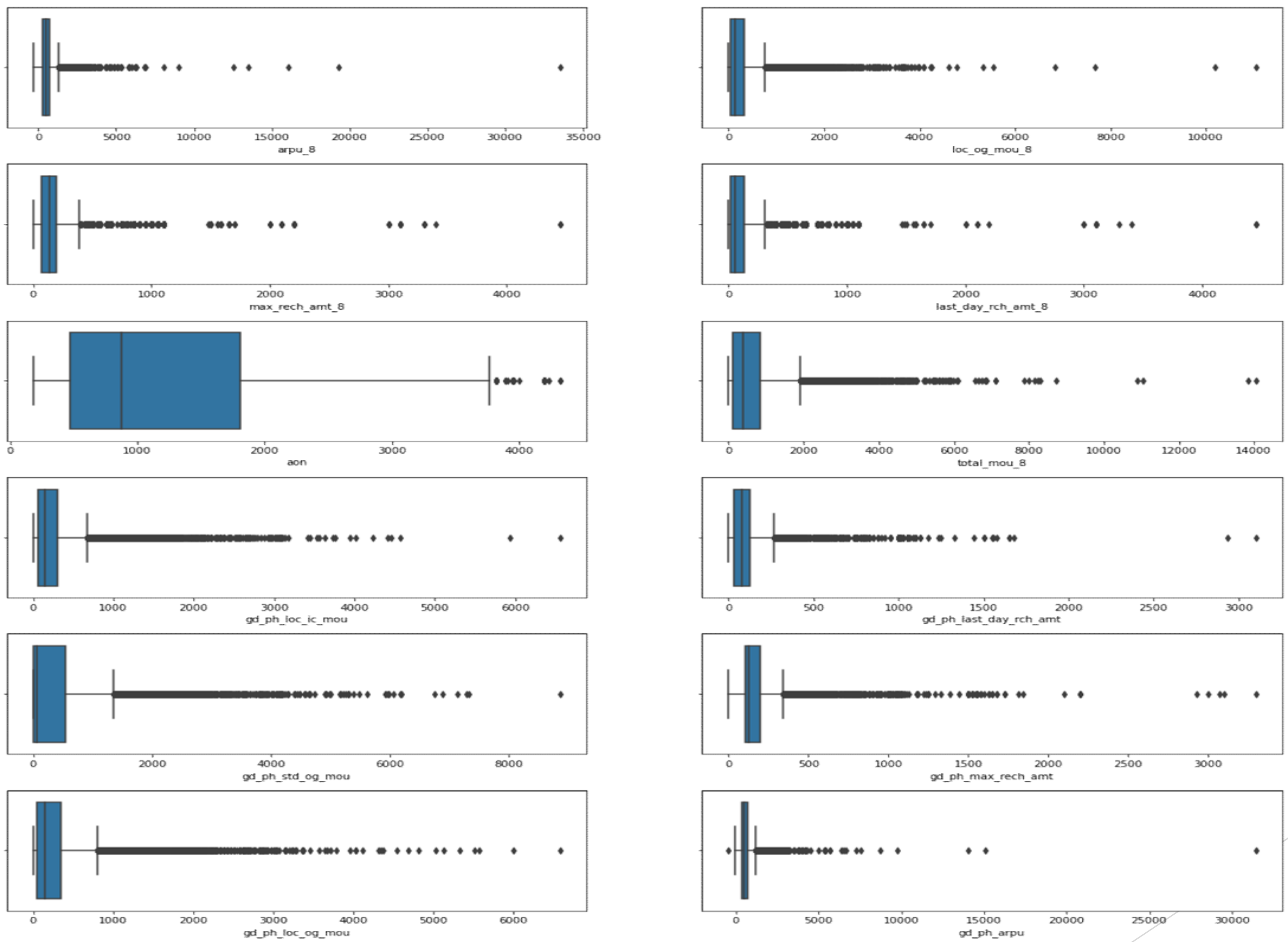
2 . Analysis Approach

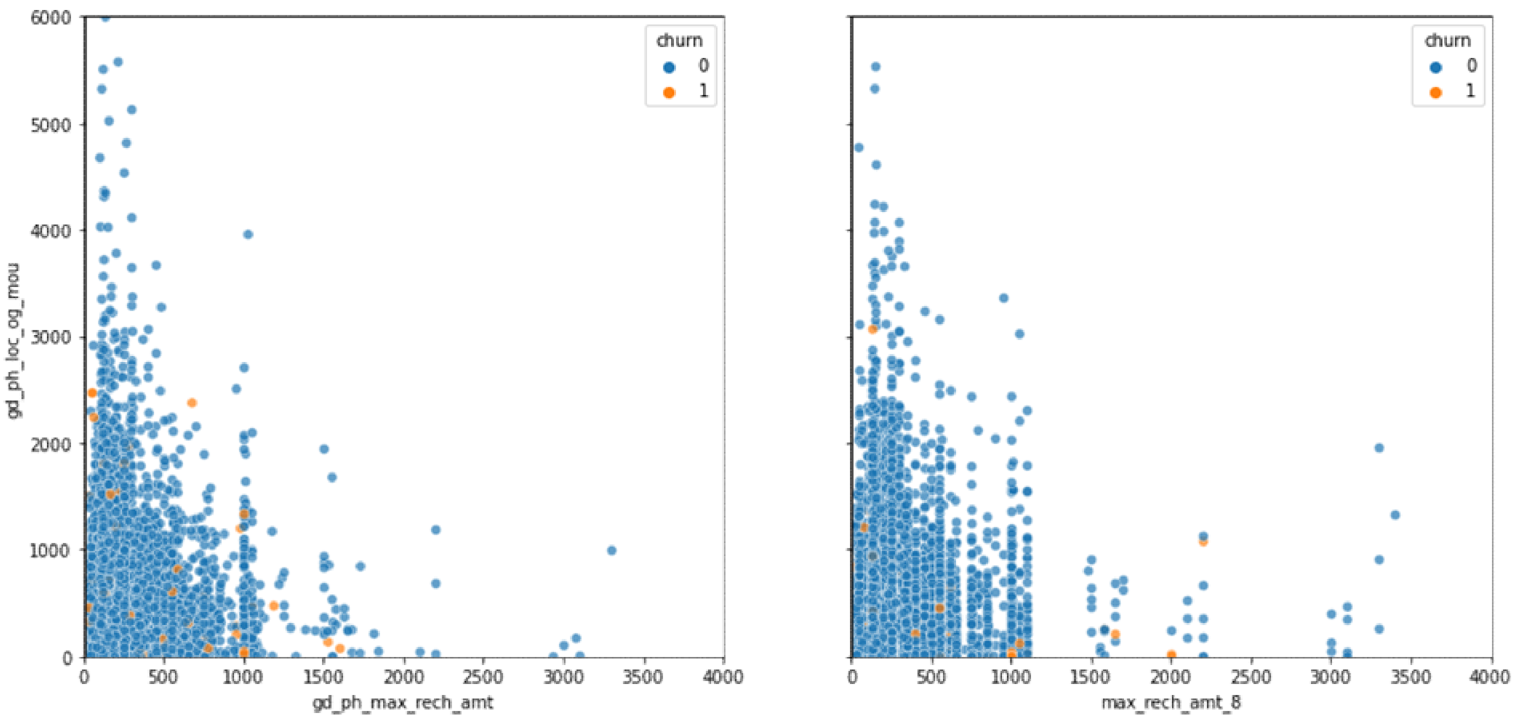
- Here we are given with 4 months of data related to customer usage. In this case study, we analyse customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn.
- Churn is predicted using two approaches. 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.
- This is a classification problem, where we need to predict whether the customers is about to churn or not.
- We have carried out Baseline Logistic Regression, then Logistic Regression with PCA, PCA + Random Forest, PCA + XGBoost.

3 Analysis Steps

- *Reading, understanding cleaning and visualising the data*
  - In this step we loaded all the necessary libraries and loaded the data set in data frame.
  - After this we identified shape of the data, number of rows columns, null values and unique values as well.
- *Preparing the data for modelling*
  - *In this step we do EDA. Carrying out steps like Univariate analysis and get the Churn rate on the*
- *basis whether the customer decreased her/his MOU in action month.*
- *Building the model*
- *Evaluate the model*

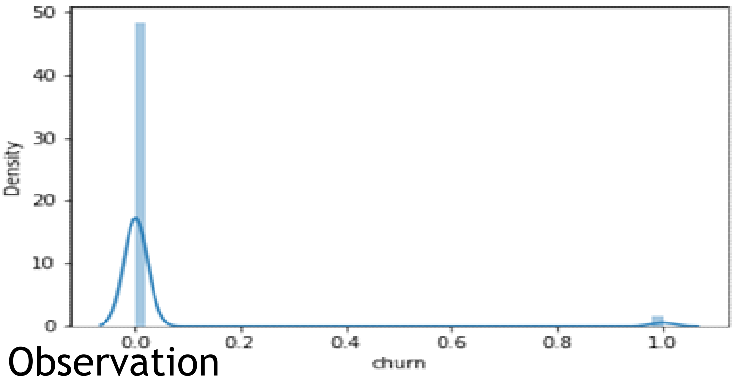
3.1. RESULTS OF PLOTS

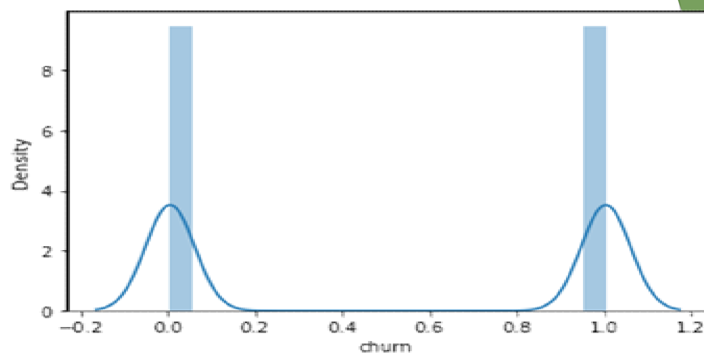




Observations

- Users who were recharging with high amounts were using the service for local uses less as compared to user who did lesser amounts of recharge
- Intuitevly people whose max recharge amount as well as local out going were very less even in the good phase churned more

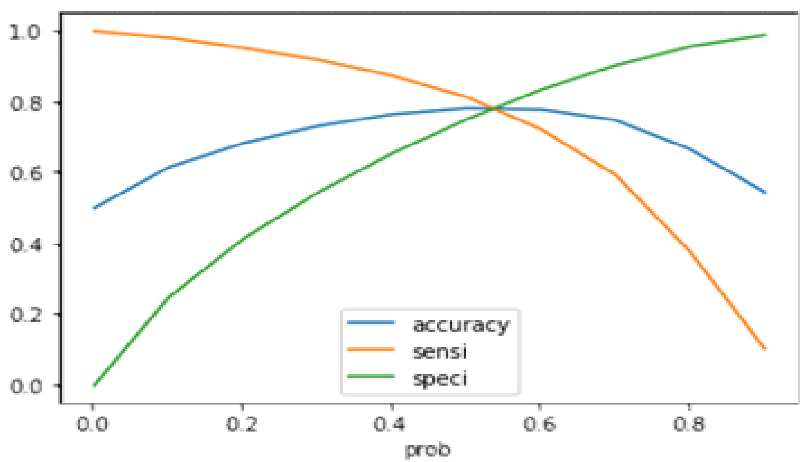




Observation

- The variable is not skewed it is highly imbalanced, the number of non-churners in the dataset is around 94%
- The Imbalance is handled using smote algorithm.

The class is balanced and the target variable is not skewed



The Optimal point is 0.5 for the model

4 Preprocessing

- Train-Test Split has been performed.
- The data has high class-imbalance with the ratio of 0.098 (class 1 : class 0).
- SMOTE technique has been used to overcome class-imbalance.
- Predictor columns have been standardized to mean - 0 and standard_deviation- 1.

5 Modelling

- Model summary
- Train set
  - Accuracy = 0.88
  - Sensitivity = 0.81
- **Test Set**
  - Accuracy = 0.90
  - Test Error = 0.68
- Overall, the model is performing well in the test set, what we learnt from the train set.
- Users maximum recharge amount is less than 200 even in the good pahse, should have a tag and re-evaluated time to time as they are more likely to churn
- Users that have been with the network less than 4 years, should be monitored time to time, as from data we can see that users who have been associated with the network for less than 4 years tend to churn more
- MOU is one of the major factors, but data especially VBC if the user is not using a data pack if another factor to look out