

Problem Statement ?



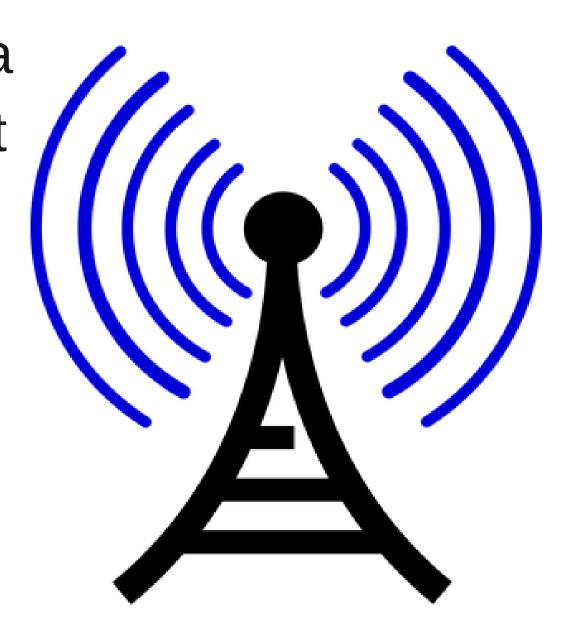
In the Indian and the southeast Asian market, approximately 80% of revenue comes from the top 20% customers (called high-value customers). Thus, if we can reduce churn of the high-value customers, we will be able to reduce significant revenue leakage.

Predict Churn In High-Valued Customers

About Dataset

The dataset contains customer-level information for a span of four consecutive months - June, July, August and September.

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.



Identify the High-Valued Customers

High-valued customers are those who have recharged more than or equal to 70th percentile of the average recharge amount in the first two months (i.e. June, July)



Data Preparation

1. Derive new features to filter high-value customers

New Feature= Data greater than or equal to 70th percentile of the average recharge amount in June and July.

2. Tag churners and remove attributes of the churn phase.

Identifying churned high value customers based on the last month(September)

1=Churned Customers
0=Non-Churned Customers

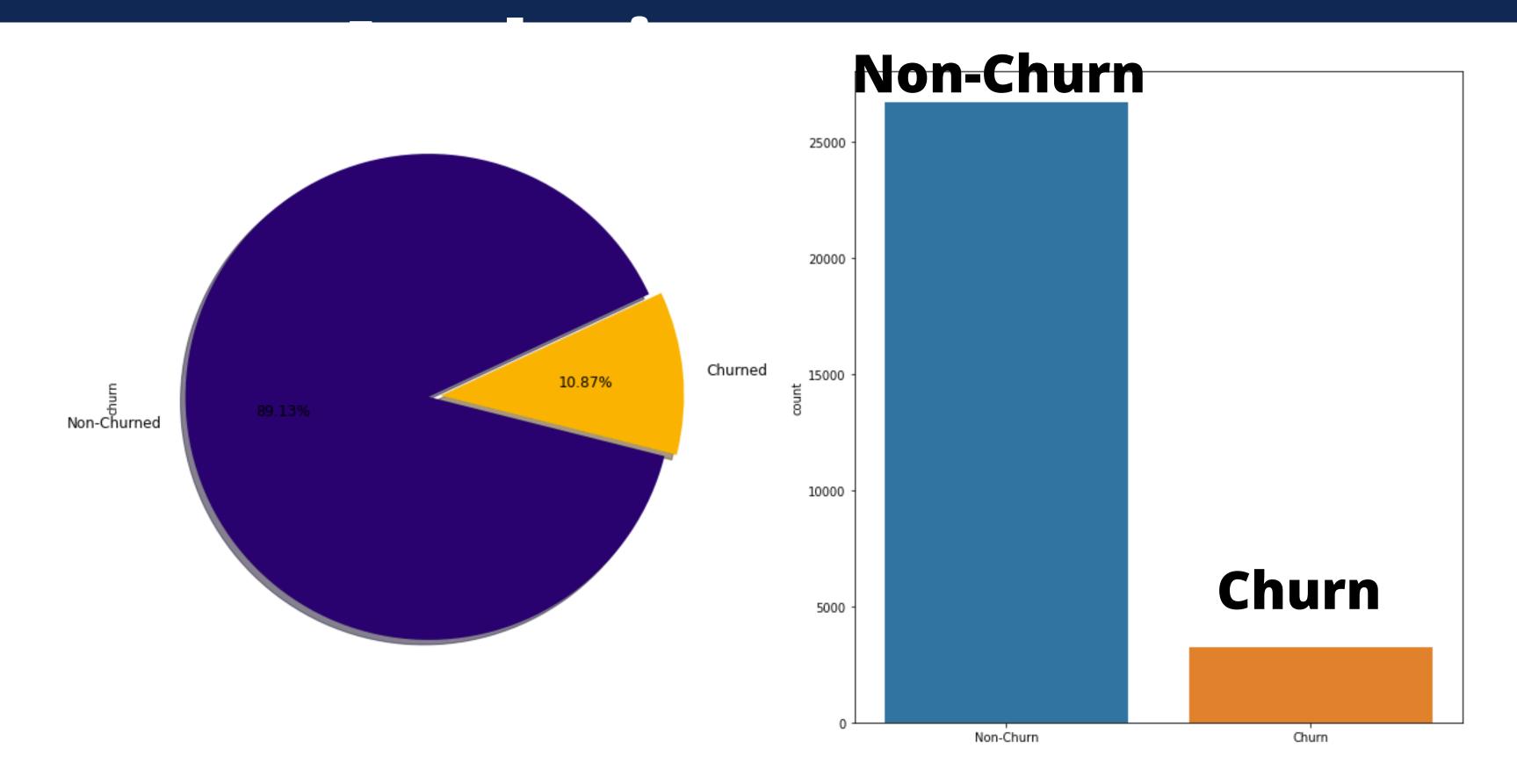


Model Building

- Preprocess data
- Exploratory Data Analysis
- Derive new features.
- Reduce the number of variables using PCA
- Handle class imbalance
- Train different models and perform appropriate hyperparameters
- Evaluate the models using appropriate evaluation metrics
- Choose a model based on some evaluation metric



Further Exploratory Data



Exploratory Data Analysis: Correlation



Deriving New Features

Out going call (June) \longrightarrow Out going call (July) 1/2Average Out going call (June and July)



3G Network (July) 72 == 3G Network (July) 72



Average 3G Network (June and July)

Building The Model

Reducing Dimensions

PCA

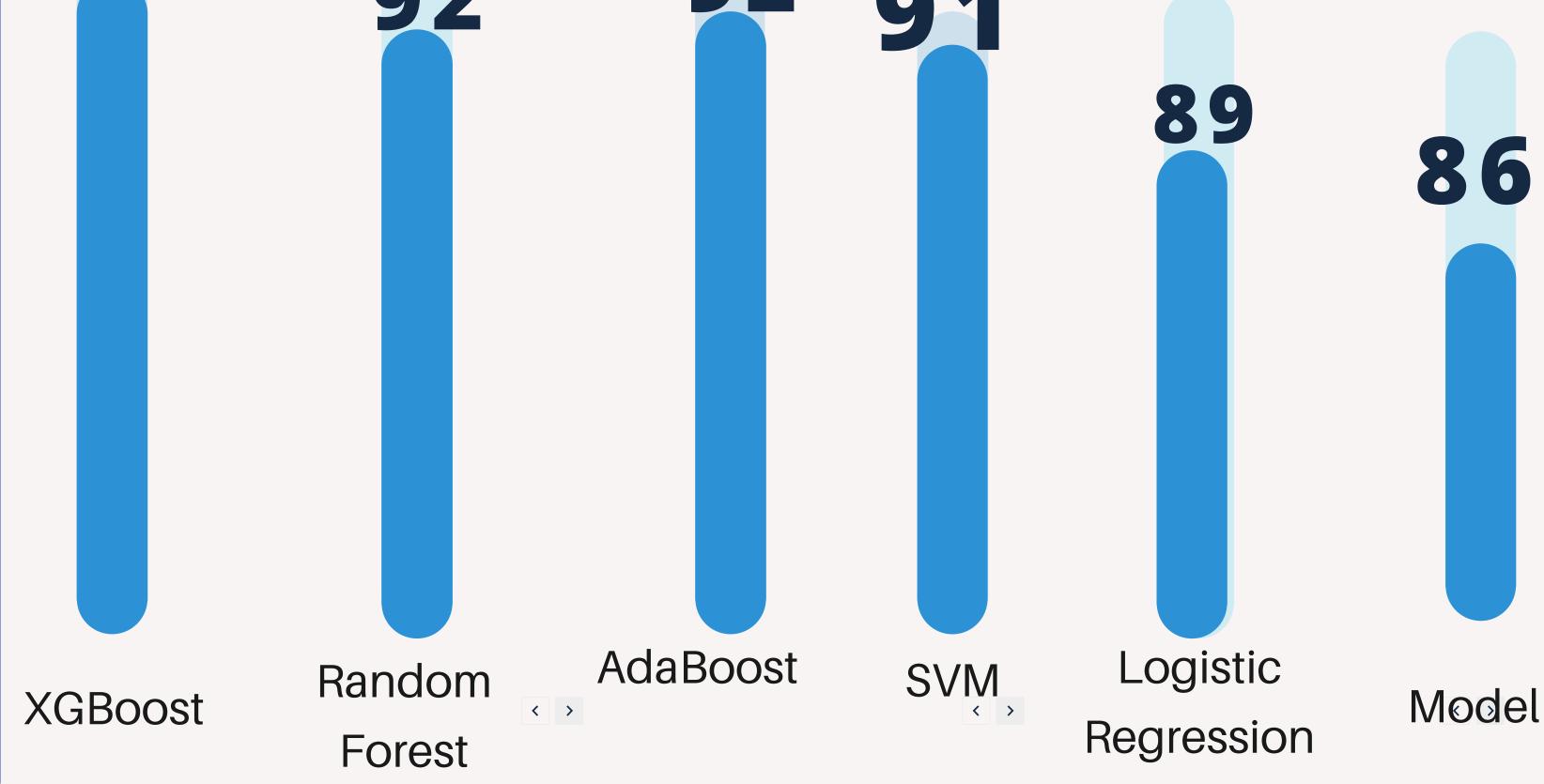
(29953, 204)

(5991, 54)

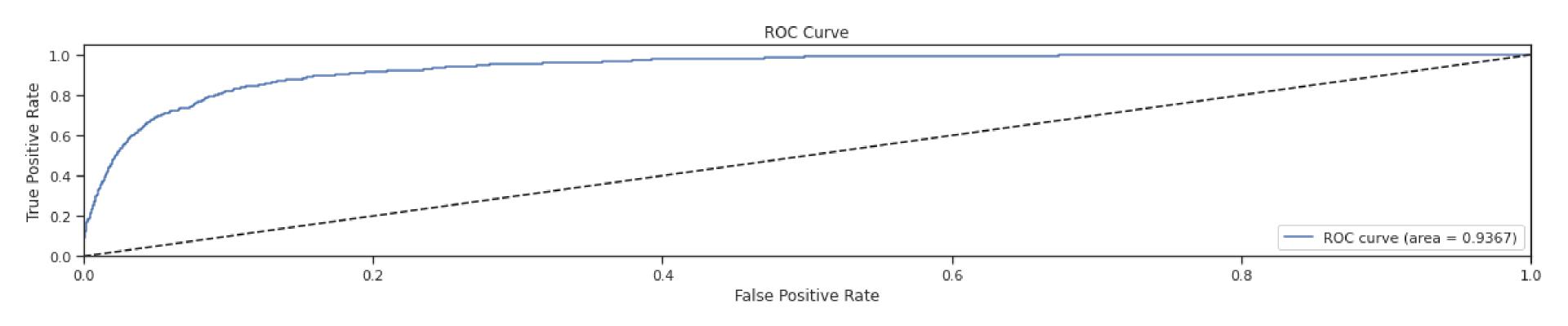
Imbalance Data

Upsampling Downsampling

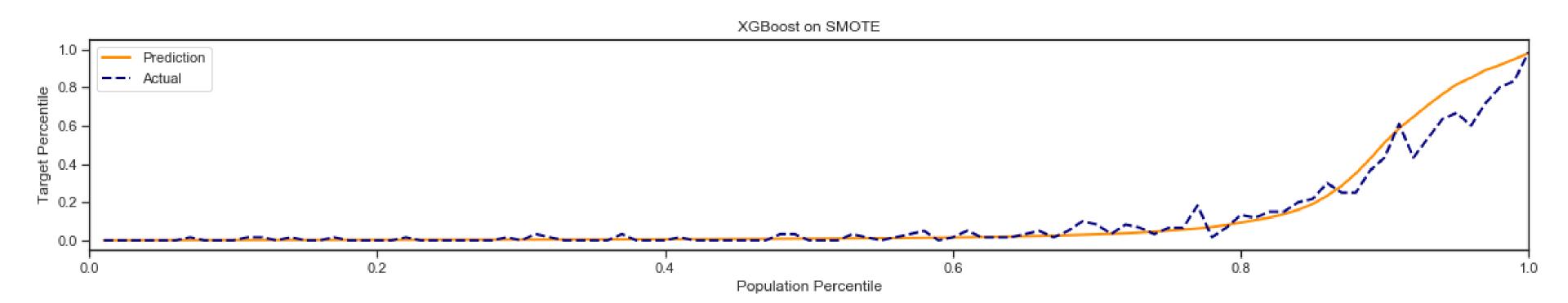
MODELS



Evaluation Metrics



Evaluation Metrics



468.8624519406 mportant Features 213.48926149800002 94.2970559418868 78.67185342563157 76.60803496049996 75.99155740658698 68.41892932363636 57.990104776666676 53.832782420000015 60.02663866857333 84.2944951275 48.540998567317075 54.48574273038462 50.52911019583331 49.10299819045456 39.754139852413786 39.589021249999995 38.092562814285714 43.20481508461538 43.24628733333333 42.45162526411765 34.301503710384615 31.795556186666662 30.967508063636373 37.9545898444444 30.580659556818183 37.69199120833333 29.663630474999998 29.299767837500003 29.0323496000000003 33.87243361714286 34.41028180599999 33.62470063195122 27.065308652499997 33.54491104166667 26.730429365000003 .506782037741935 23.748748689090906 24.15502070833334 52.2441139 50.0271606 38.6190453 28.04109195 31.6959305 29.7415743 28.7898534 28.3071556 23.9896876 24.0976105 aug_vbc asn_d ф bc_ic_mou_8c std_og_t2t_mou_8c max_rech_dat offinet_mol bc_ic_t2t_mor bc_ic_t2m_mol max_rech_arr roam_ic_mo std_og_mou_8c btal_rech_amt_8c date_of_last_rech_8r roam_og_mou_8c roam_og_mo bc_og_t2m_mo last_day_rch_arr roam_ic_mou_8c wol_3g_m loc_og_t2t_mo s_loc_og_t2m_mou_ late_of_last_rech_da loc_og_mo B_loc_ic_mou_ is_loc_ic_t2f_mou_ is_loc_ic_t2t_mou_ is_std_ic_mou_ is_loc_og_t2t_mou_ s_roam_og_mou_ is_std_og_mou_ is_roam_ic_mou_ AVG_std_og_mou_ s_loc_ic_t2m_mou_ s loc og t2f mou std_og_t2m_mou_ s_loc_og_mou_ s std ic t2m mou s_vol_3g_mb_ s_total_ic_mou_ AVG_offnet_mou_ is_offinet_mou_ is_vol_2g_mb_ is_total_rech_amt _nom_go_lds_a is ic others std_ic_t2t_mou_ is_onnet_mou_ is_total_og_mou

500

400

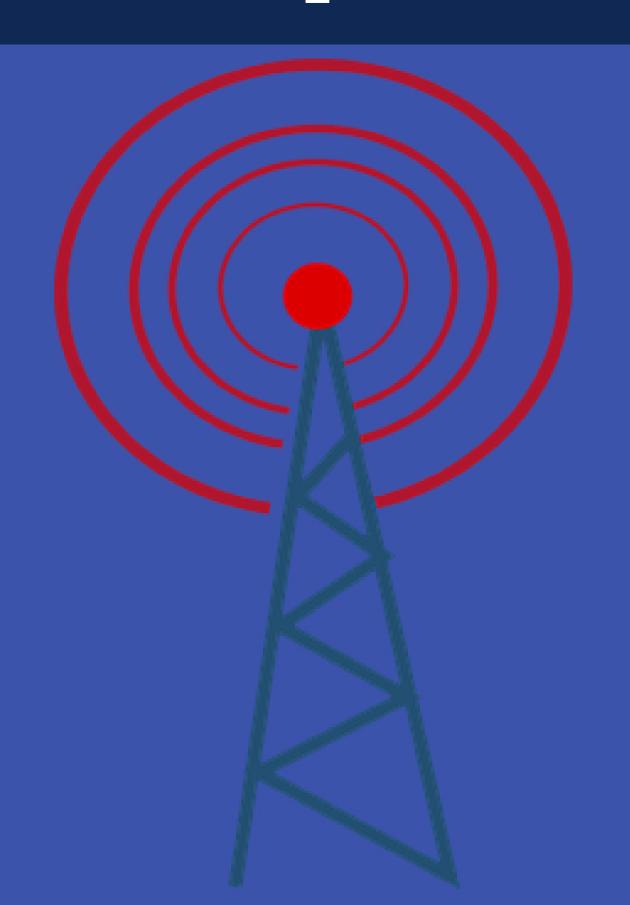
300

200

100

score

Sample Important Features



Data of last recharge in the 8th month

The last day recharge amount in 8th month

Total In-coming and out-going calls

Last day recharge amount in the 6th month

Average Revenue Per User

Recommendations to Business

We can infer from our analysis that there are certain features which are high indicaters of whether a customer will churn or not. In order for Telecom to reduce the number of churn and maintain these high-valued customers, it has to take note of these features and monitor it quite consistently in order to identify the stage those customers that are likely to churn and address their pain-point as soon as possible to prevent them from churning.

Another key factor to be consider is that if the recharge amount of the customer starts to reduce at the month where the customer has shown consistenet activeness, then it's a sign of churn and must be address as soon as possible.

Age on network is also a key indicator for identifying the churn, if age is less than 500 days and their usage is reduce then the customer is likely to churned. The other key indicater that the company should consider is data usage & amount.



Thank You