TELECOM CASE STUDY

BY - Geeta

Varun

Saketha

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.

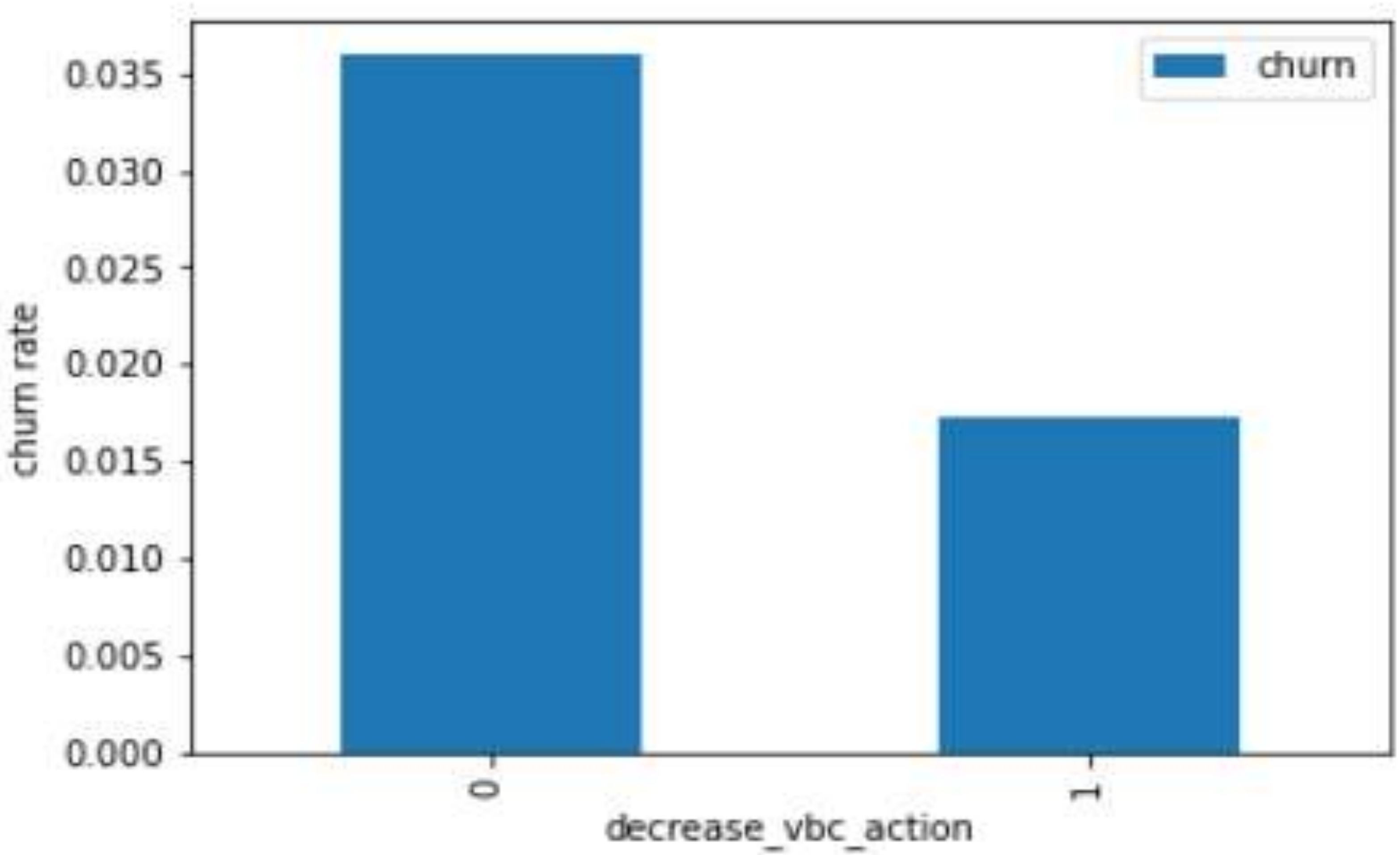
For many incumbent operators, retaining high profitable customers is the number one business goal.

To reduce customer churn, telecom companies need to predict which customers are at high risk of churn.

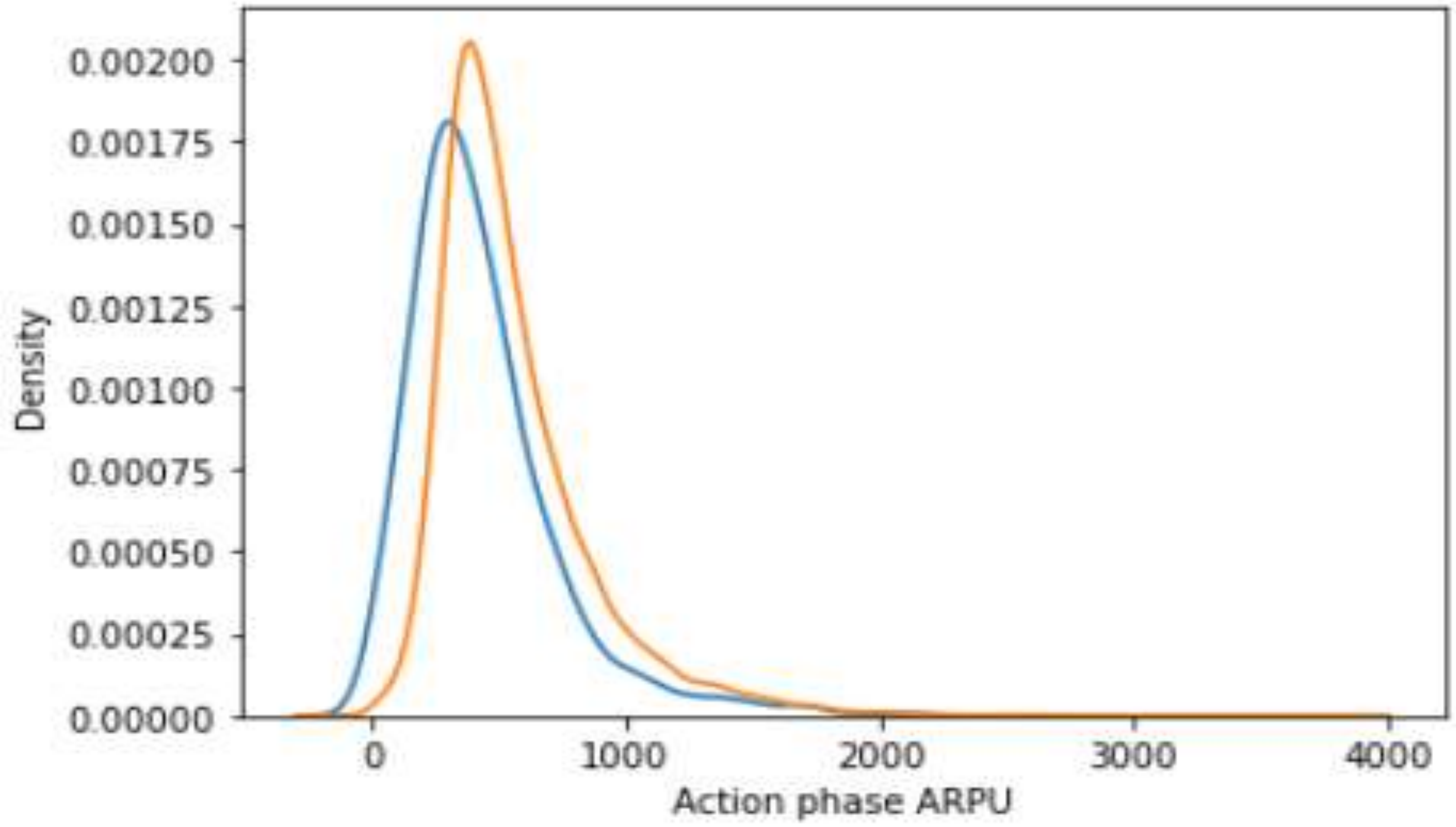
In this project, you will analyze customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn.

STEPS FOLLOWED :

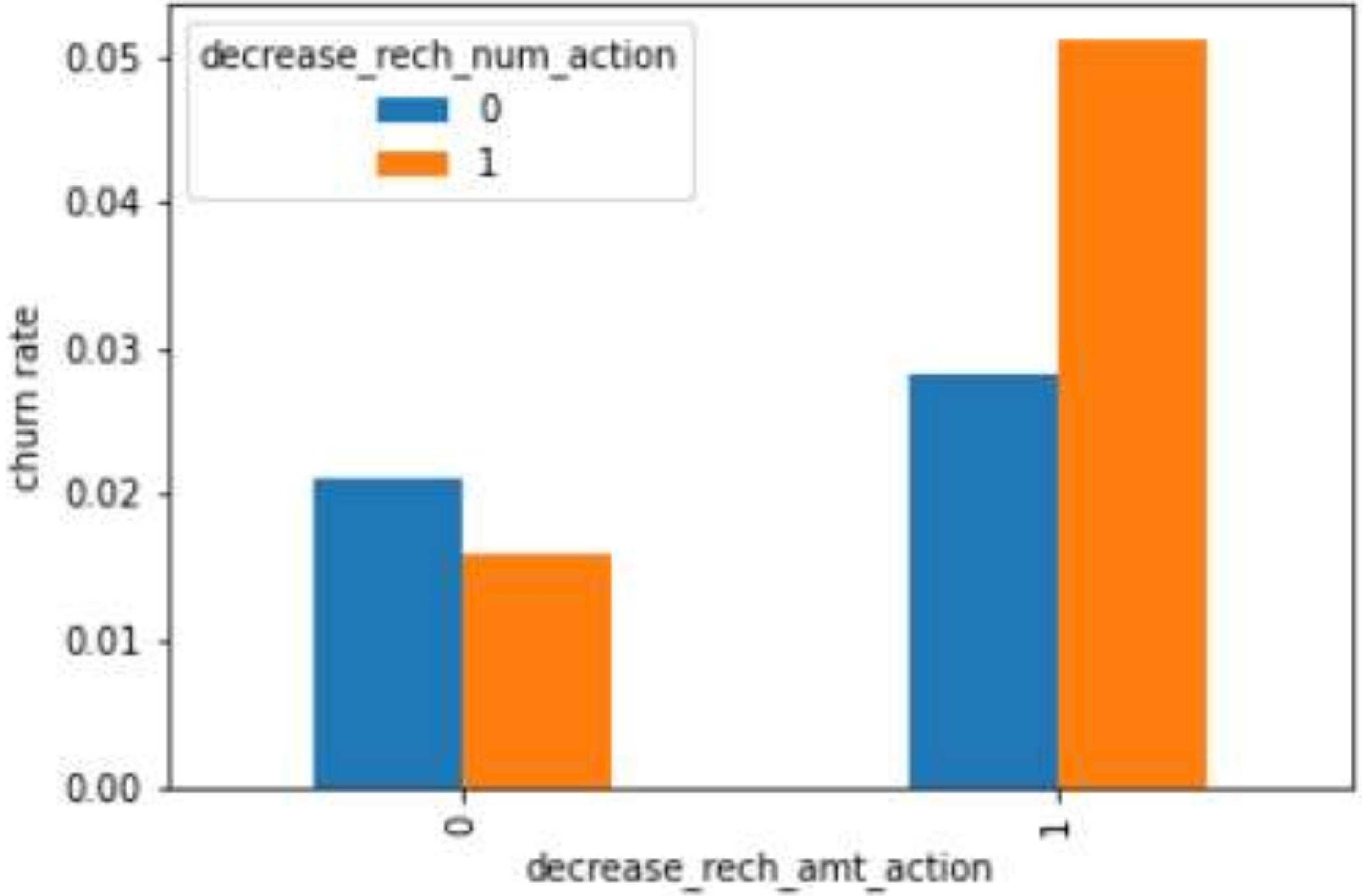
- ❖ Understand the domain/variables
- ❖ Import/load the data
- ❖ Clean and Prepare the data
- ❖ Exploratory Data Analysis
- ❖ Dealing with data imbalance
- ❖ Splitting the data into Train and Test sets
- ❖ Feature Scaling
- ❖ Using PCA technique for feature selection
- ❖ Building Multiple Models
- ❖ Model Evaluation



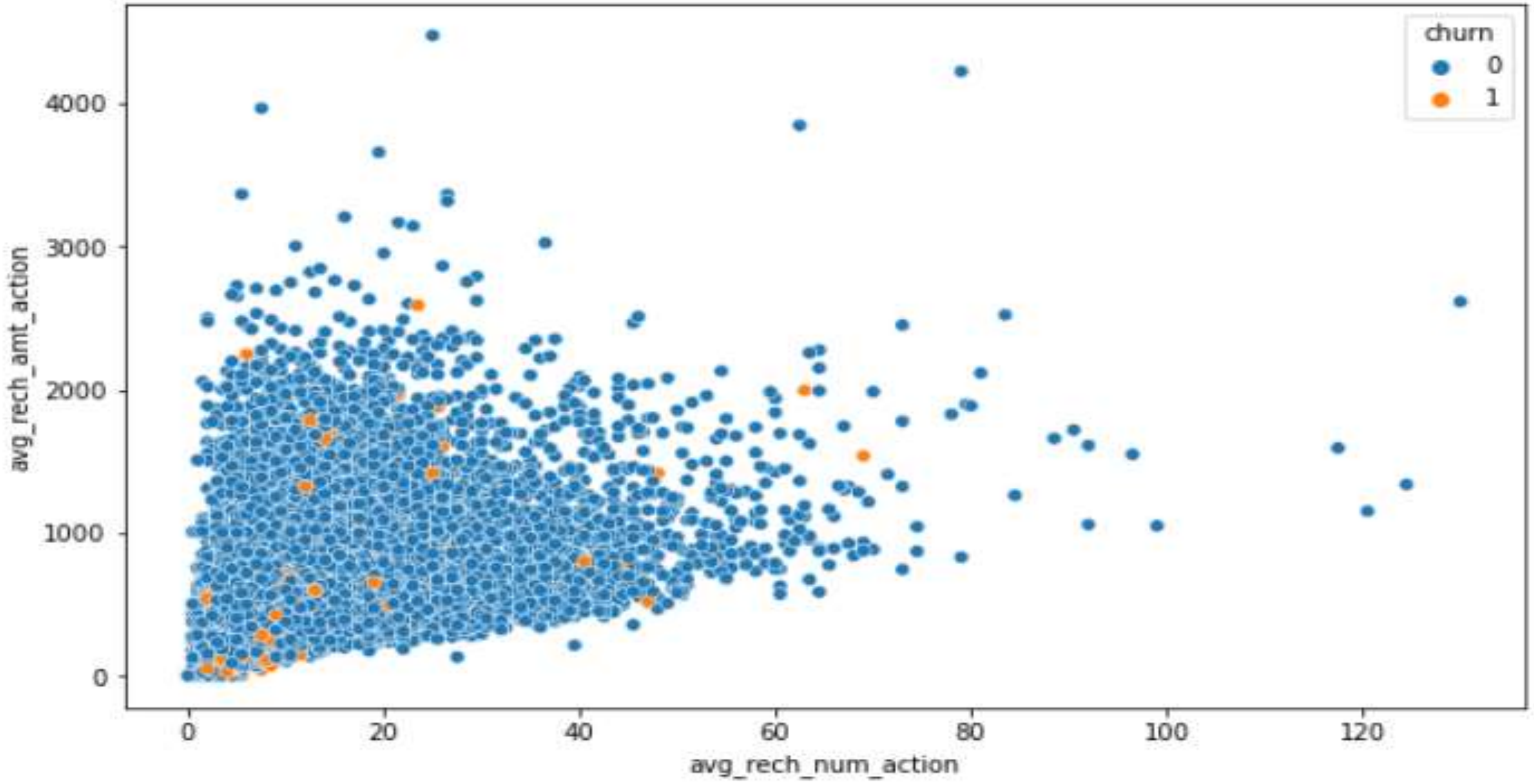
❖ Here we see the expected result. The churn rate is more for the customers, whose volume based cost in action month is increased. That means the customers do not do the monthly recharge more when they are in the action phase.



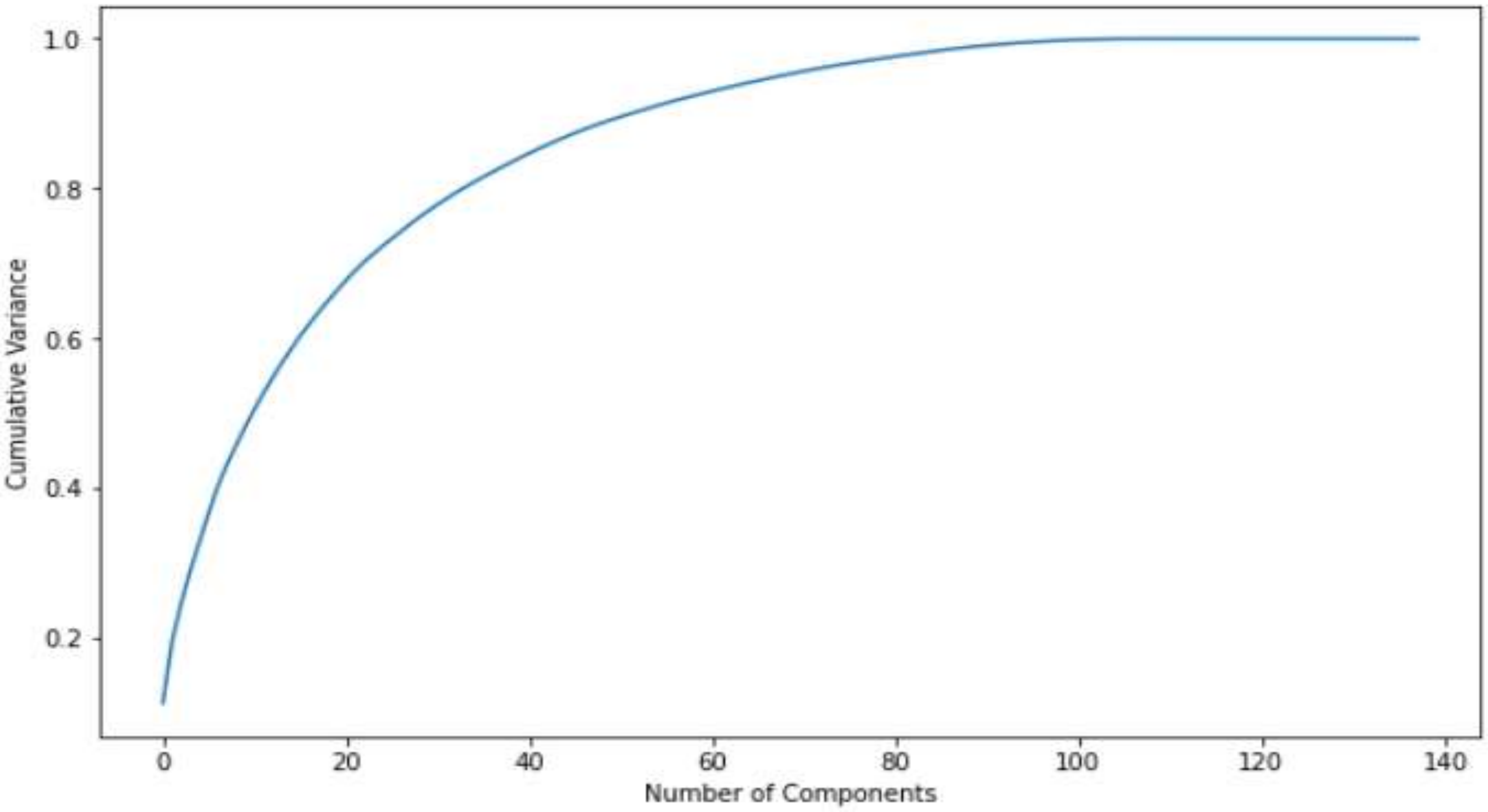
- ❖ Average revenue per user (ARPU) for the churned customers is mostly dense on the 0 to 900. The higher ARPU customers are less likely to be churned.
- ❖ ARPU for the not churned customers is mostly dense on the 0 to 1000.



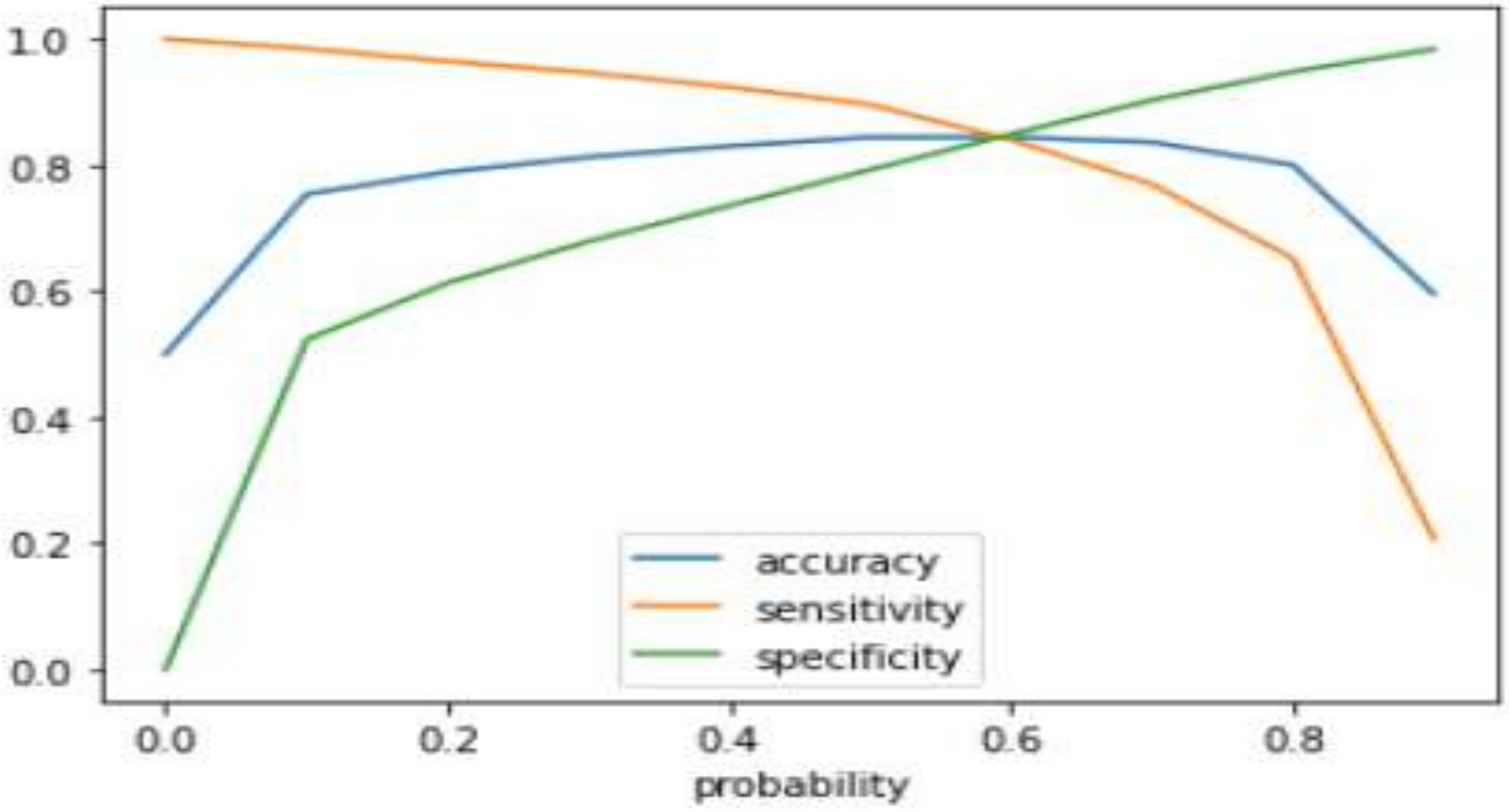
❖ We can see from the above plot, that the churn rate is more for the customers, whose recharge amount as well as number of recharge have decreased in the action phase than the good phase.



❖ We can see from the above pattern that the recharge number and the recharge amount are mostly proportional. More the number of recharge, more the amount of the recharge.

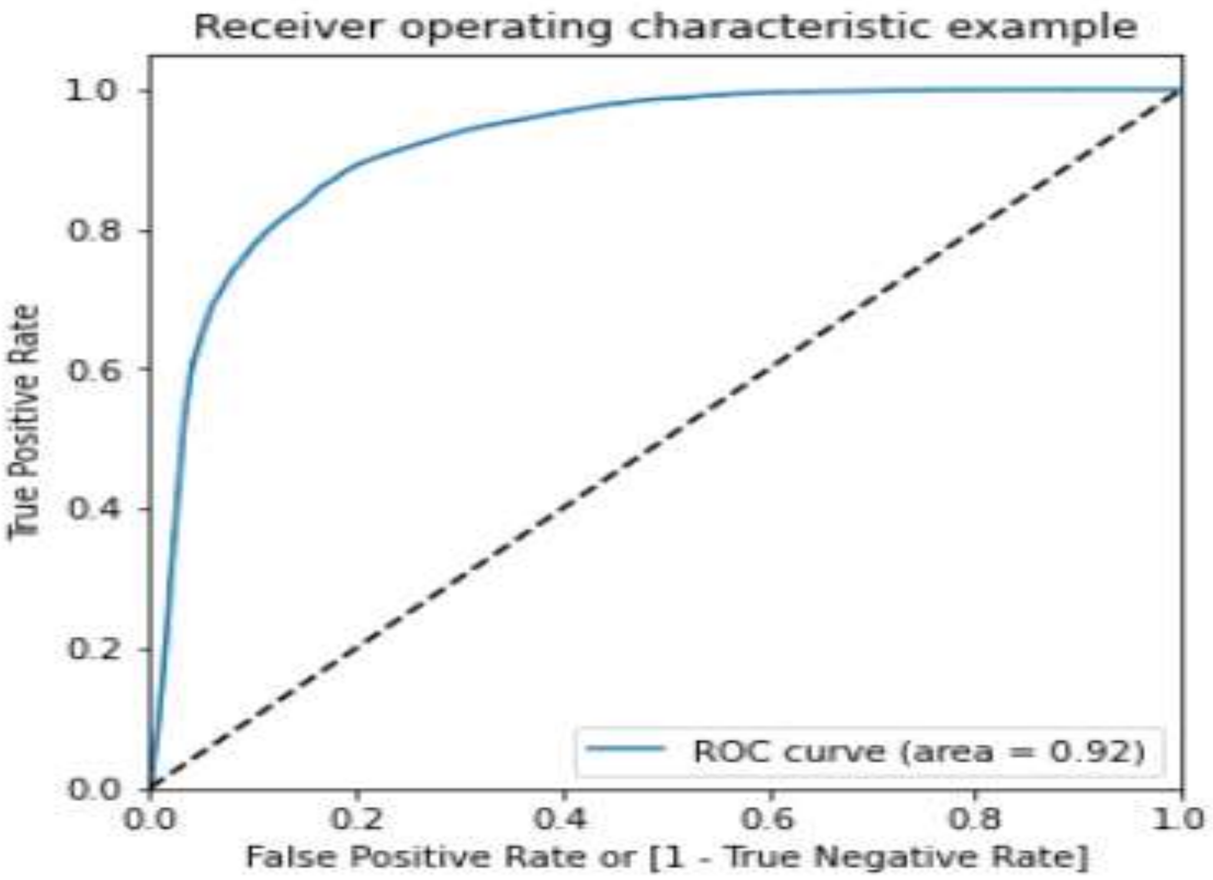


❖ We can see that 60 components explain almost more than 90% variance of the data. So, we will perform PCA with 60 components.



- At point 0.6 where the three parameters cut each other, we can see that there is a balance between sensitivity and specificity with a good accuracy.

- Here we are intended to achieve better sensitivity than accuracy and specificity. Though as per the above curve, we should take 0.6 as the optimum probability cut-off, we are taking *0.5* for achieving higher sensitivity, which is our main goal.



❖ We can see the area of the ROC curve is closer to 1, which is the Gini of the model.

TOP VARIABLES SELECTED IN LOGISTIC REGRESSION :

Variables	Coefficients
loc_ic_mou_8	-3.3287
og_others_7	-2.4711
ic_others_8	-1.5131
isd_og_mou_8	-1.3811
decrease_vbc_action	-1.3293
monthly_3g_8	-1.0943
std_ic_t2f_mou_8	-0.9503
monthly_2g_8	-0.9279
loc_ic_t2f_mou_8	-0.7102
roam_og_mou_8	0.7135

- We can see most of the top variables have negative coefficients. That means, the variables are inversely correlated with the churn probability.

E.g.:-

If the local incoming minutes of usage (loc_ic_mou_8) is lesser in the month of August than any other month, then there is a higher chance that the customer is likely to churn.

BUSINESS RECOMMENDATION

- Target the customers, whose minutes of usage of the incoming local calls and outgoing ISD calls are less in the action phase (mostly in the month of August).
- Target the customers, whose outgoing others charge in July and incoming others on August are less.
- Also, 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.
- Customers, whose monthly 3G recharge in August is more, are likely to be churned.
- Customers having decreasing STD incoming minutes of usage for operators T to fixed lines of T for the month of August are more likely to churn.
- Customers decreasing monthly 2g usage for August are most probable to churn.
- Customers having decreasing incoming minutes of usage for operators T to fixed lines of T for August are more likely to churn.
- roam_og_mou_8 variables have positive coefficients (0.7135). That means for the customers, whose roaming outgoing minutes of usage is increasing are more likely to churn.