Telecom Churn Case Study

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

Understanding and defining churn

There are two main models of payment in the telecom industry - postpaid (customers pay a monthly/annual bill after using the services) and prepaid (customers pay/recharge with a certain amount in advance and then use the services).

In the postpaid model, when customers want to switch to another operator, they usually inform the existing operator to terminate the services, and you directly know that this is an instance of churn.

However, in the prepaid model, customers who want to switch to another network can simply stop using the services without any notice, and it is hard to know whether someone has actually churned or is simply not using the services temporarily (e.g. someone may be on a trip abroad for a month or two and then intend to resume using the services again).

Thus, churn prediction is usually more critical (and non-trivial) for prepaid customers, and the term 'churn' should be defined carefully. Also, prepaid is the most common model in India and Southeast Asia, while postpaid is more common in Europe in North America.

This project is based on the Indian and Southeast Asian market.

3

SOLUTION

METHODOLOGY

1. Data Cleaning and Manipulation

- Handling missing values in rows
- Deleting the date columns as the date columns are not required in our analysis
- Filter high-value customers
- Tagging churners
- Deleting all the attributes corresponding to the churn phase
- Outliers' treatment
- Derive new features

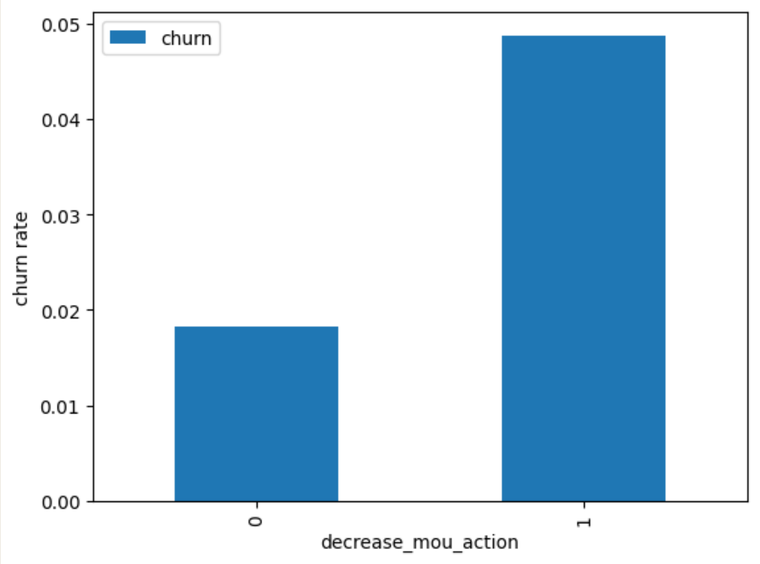
2. EDA

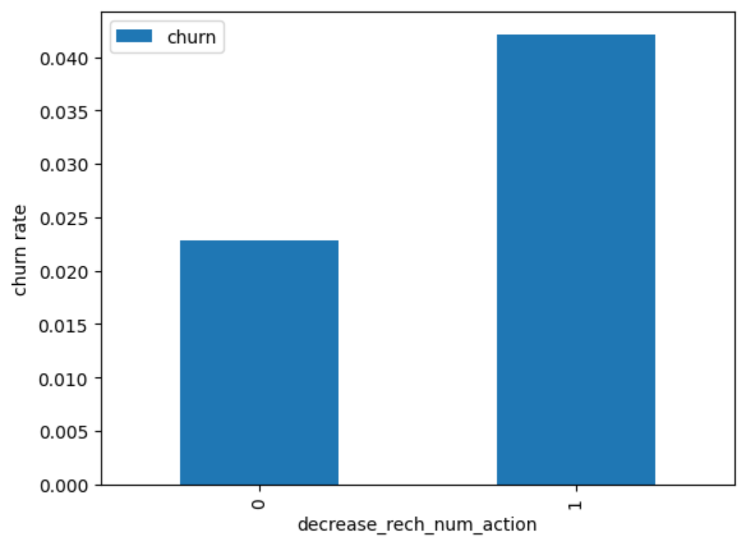
- Univariate Analysis: Conducted assessments of distributions and value counts to understand variable behavior.
- Bivariate Analysis: Evaluated correlations and patterns between variables to identify key relation

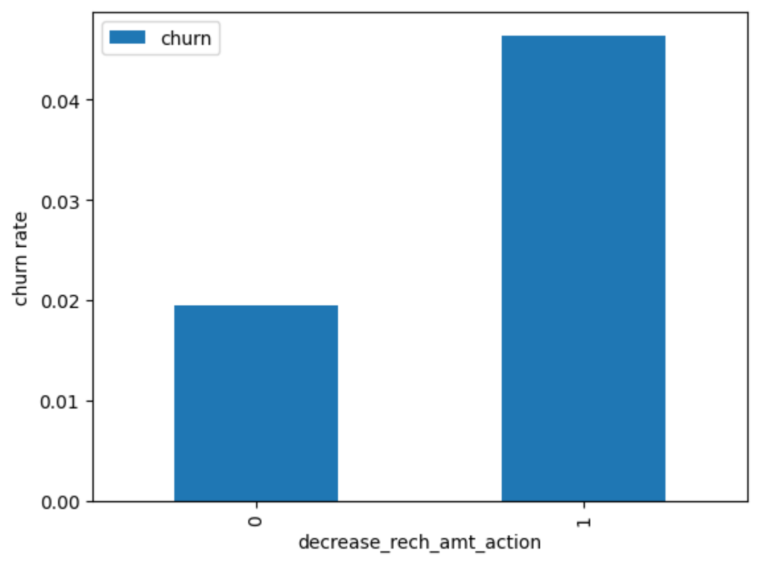
3. Model Development

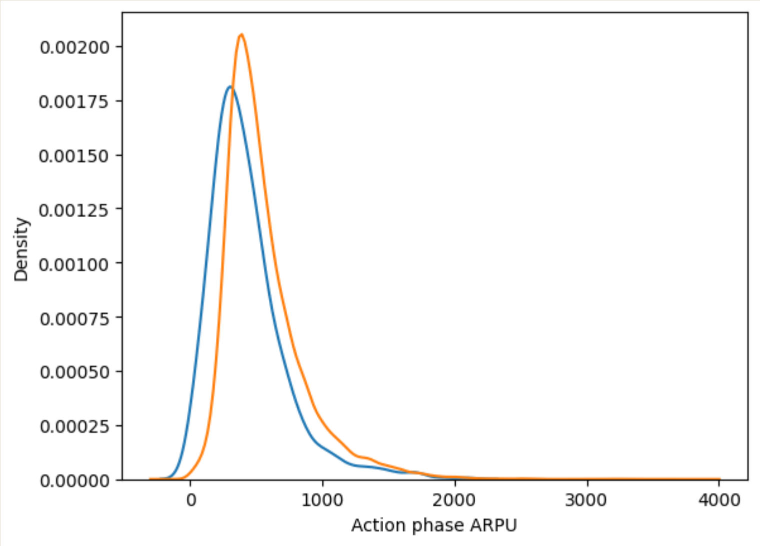
- Train-Test Split: Divided the dataset into training and testing sets in 80:20 ratio.
- Model with PCA : Logistic regression with PCA || Support Vector Machine(SVM) with PCA || Decision tree with PCA || Random forest with PCA
- Model Without PCA: Logistic regression with No PCA

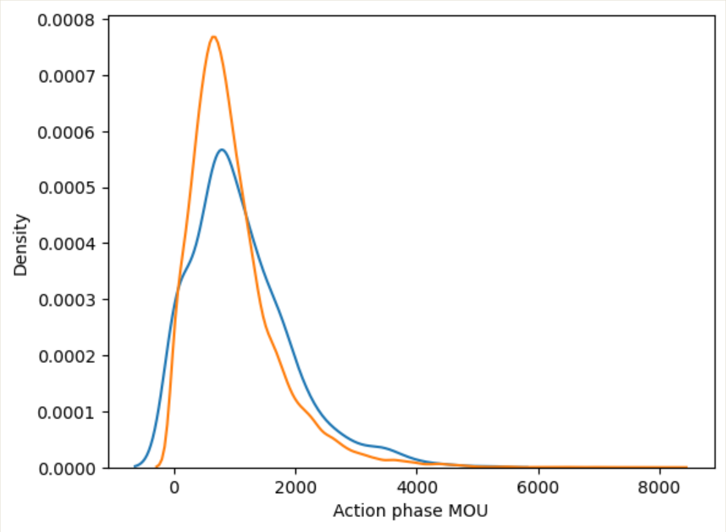
EDA:Univariate analysis

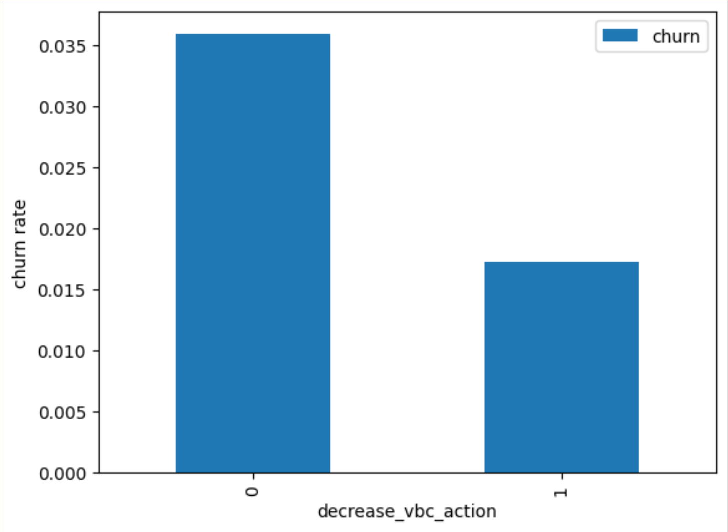




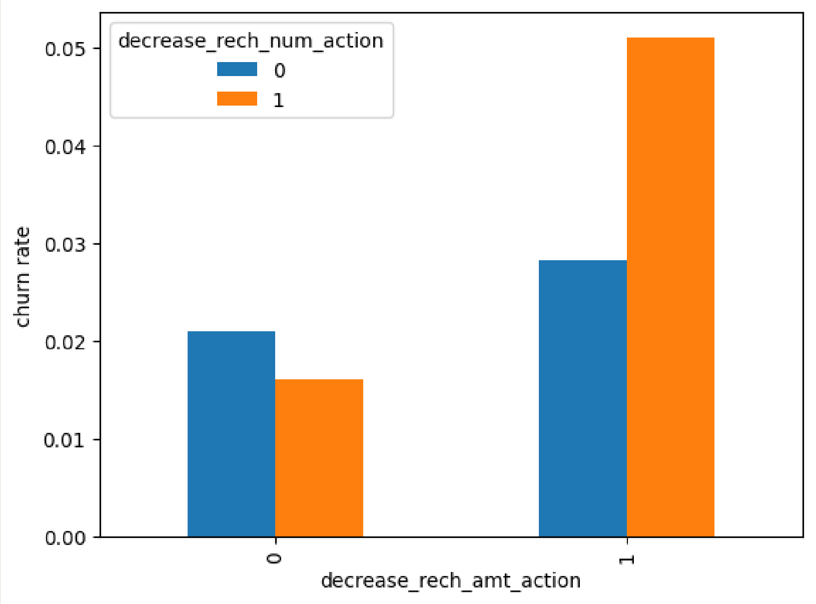


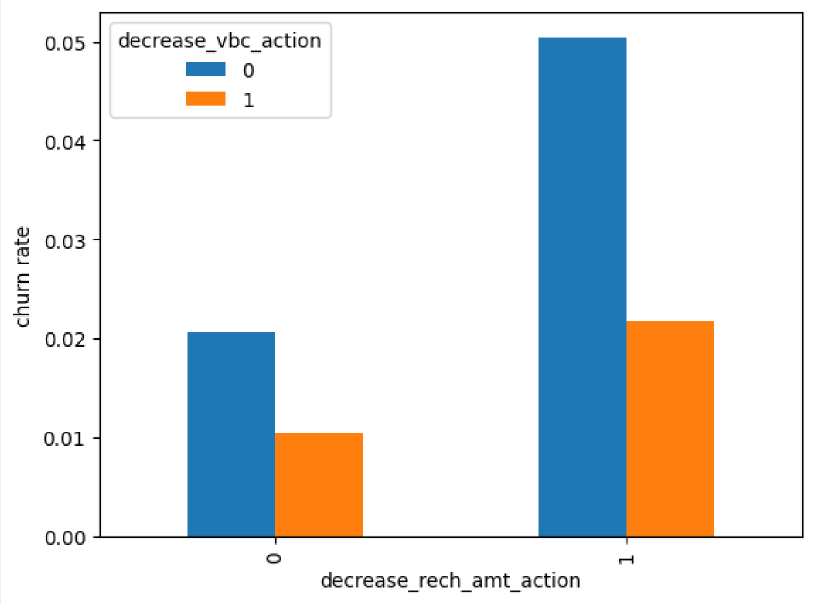


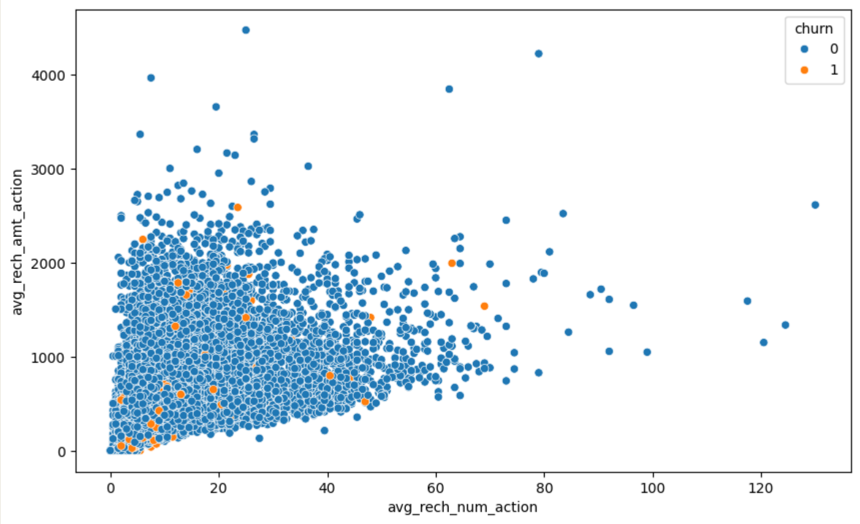




EDA:Bivariate analysis

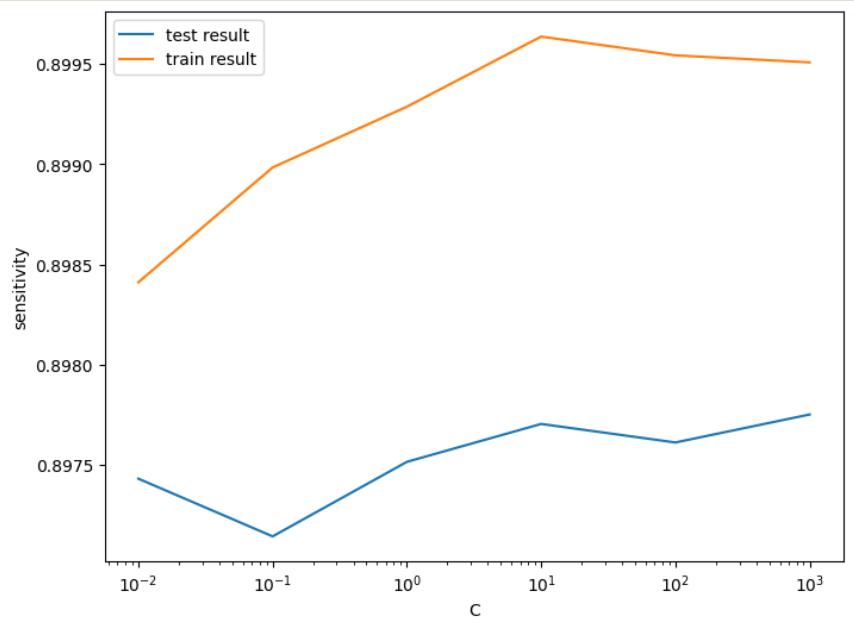


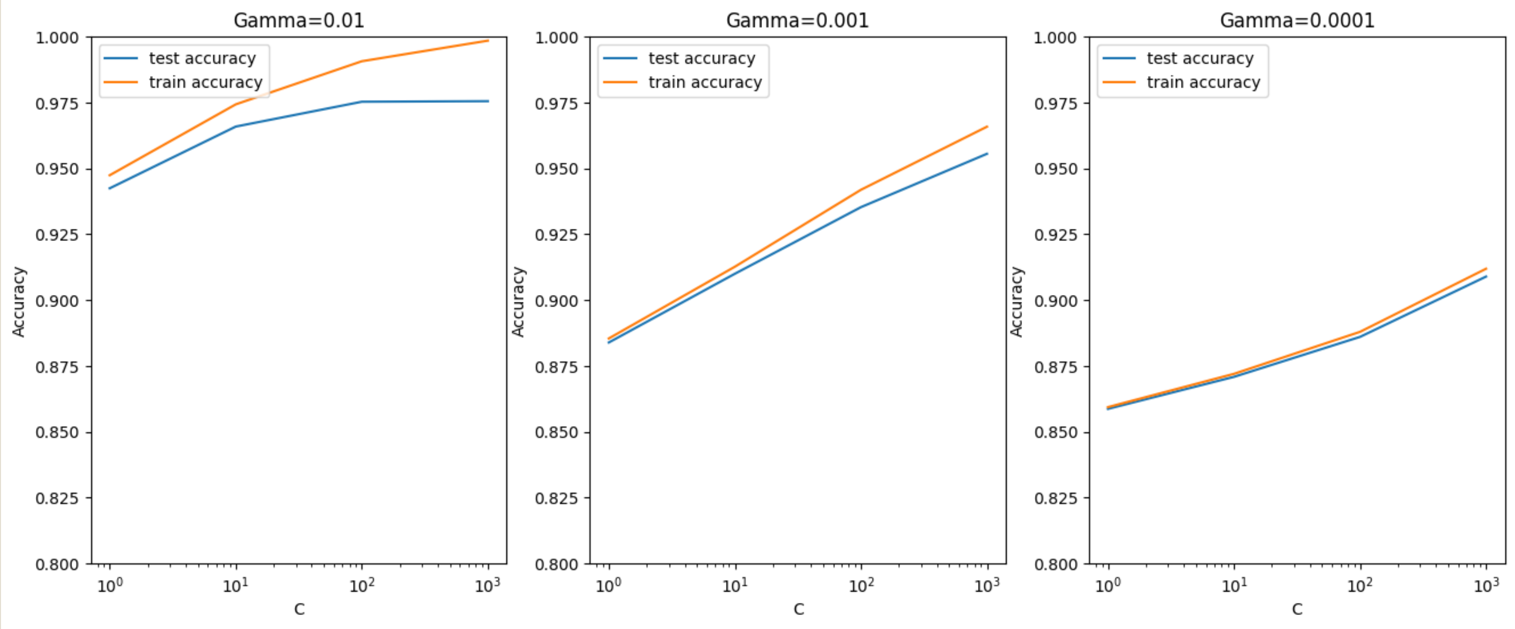




4

Model with PCA





Logistic regression

Accuracy:- 0.8314383685255369
Sensitivity:- 0.8134715025906736
Specificity:- 0.8320867614061331

SVM

Accuracy:- 0.8507489622811767
Sensitivity:- 0.8134715025906736
Specificity:- 0.8520942408376964

Decision Tree

Accuracy:- 0.8603140227395777
Sensitivity:- 0.6994818652849741
Specificity:- 0.8661181750186986

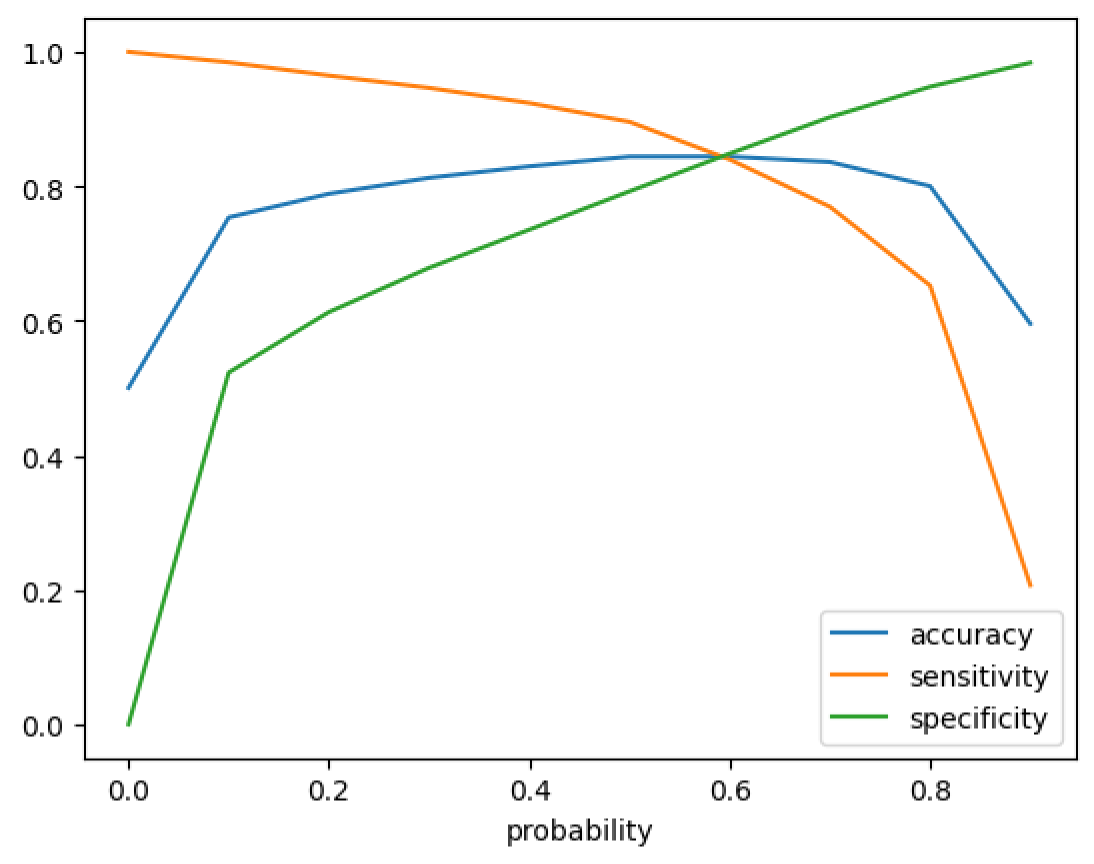
Random forest

Accuracy:- 0.7993142032124165
Sensitivity:- 0.7564766839378239
Specificity:- 0.8008601346297681

5

Model with

No PCA



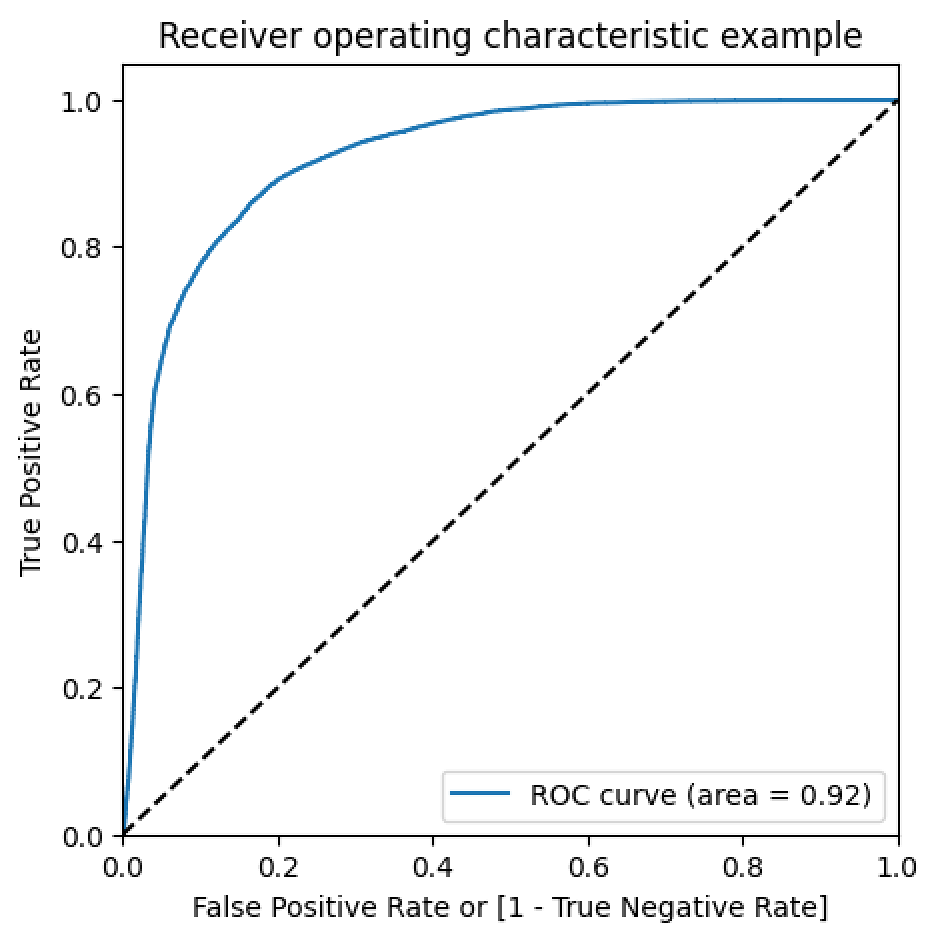
Analysis of the curve

Accuracy - Becomes stable around 0.6

Sensitivity - Decreases with the increased probablity.

Specificity - Increases with the increasing probablity.

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



ROC Curve

the area of the ROC curve is closer to 1, which is the Gini of the model.

Model summary

- Train set
  - Accuracy = 0.84
  - Sensitivity = 0.81
  - Specificity = 0.83
- Test set
  - Accuracy = 0.78
  - Sensitivity = 0.82
  - Specificity = 0.78

6

Business Recomendation

Top Predictors of Churn

The table below presents key variables identified in the logistic regression model that significantly contribute to predicting customer churn.

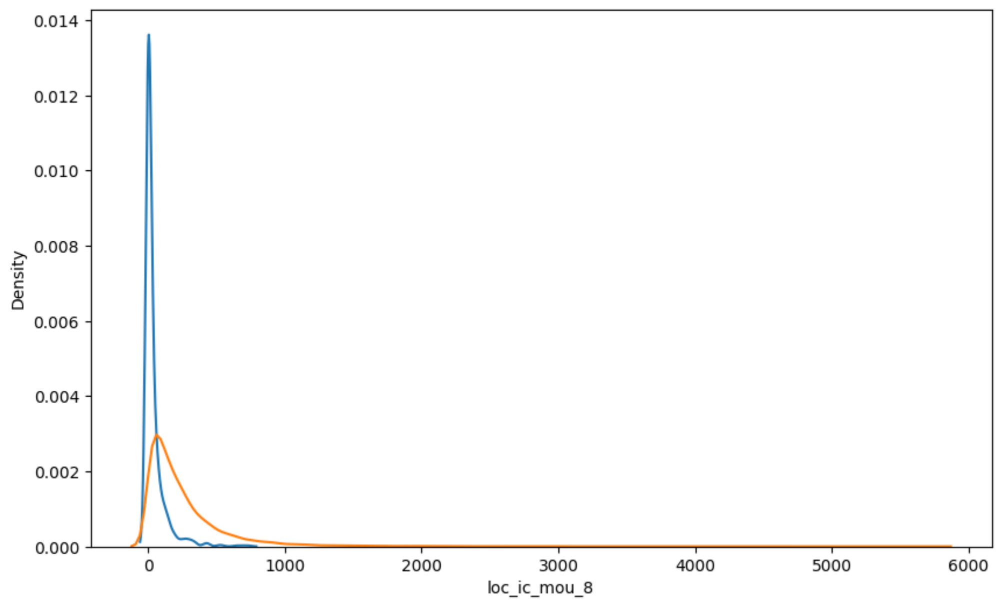
Variable	Coefficient
loc_ic_mou_8	-3.3287
og_others_7	-2.4711
ic_others_8	-1.5131
isd_og_mou_8	-1.3811
decrease_vbc_action	-1.3293
monthly_3g_8	-1.0943
std_ic_t2f_mou_8	-0.9503
monthly_2g_8	-0.9279
loc_ic_t2f_mou_8	-0.7102
roam_og_mou_8	0.7135

The majority of these variables exhibit negative coefficients, indicating an inverse relationship with churn probability.

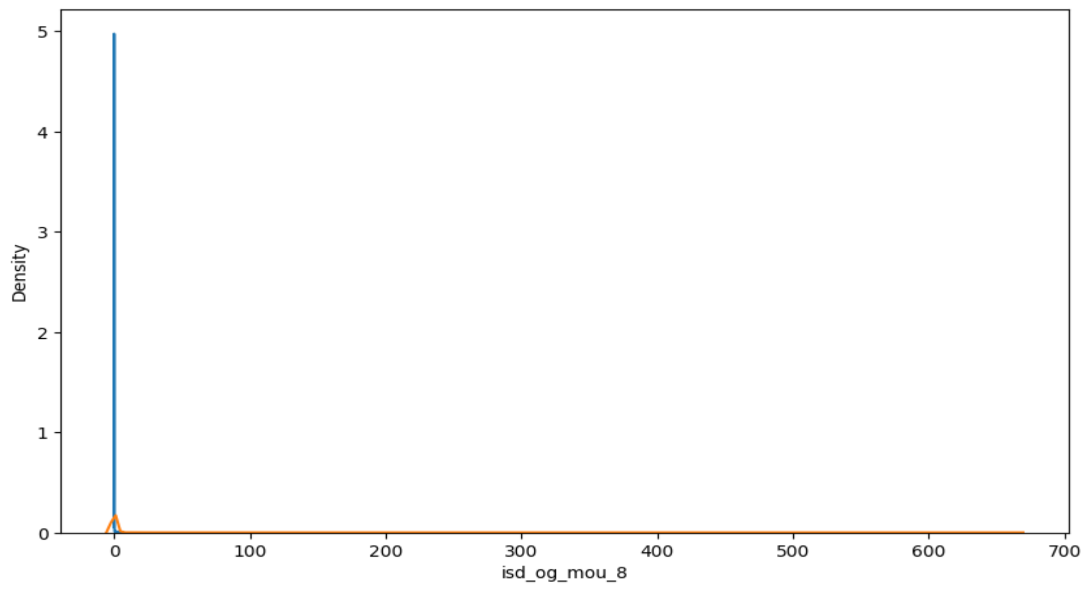
For example, a decrease in local incoming minutes of usage (loc_ic_mou_8) in August correlates with a higher likelihood of churn.

Recommendations for Customer Retention

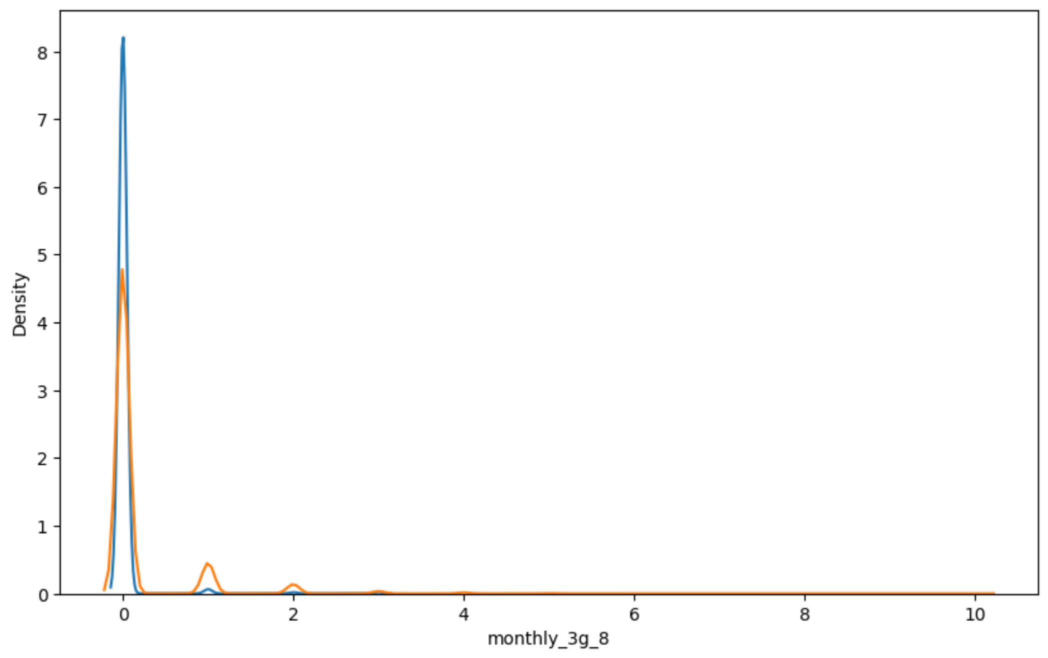
- Target customers with reduced local incoming and outgoing ISD call usage during the action phase (August).
- Identify customers with decreased outgoing charges in July and reduced incoming charges in August, as they exhibit a higher churn probability.
- Monitor customers whose value-based cost has increased during the action phase, as they are more likely to churn. Offering tailored incentives may help retain them.
- Customers who have increased their monthly 3G recharge in August are at risk of churning. Consider offering data-based retention plans.
- Customers exhibiting a decline in STD incoming minutes of usage (T to fixed-line) in August should be prioritized for retention strategies.
- A reduction in monthly 2G usage in August is also a strong indicator of potential churn.
- A decrease in incoming minutes of usage for T to fixed-line calls in August is another key churn predictor.
- Customers with increasing roaming outgoing minutes of usage (roam_og_mou_8, coefficient: 0.7135) are more likely to churn, necessitating targeted engagement strategies.
- By addressing these predictive factors and implementing targeted retention strategies, customer churn can be effectively mitigated.



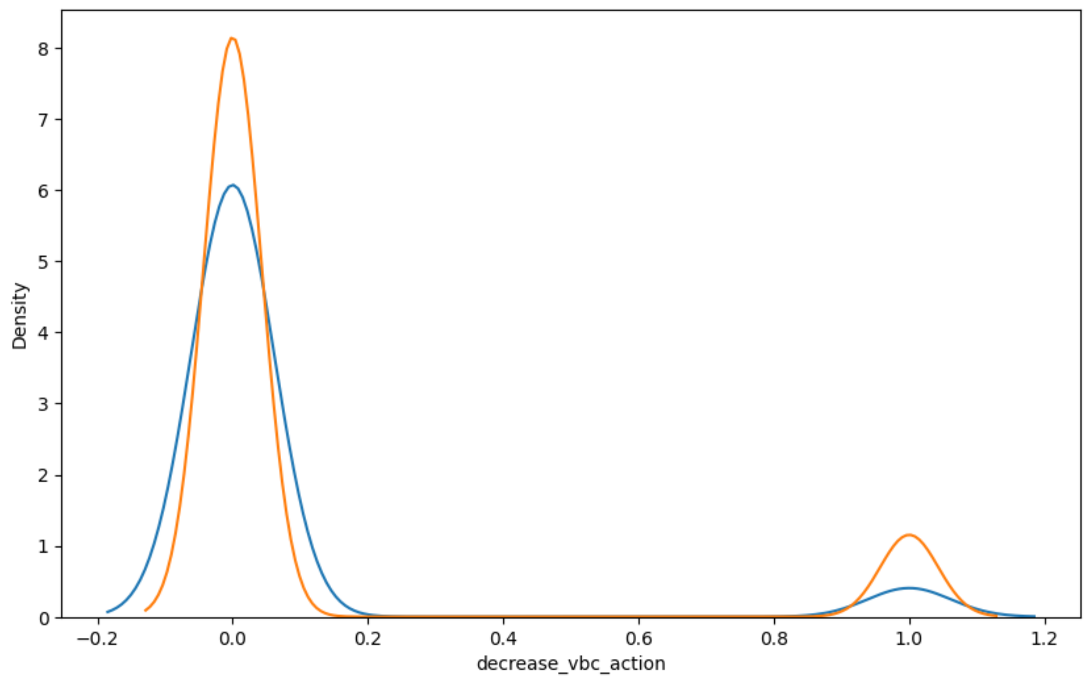
Density
loc_ic_mou_8



Density
isd_og_mou_8



Density
monthly_3g_8



Density
decrease_vbc_action

THANK YOU !