Telecom Churn Case Study

Presented by

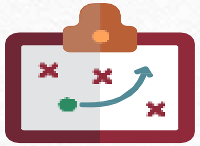
Manali Shah

Tummalacheruvu Baji Babu

Srinivasacharyulu Daroori

Problem Statement

- Customer retention has now become even more important than customer acquisition To reduce customer churn, telecom companies need to predict which customers are at high risk of churn. In the Indian and Southeast Asian markets, approximately 80% of revenue comes from the top 20% of customers (called high-value customers). churn prediction is usually more critical (and non-trivial) for prepaid customers - Usage based churn Thus, if we can reduce the churn of high-value customers, we will be able to reduce significant revenue leakage. High-value customers to be defined based on a certain metric and predict churn only on high-value customers.



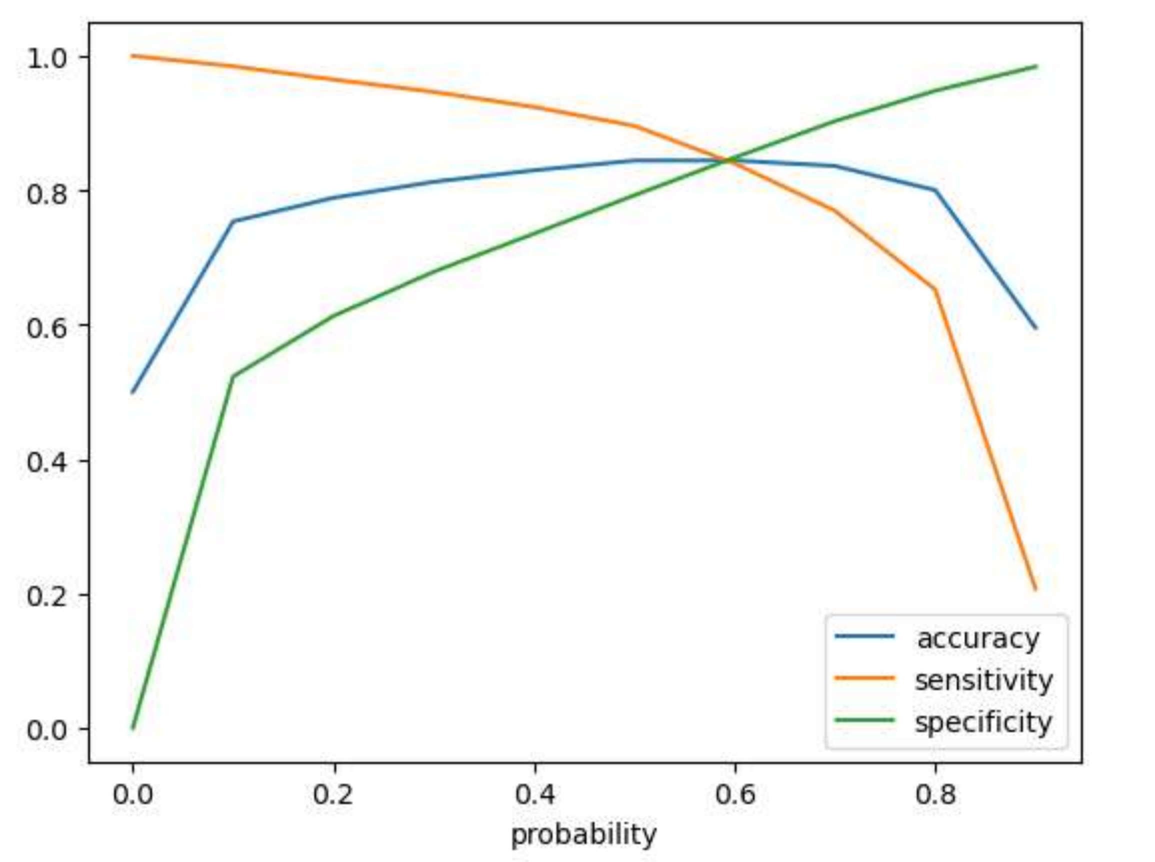
Objective

- To predict customer churn.
- Highlight the main variable / factors influencing customer churn
- Use various ML algorithm to build prediction models, evaluate the accuracy and performance of these models.
- Finding out the best model for our business case & providing executive summary.

APPROACH

- **Problem Statement**
- Data Understanding, cleaning and preparation
- **High value customers**
- **Outlier treatment**
- Exploratory Data Analysis
- **Dealing with Imbalanced Data**
- **Logistic regression with PCA / Logistic regression with No PCA**
- **Logistic regression with optimal C**
- **Support Vector Machine(SVM) with PCA**
- **Decision tree with PCA**
- **Checking VIF for Model-2**
- **Model performance on the train set**

Model performance on the train set



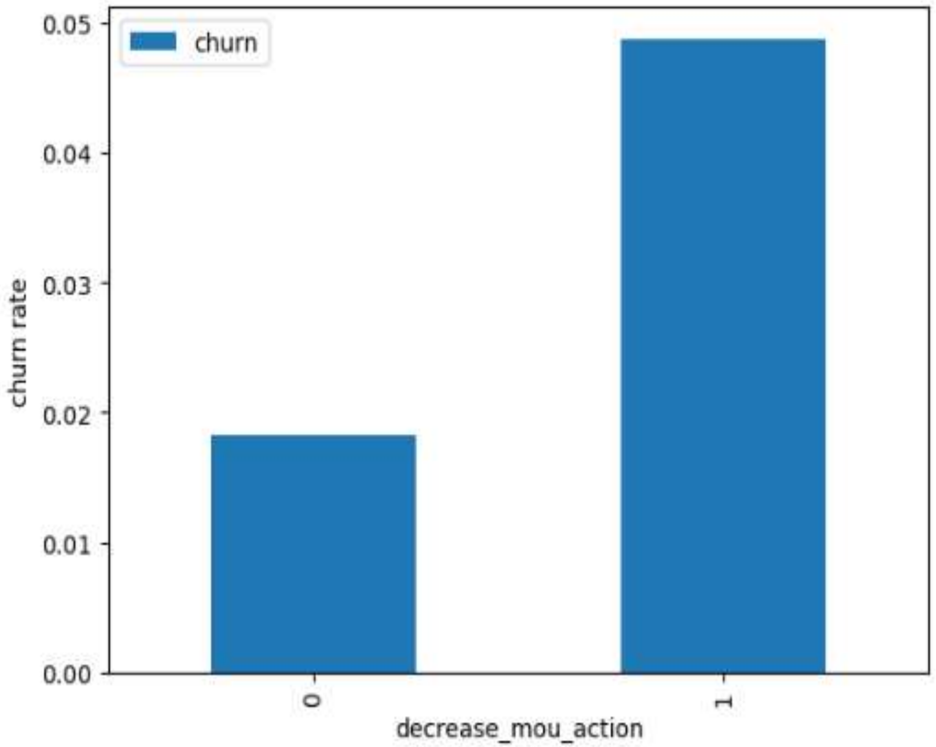
We can see that for the churn customers the minutes of usage for the month of August is mostly populated on the lower side than the non churn customers.

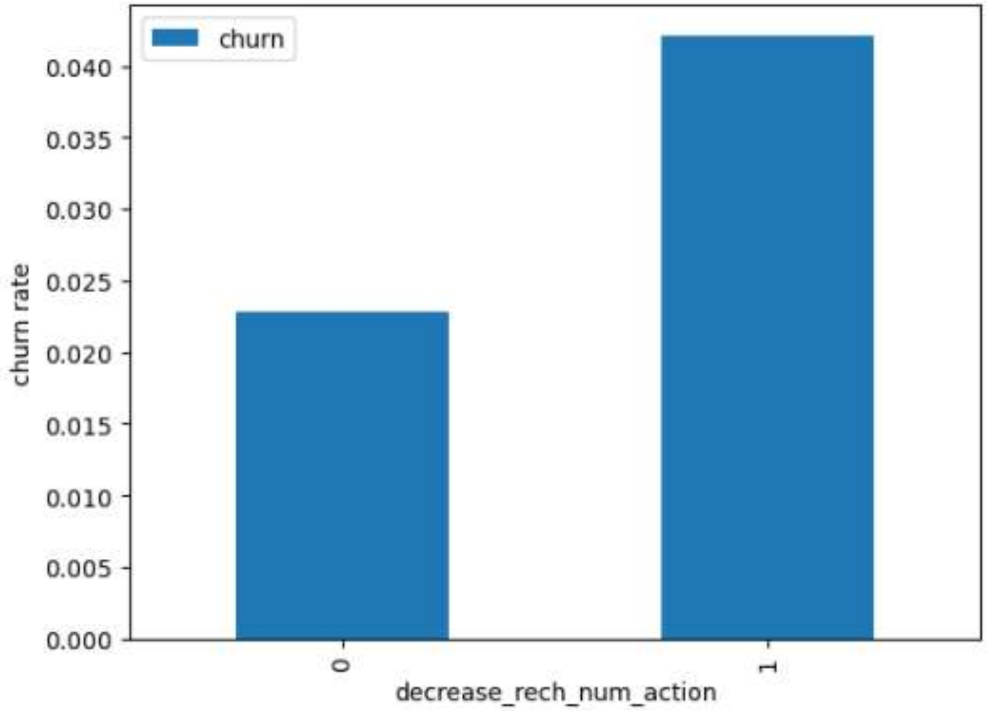
MODEL PERFORMANCE ON THE TRAIN SET

- 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.

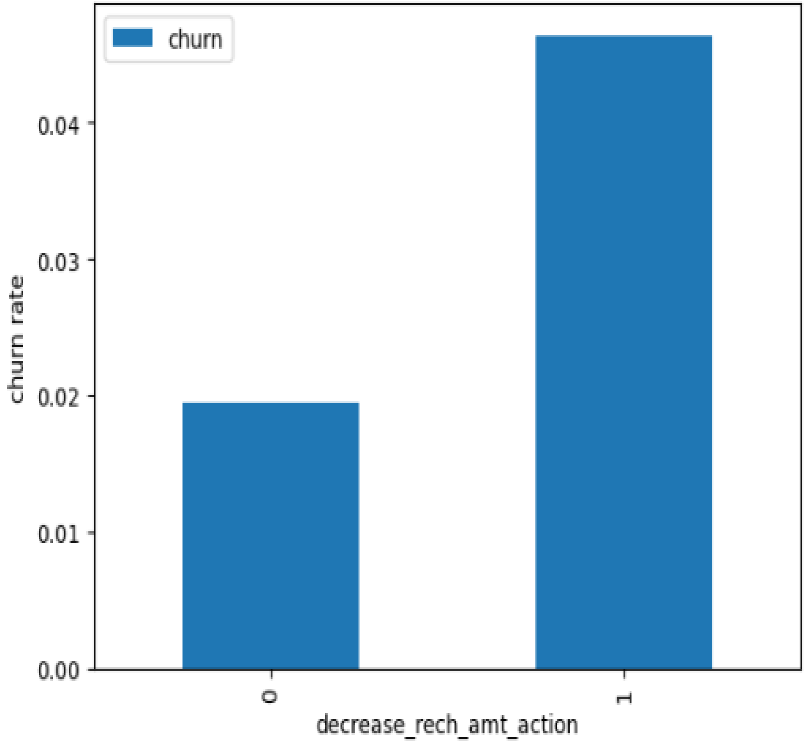
- Here we are intended to acheive better sensitivity than accuracy and specificity. Though as per the above curve, we should take 0.6 as the probability cutoff, we are taking 0.5 for acheiving higher sensitivity, which is our main goal.

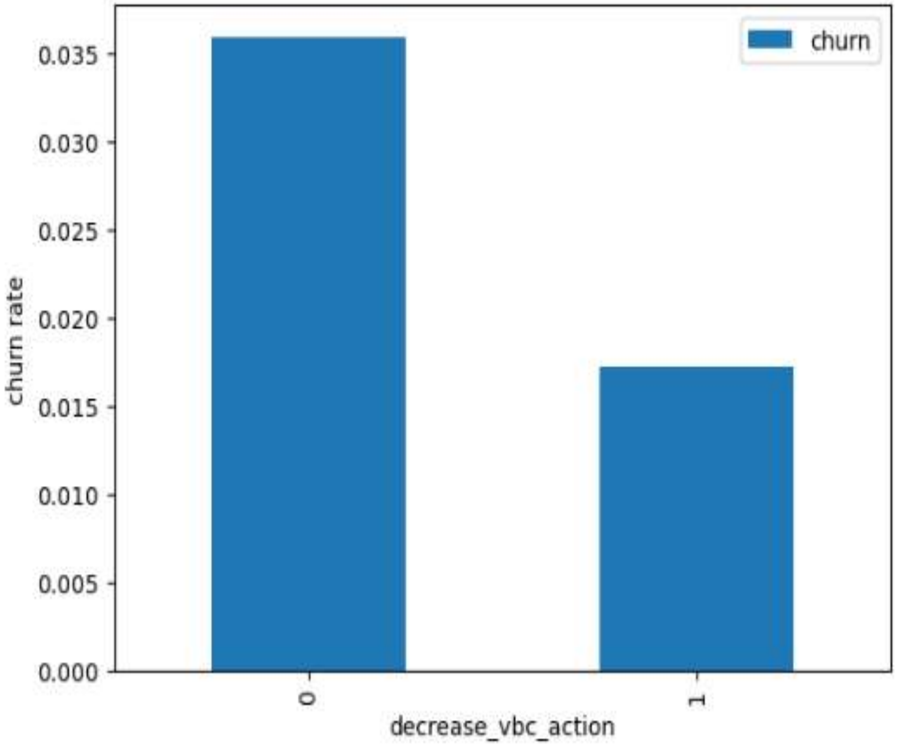
Exploratory Data Analysis





•**Decrease in Usage (MOU):**
Customers with reduced usage are more likely to churn.
•**Decrease in Recharges:**
A drop in recharge frequency strongly correlates with higher churn rates.
Key Insight:
Monitoring and addressing declines in usage or recharges can help reduce churn.





Decrease in Recharge Amount:

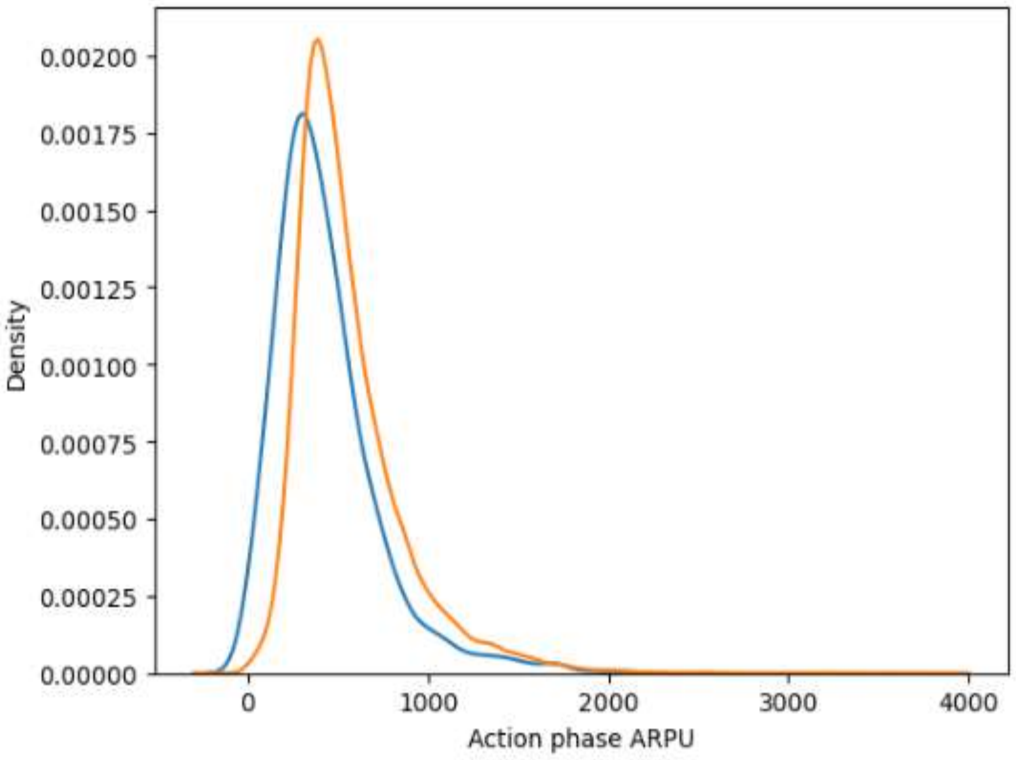
Customers with reduced recharge amounts (decrease_rech_amt_action = 1) show a higher churn rate compared to those with stable recharge amounts.

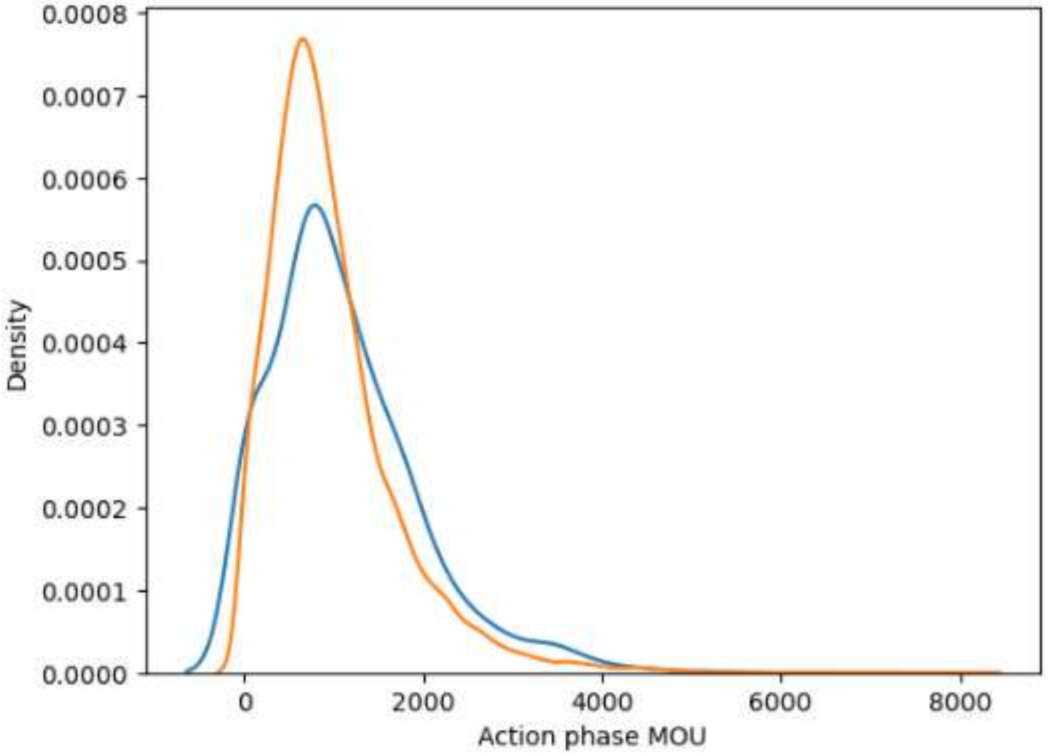
Decrease in Volume-Based Charges (VBC):

Surprisingly, customers with decreased VBC usage (decrease_vbc_action = 1) exhibit a lower churn rate, indicating a potential shift in usage behavior rather than dissatisfaction.

Key Insight:

Declining recharge amounts are strongly linked to churn, while a decrease in VBC usage might not necessarily indicate dissatisfaction.





•**Action Phase ARPU (Average Revenue Per User):**

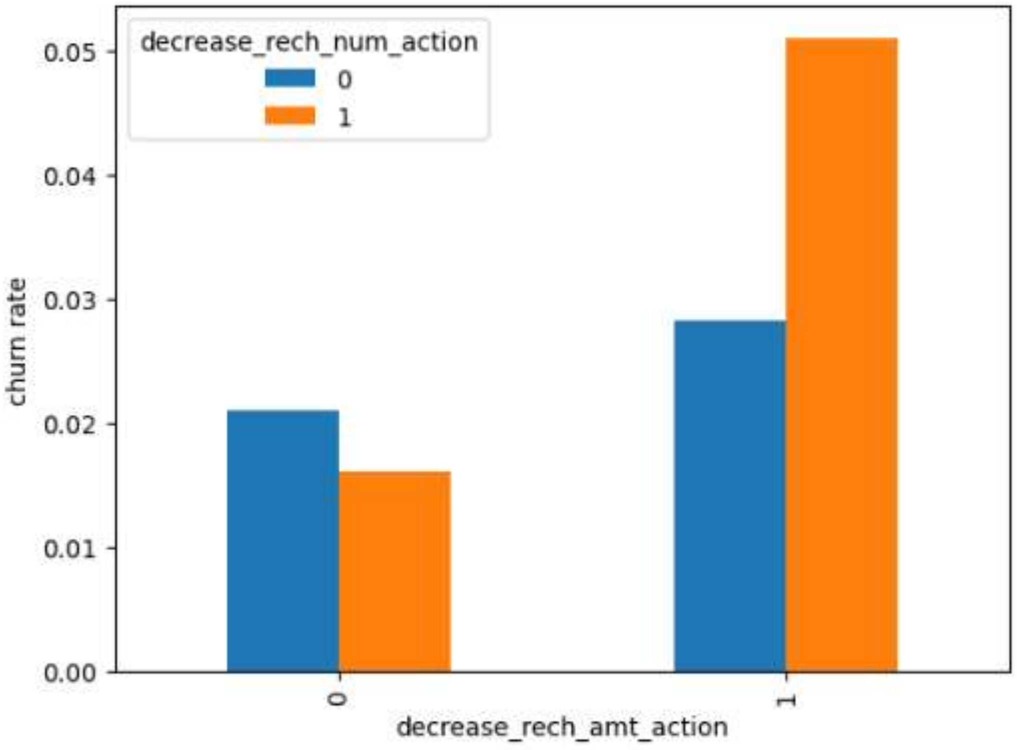
Customers who churn (orange line) generally have lower ARPU compared to non-churners (blue line).

•**Action Phase MOU (Minutes of Usage):**

Churned customers (orange line) tend to have slightly lower MOU distributions compared to non-churners (blue line).

Key Insight:

Lower ARPU and MOU during the action phase are indicators of customer churn.



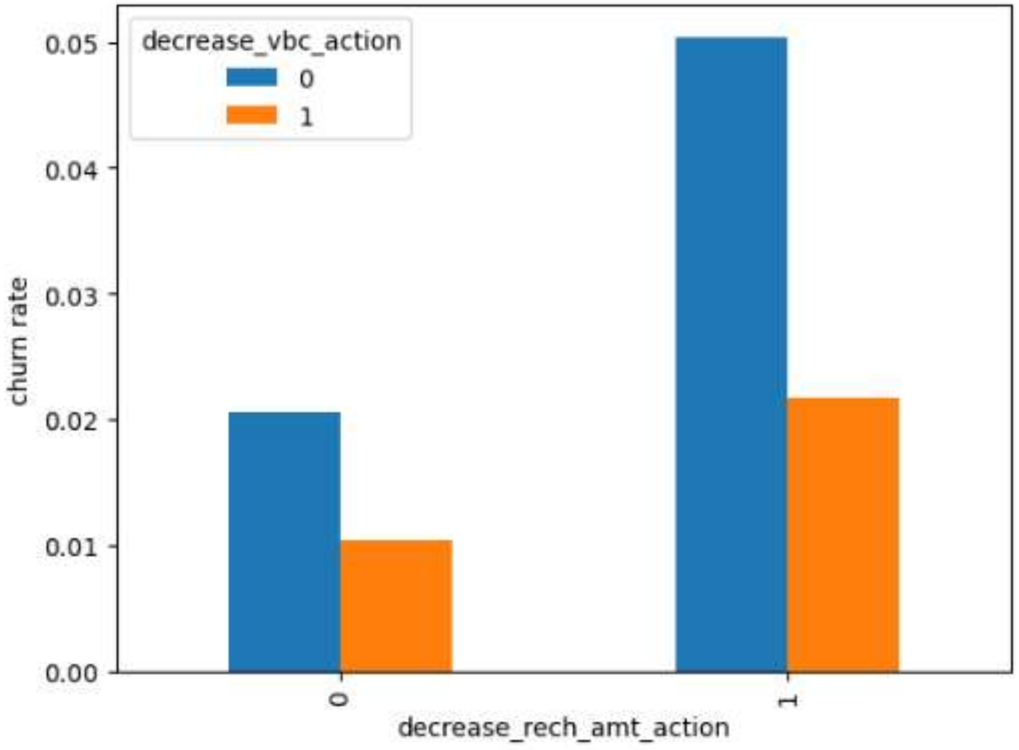
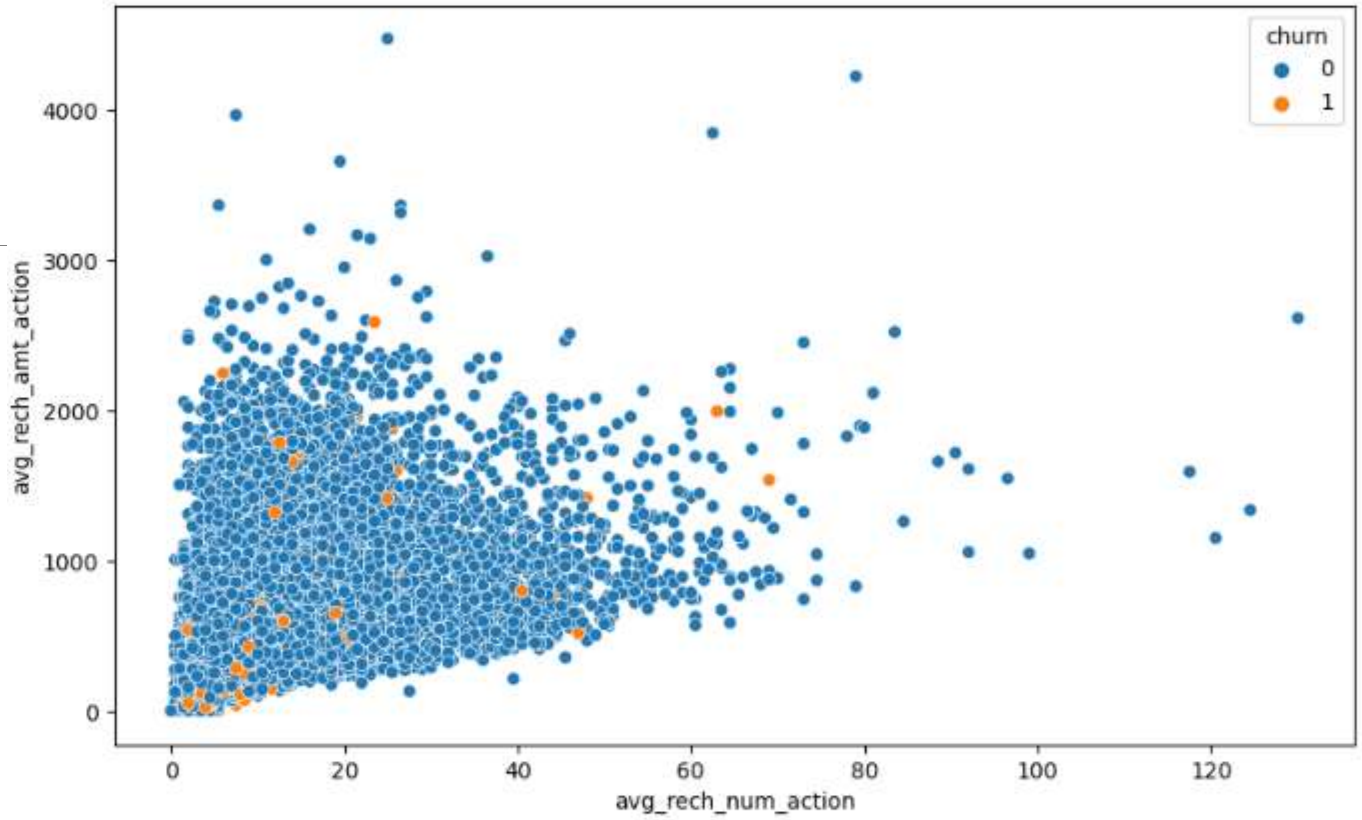
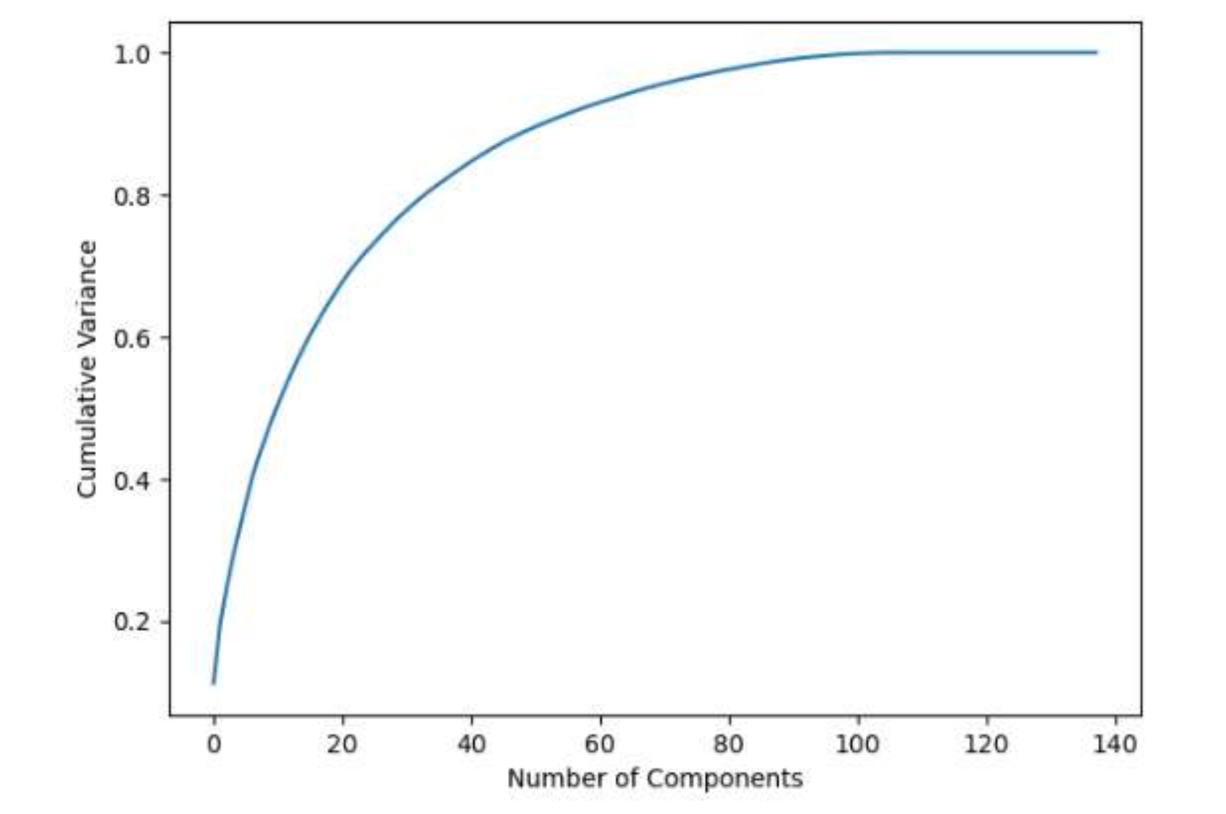


Image 1: Customers who decrease recharge frequency are more likely to churn.
Image 2: Customers who decrease recharge amount are more likely to churn.



The scatter plot shows a negative correlation between recharge amount and churn, indicating that customers with higher recharge amounts are less likely to churn.

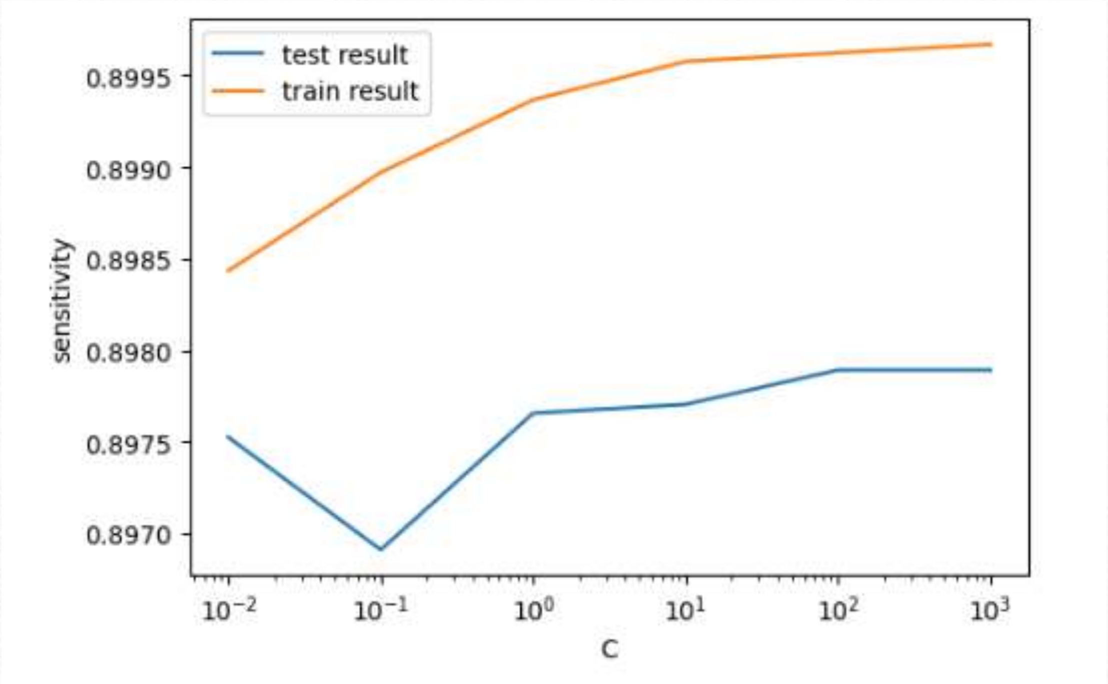
MODEL WITH PCA



Transformation with the already fitted data on the train set.

The plot shows the cumulative variance explained by different numbers of principal components. This helps us choose the optimal number of components to retain most of the information while reducing dimensionality.

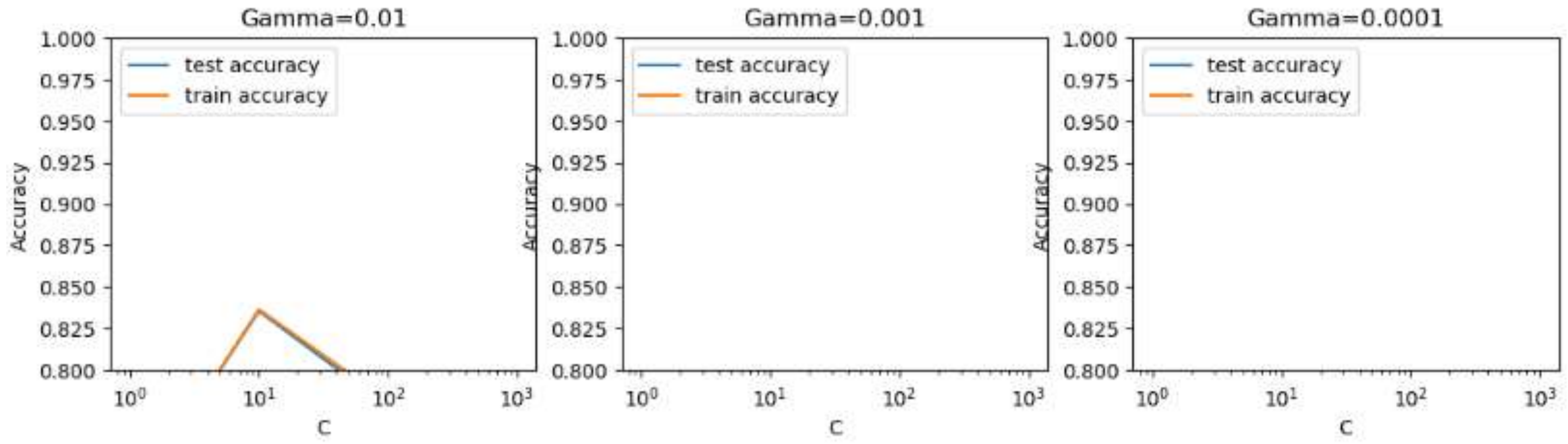
Logistic regression with PCA



The plot shows the sensitivity of the logistic regression model with PCA for different values of the regularization parameter C. The highest test sensitivity is achieved at C = 100.

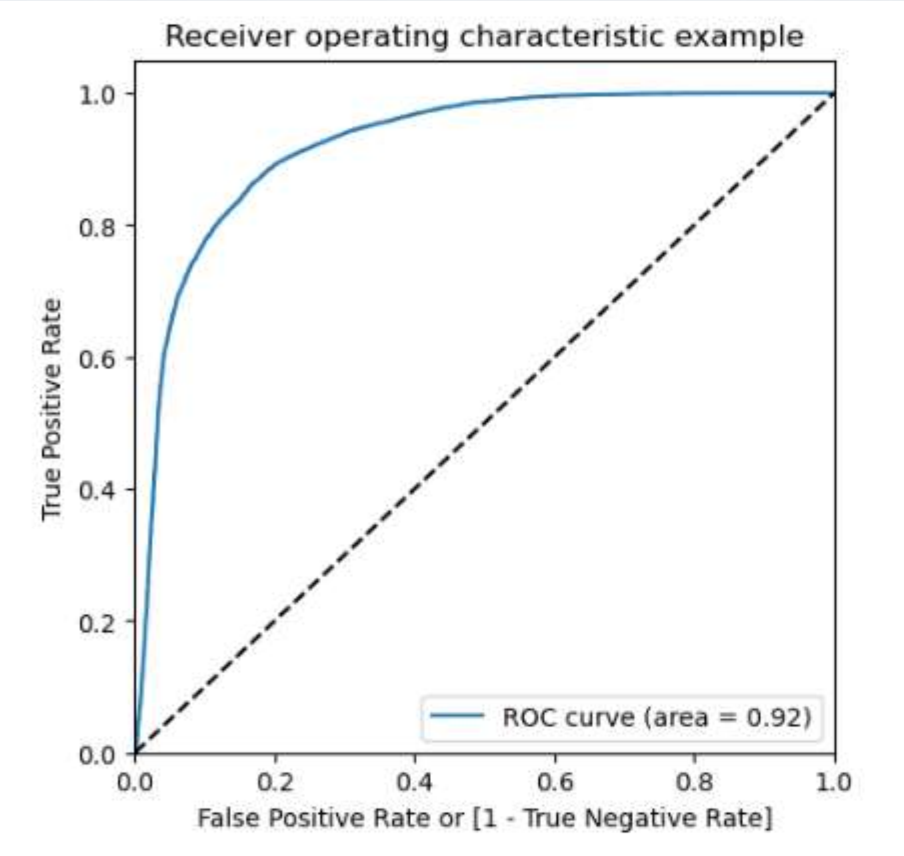
The highest test sensitivity is 0.8978916608693863 at C = 100

PLOTTING THE ACCURACY WITH VARIOUS C AND GAMMA VALUES



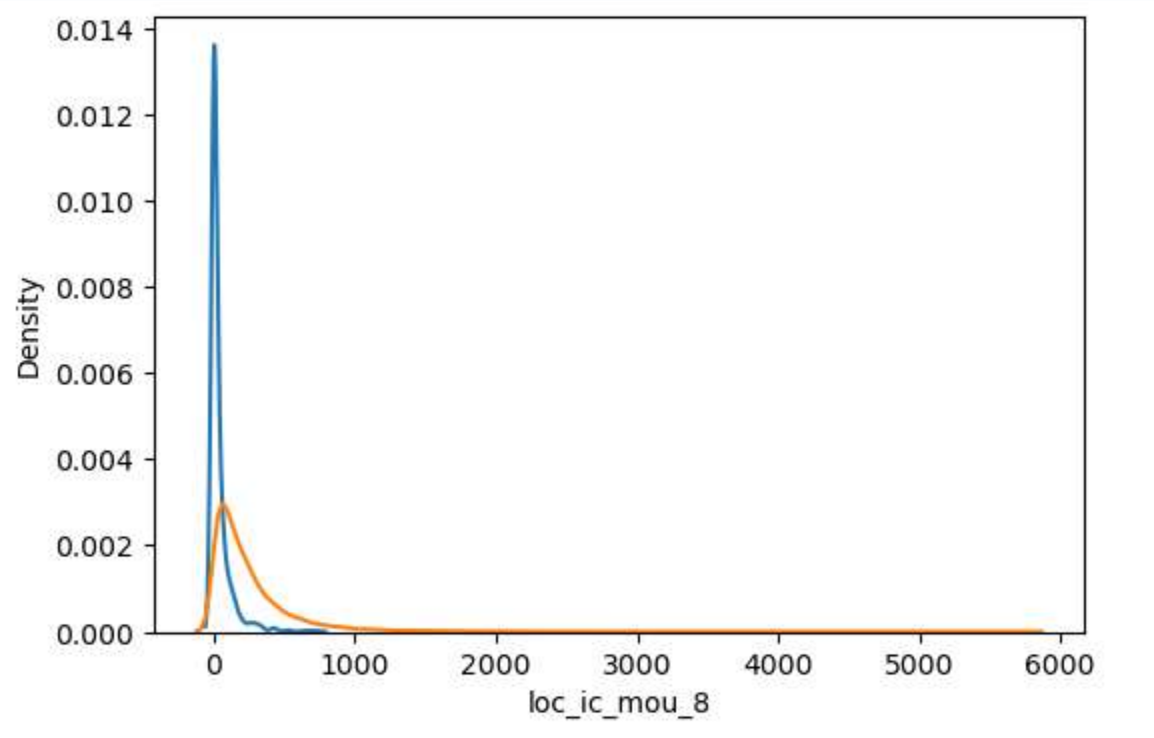
The best test score is 0.835613160694502 corresponding to hyperparameters {'C': 10, 'gamma': 0.01}

RECEIVER OPERATING CHARACTERISTIC EXAMPLE



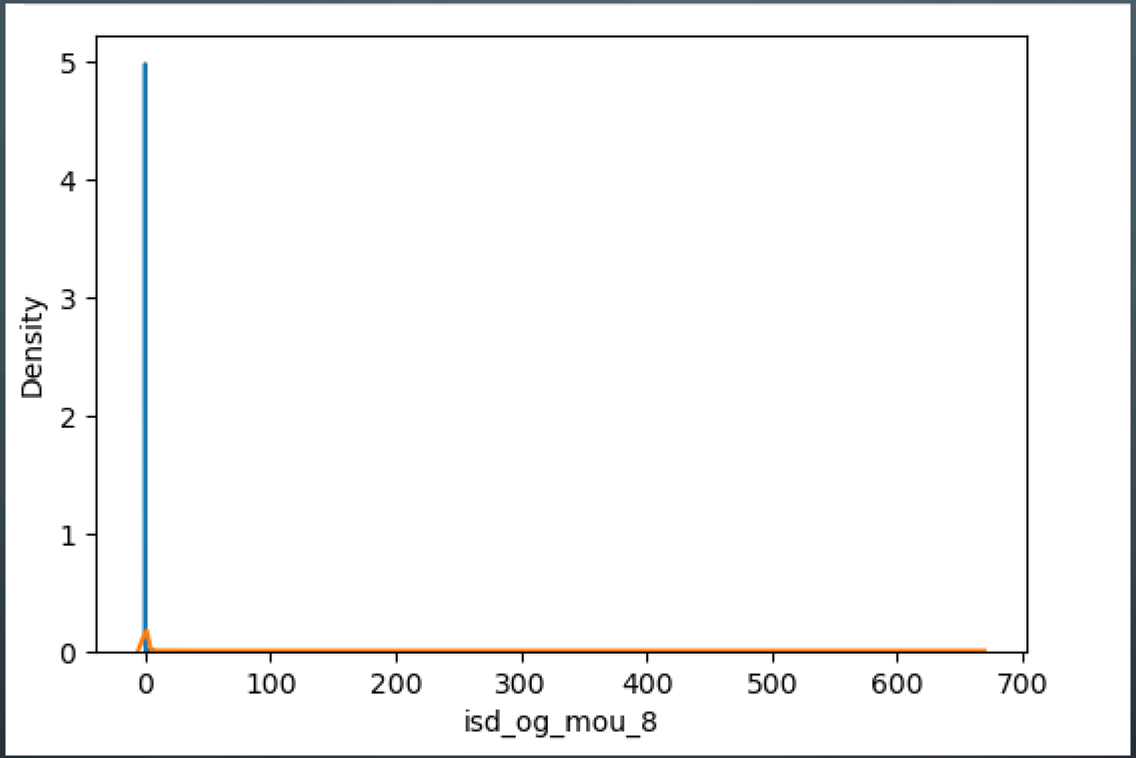
We can see the area of the ROC curve is closer to 1, which is the Gini of the model.

Plots of important predictors for churn and non churn customers

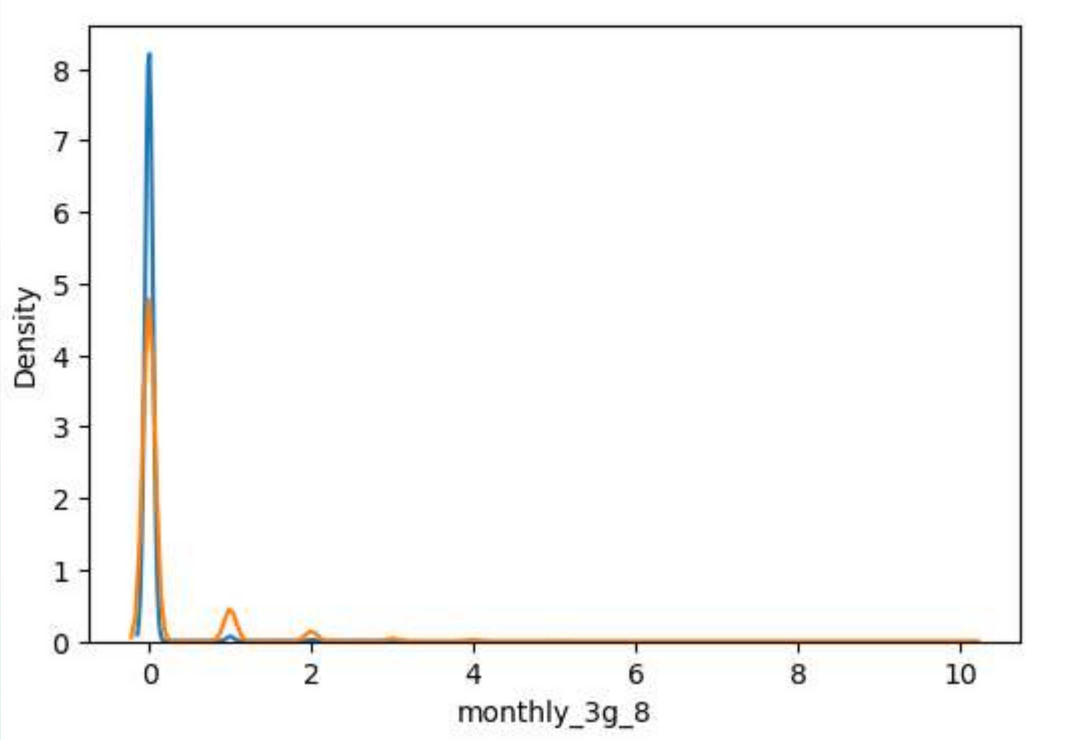


We can see that for the churn customers the minutes of usage for the month of August is mostly populated on the lower side than the non churn customers.

PLOTS OF IMPORTANT PREDICTORS FOR CHURN AND NON CHURN CUSTOMERS



Plots of important predictors for churn and non churn customers



The plot shows that churned customers tend to have higher monthly 3G usage compared to non-churned customers. This suggests that high 3G usage could be a factor contributing to customer churn.

MODEL SUMMARY

Overall, the model is performing well in the test set, what it had learnt from the train set.

	Train Data:	Test Data:
Accuracy :	0.84	0.78
Sensitivity :	0.89	0.82
Specificity :	0.79	0.78

THANK YOU