Telecom Churn Case Study

ABINESH INBARAJAN

SATYAM SHAH

HARSH KANT TIWARI

Problem Statement

Within the telecommunications sector, customers possess the freedom to opt for various service providers, often transitioning between operators. In this fiercely competitive landscape, the industry encounters an annual churn rate averaging between 15-25%. Considering that acquiring new customers incurs expenses 5-10 times greater than retaining existing ones, prioritizing customer retention has escalated in significance over customer acquisition. To combat customer attrition, telecom companies must proactively anticipate and identify customers predisposed to churn.

Business Goal

- This project involves analyzing customer-level data from a prominent telecom company, constructing predictive models to pinpoint customers with a high likelihood of churning, and determining the primary indicators linked to customer attrition.

Steps

01

Reading, understanding and visualising the data

02

Preparing the data for modelling

03

Building the model

04

Evaluate the model

Methodology

- Data cleaning and data manipulation.
- Handle duplicate data.
- Handle NA values and missing values.
- Drop columns, if it contains large number of missing values.
- Imputation of the values.
- Handle outliers in data.
- EDA
- Univariate data analysis
- Bivariate data analysis

Methodology

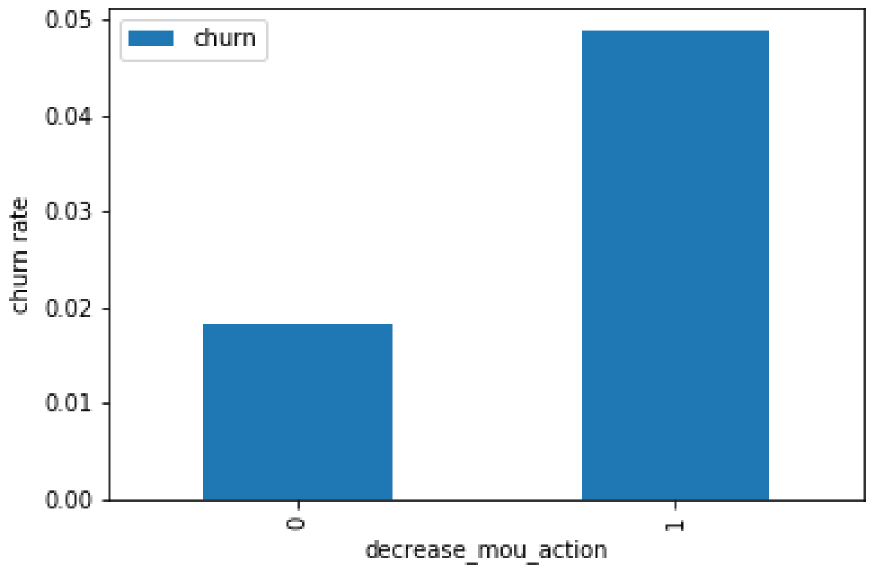
The data preprocessing stage involves Feature Scaling and the creation of Dummy Variables for appropriate data encoding.

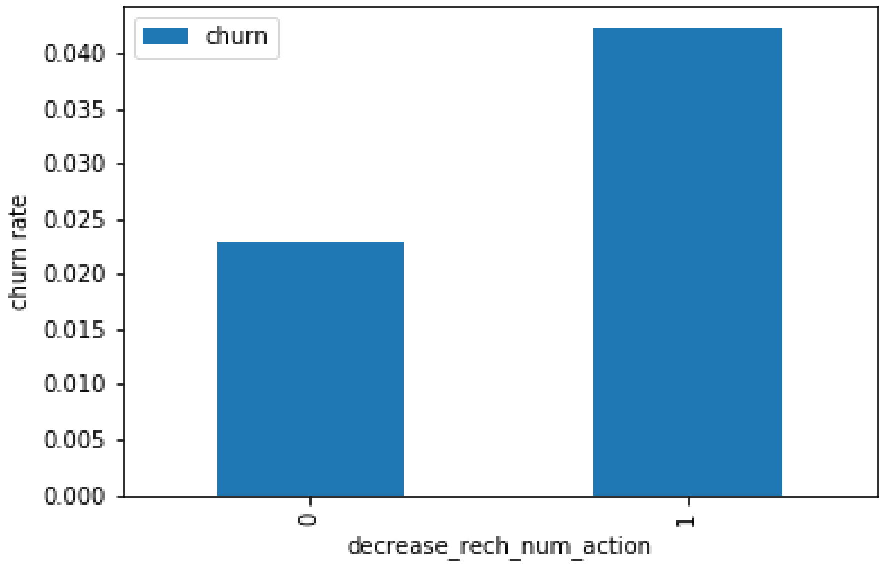
Subsequently, the chosen classification technique, Logistic Regression, is employed for model construction and prediction.

A rigorous validation process is conducted to assess the model's accuracy and effectiveness.

Following validation, the model's findings and insights are presented comprehensively.

In conclusion, the study offers valuable insights, and based on these findings, recommendations are provided to guide informed decision-making and further actions.

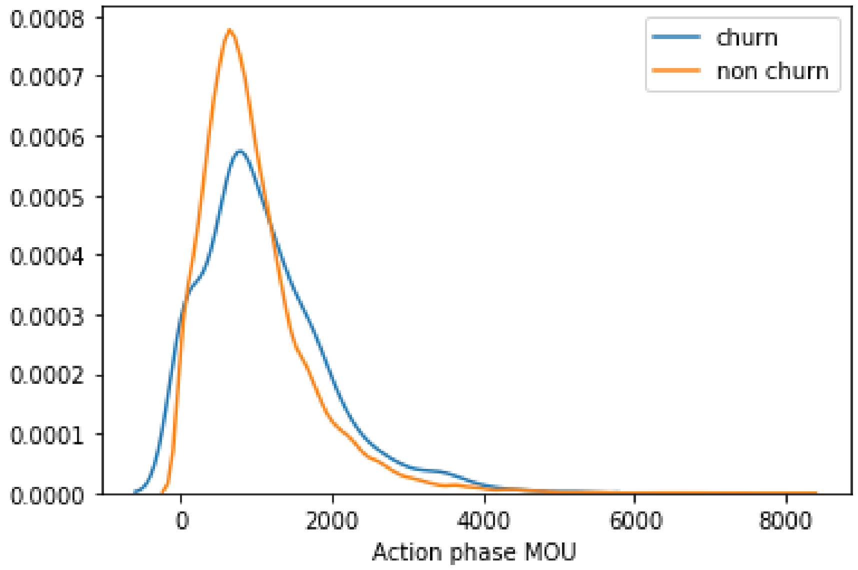


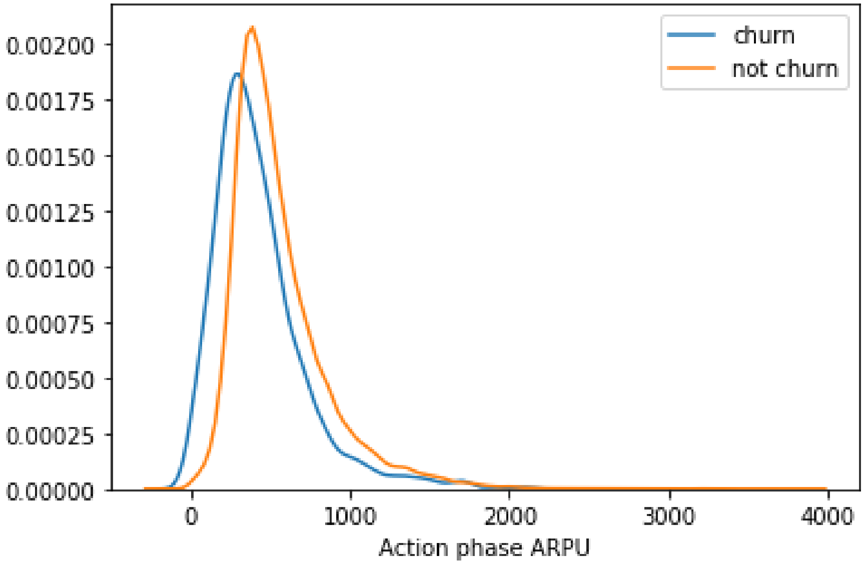


Univariate Analysis

In the first plot, it is observed trend indicates a higher churn rate among customers whose usage minutes (MOU) declined during the action phase compared to the good phase.

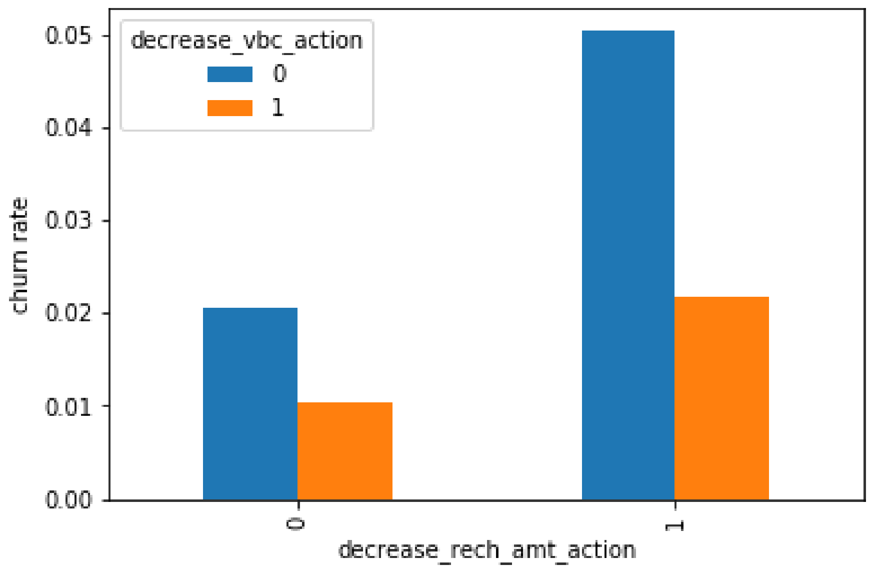
In the second plot, it is also observed that customers exhibiting a lower number of recharges during the action phase experience a higher churn rate compared to those in the good phase.

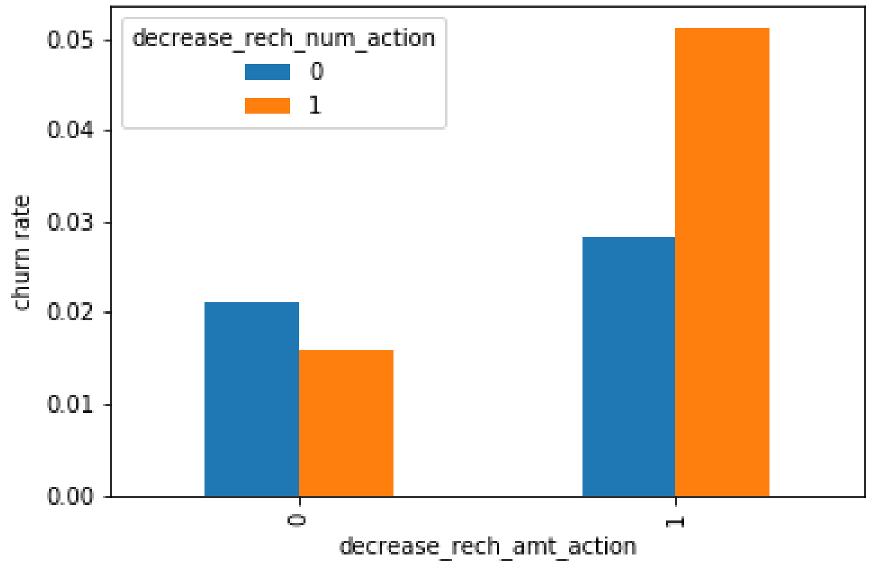




Univariate Analysis

- In the first plot, the distribution of Minutes of Usage (MOU) among churned customers primarily falls within the range of 0 to 2500. Additionally, there's an inverse relationship between MOU and churn probability, suggesting that higher MOU values correlate with lower probabilities of churn.
- In the second plot, the data shows that the Average Revenue Per User (ARPU) among churned customers predominantly concentrates in the range of 0 to 900. Moreover, customers with higher ARPU tend to exhibit lower churn rates.
- Conversely, for non-churned customers, the ARPU is largely clustered within the range of 0 to 1000.

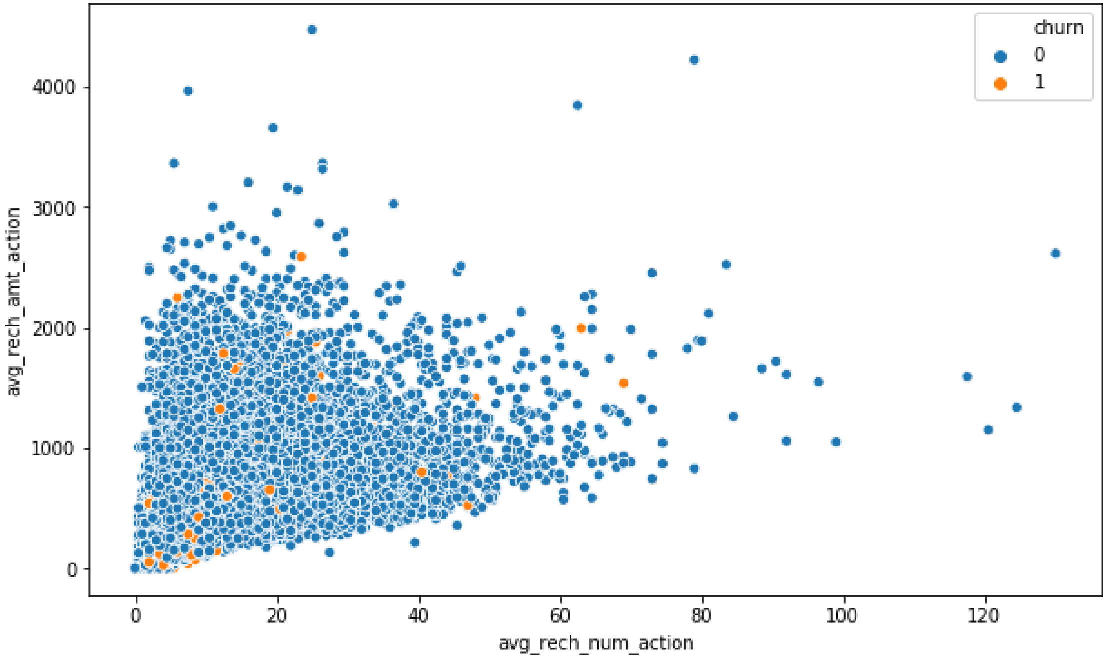




Bivariate Analysis

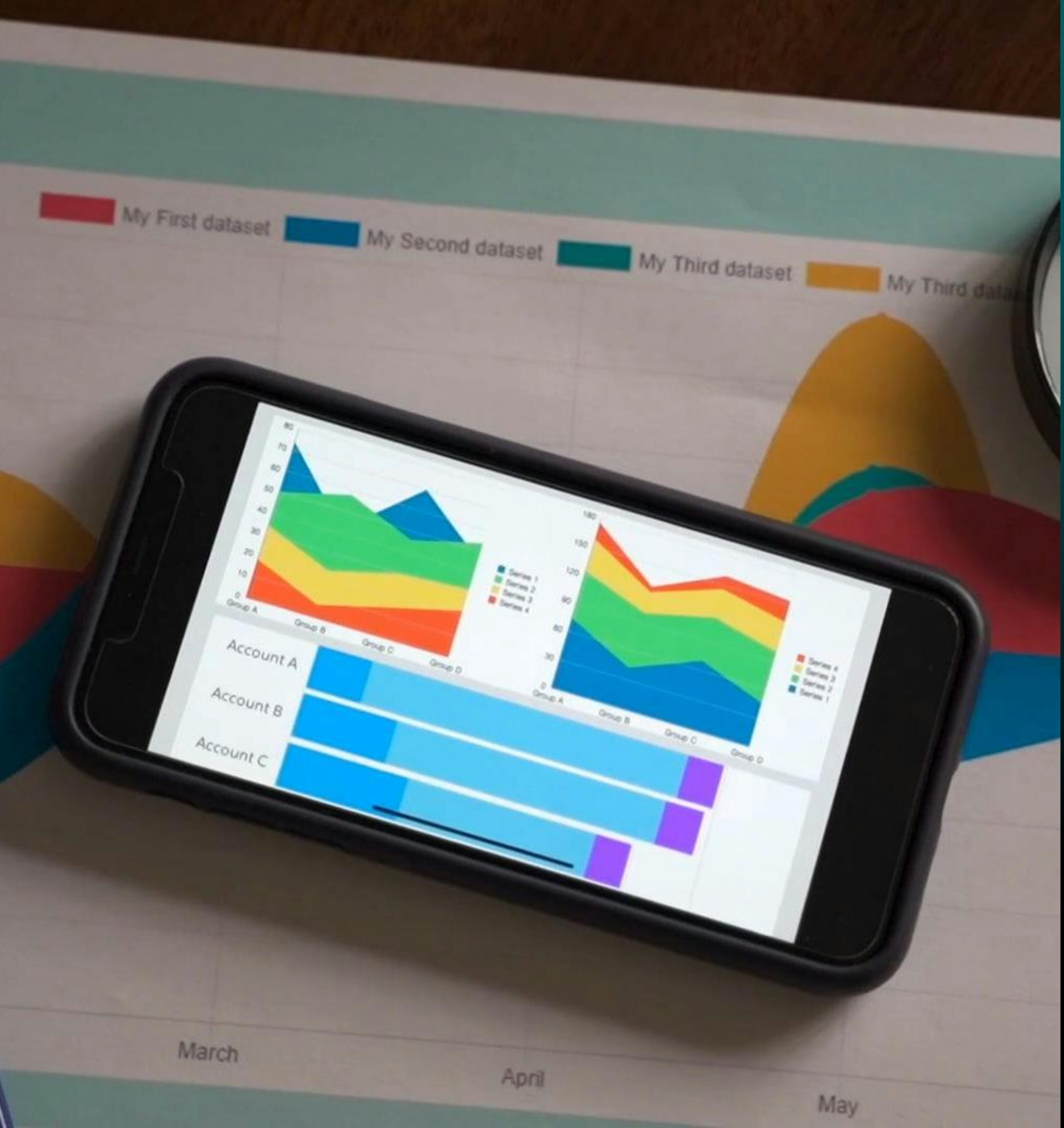
- It is observed from the first plot that , it's evident that customers facing a decrease in recharge amounts coupled with an increase in volume-based costs during the action month tend to exhibit a higher churn rate.

- As depicted in the second plot, customers who experience a decrease in both recharge amount and the number of recharges during the action phase exhibit a higher churn rate compared to those in the good phase.



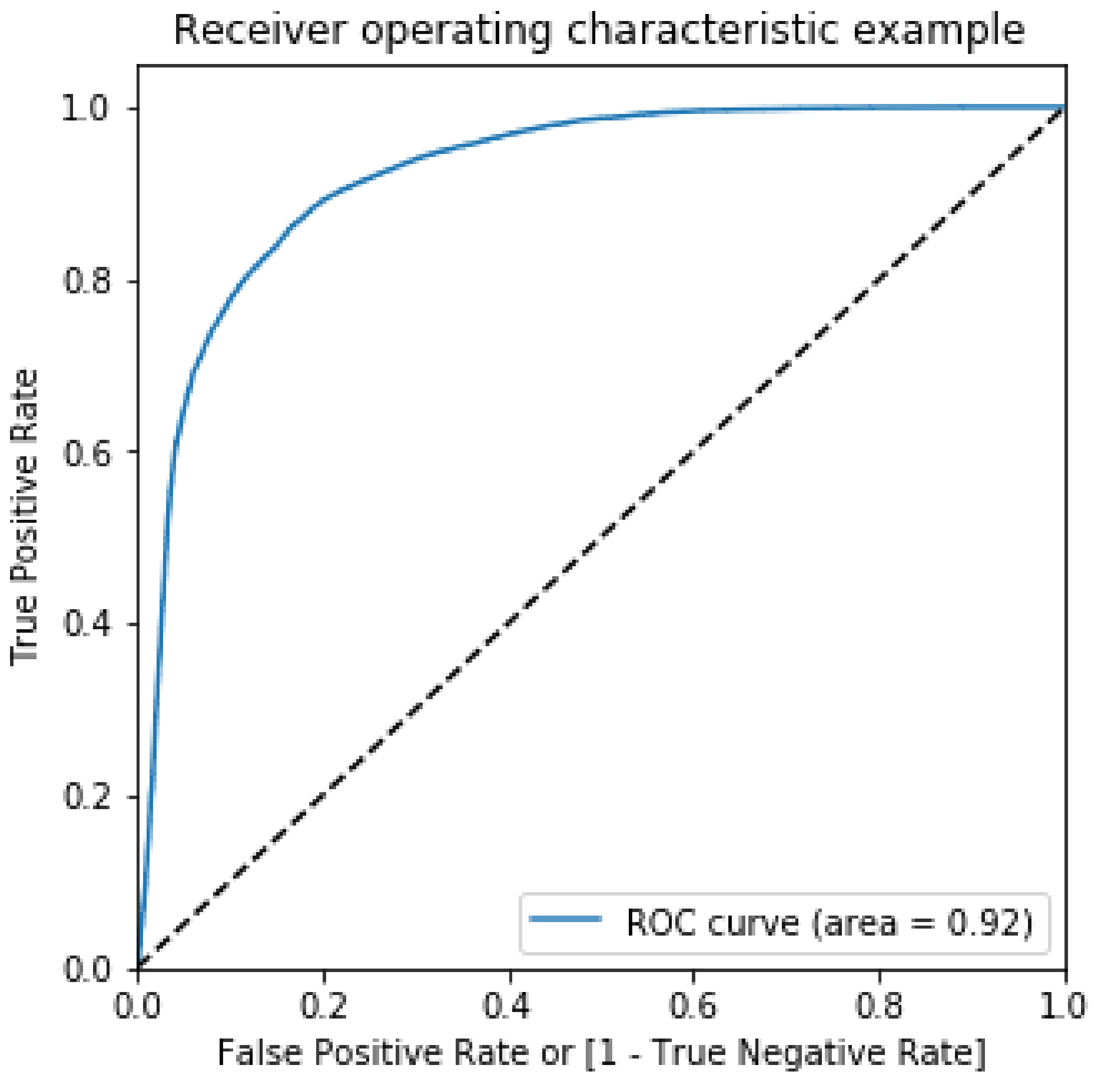
Bivariate Analysis

- The observed pattern indicates a proportional relationship between the number of recharges and the recharge amount. Specifically, as the number of recharges increases, so does the total amount spent on recharging.



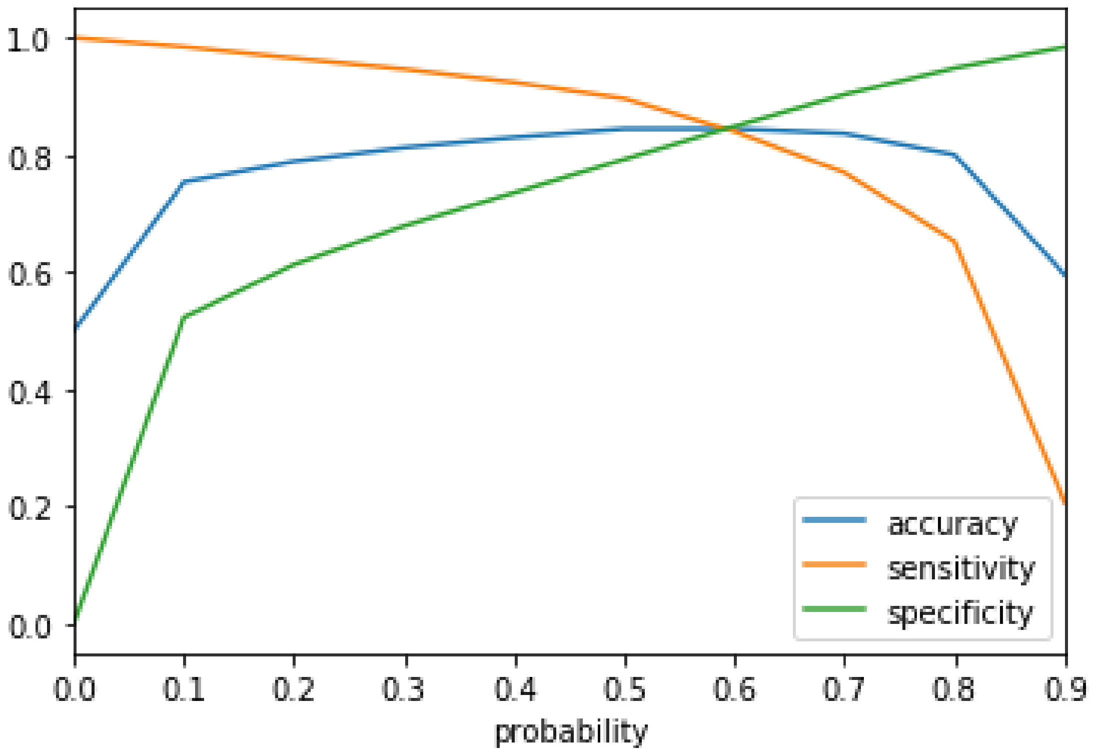
Model Building

- The initial step in our regression analysis entails partitioning the dataset into distinct training and testing subsets. We have opted for a partition ratio of 70:30, where 70% of the data is allocated to the training set.
- Following the data split, we employ Recursive Feature Elimination (RFE) as a feature selection technique. RFE is executed with the objective of retaining the 15 most relevant variables.
- Subsequently, we construct the regression model by iteratively eliminating variables. All VIFs are < 2.0 and p values are all variables are 0.
- We apply this model to make predictions on the test dataset, ultimately achieving an overall predictive accuracy of 78.48%.



Plotting the ROC Curve

- It appears that the ROC curve's area closely approximates a value of 1, indicative of the model's Gini coefficient.



Finding the optimal Cut point

- Accuracy stabilizes around 0.6.
- Sensitivity decreases as the probability increases.
- Specificity increases as the probability rises.
- At the point where these three parameters intersect around 0.6, a balance is observed between sensitivity and specificity, resulting in good accuracy.
- Despite the optimal probability cutoff at 0.6, the primary objective is to achieve higher sensitivity over accuracy and specificity. Therefore, although 0.6 is suggested by the curve, a probability cutoff of *0.5* is being chosen to prioritize higher sensitivity, aligning with our primary goal.

Recommendation

- Target customers exhibiting decreased usage in specific call types during the action phase, particularly in August:
- Prioritize those with reduced incoming local calls and outgoing ISD calls.
- Focus on customers showing lower charges for outgoing others in July and reduced incoming others in August.
- Customers experiencing increased value-based costs during the action phase are potential targets for tailored offers.
- Identify customers with higher monthly 3G recharge in August, indicative of potential churn.
- Pay attention to customers with decreasing STD incoming minutes for T-operator fixed lines in August.
- Customers displaying reduced monthly 2G usage in August are more prone to churn.
- Highlight those with declining incoming minutes for T-operator fixed lines in August.
- Customers with rising outgoing roaming minutes are more likely to churn, indicated by the positive coefficients of the roam_og_mou_8 variables.


Thank You