Telecom Customer Churn Classification Case Study

Data – Driven Insights and Model – Based Strategies

Submitted by – Sini Keloth, Rupam K

Business Problem Statement

In the telecom industry, customers have the freedom to choose between various service providers and frequently switch from one to another. This intense competition results in an average annual churn rate of 15-25%. Since acquiring a new customer is 5-10 times more expensive than retaining an existing one, customer retention has become a higher priority than acquisition.

For many established telecom operators, keeping highly profitable customers is the top business objective.
To minimize churn, telecom companies need to predict which customers are at high risk of churn.

The primary goal is to analyze customer-level data from a leading telecom company, develop predictive models to identify high-risk churn customers, and determine the key factors driving customer churn.

Telecom Customer Churn Prediction

Customer churn is a major challenge for the telecom industry, especially in competitive markets like India and Southeast Asia. This project focuses on building predictive models to identify high-risk customers and understand key drivers of churn.

Business Objectives:

1 **Predict Churn**

Develop models to forecast which high-value customers are likely to churn in the near future, enabling timely intervention strategies.

2 **Identify Churn Drivers**

Analyze significant predictor variables that contribute to churn, providing insights into customer behavior and potential areas for improvement.

Analysis Approach

The analysis approach involved a comprehensive process of data collection, preprocessing, feature engineering, model selection, and evaluation. The focus was on leveraging machine learning techniques to build a predictive model that could accurately assign scores to leads based on their likelihood of conversion.

1. Data Collection & Preprocessing:
Thoroughly analyzed the dataset to understand its structure and content. Clean the provided leads dataset, addressing any null values and filled the missing values appropriately. This step is crucial for ensuring the data's integrity.

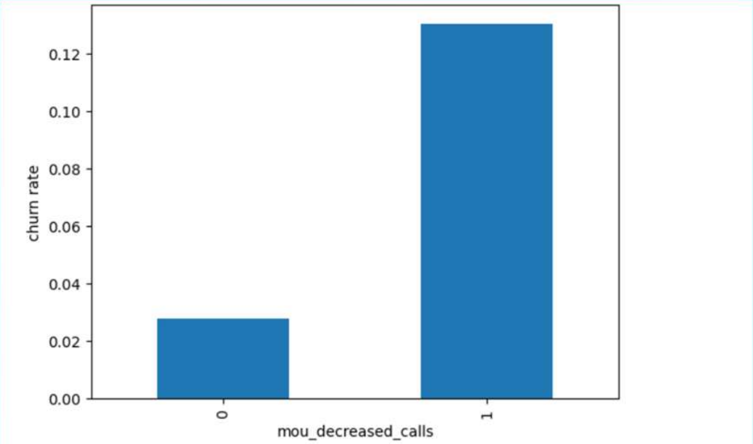
2. Exploratory Data Analysis (EDA)
Conduct an EDA to understand the relationships between various attributes (Lead Source, Total Time Spent, Last Activity, etc.) and the lead conversion outcome. Visualizations will help in interpreting these relationships.

3. Handling Class Imbalance
Class imbalance refers to a situation in classification problems where one class is significantly more prevalent than others. In other words, the dataset contains a disproportionate number of instances for each class.

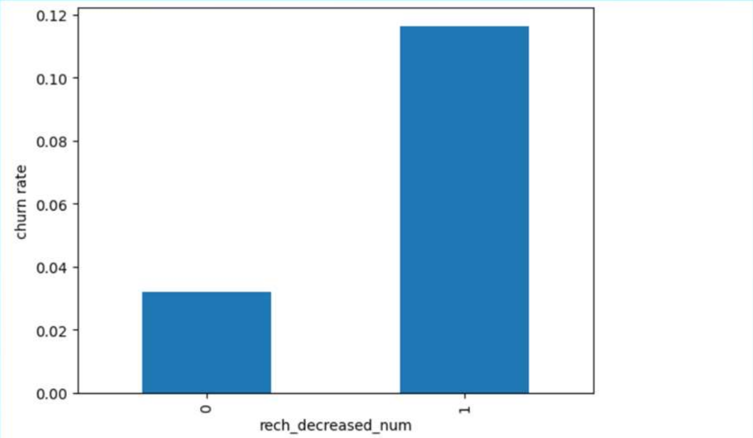
4. Model Evaluation and Tuning
Evaluate the performance of the model using metrics like accuracy, precision, and recall. Tune the model to improve its performance.

Key Insights from EDA

Exploratory data analysis (EDA) revealed valuable insights into the characteristics of high-churning rate of customers. Analyzing data patterns and trends helped identify key features that contributed to Churn.

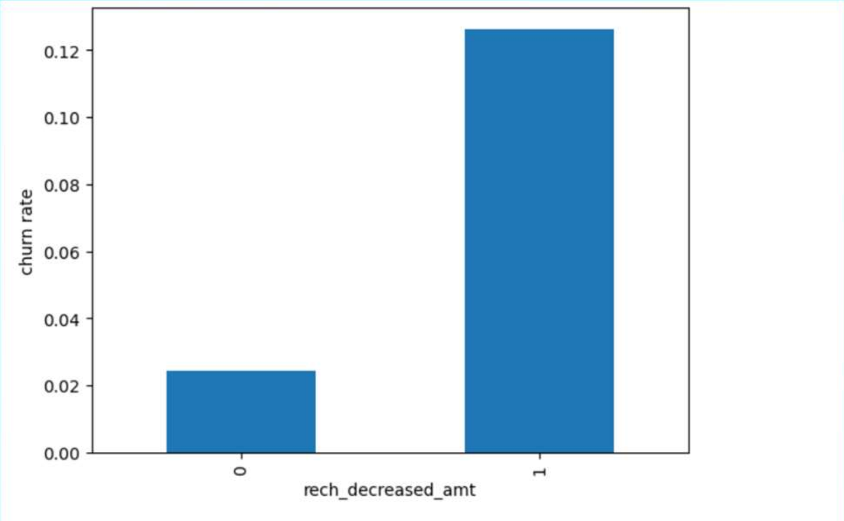


We can see that Churn rate is higher for cutomers whose mou decreased in action phase



Churn rate is higher for customers who has decreased in number of recharge in action phase

Key Insights from EDA

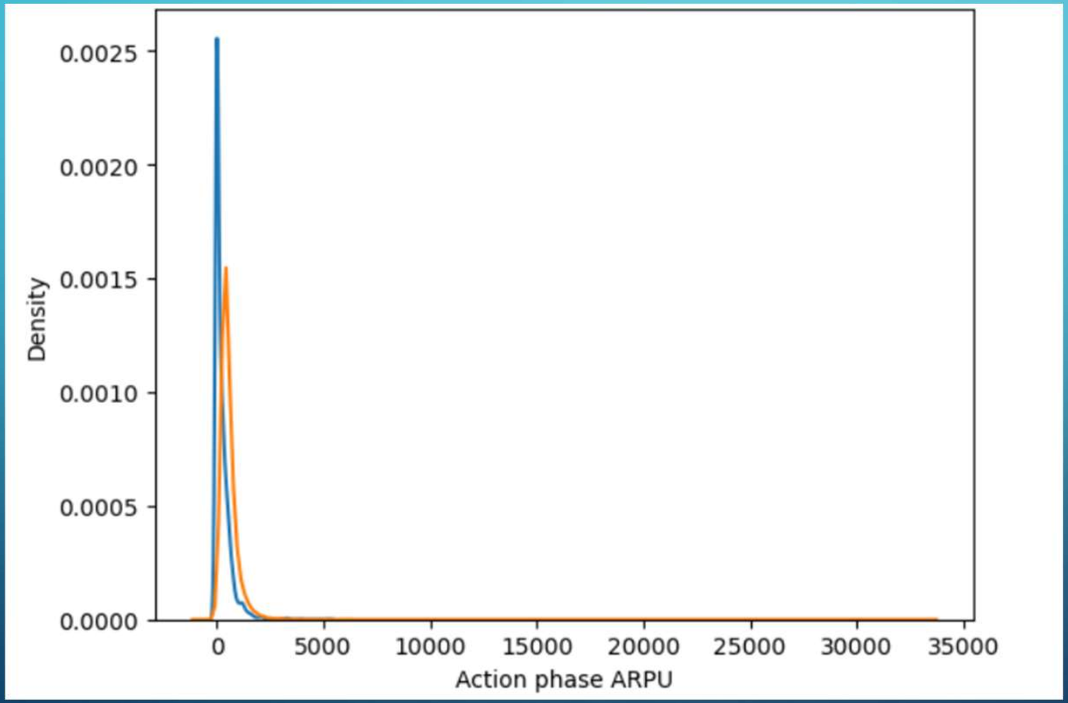


Churn rate of customers is higher whose amount of recharge has reducted in action phase.

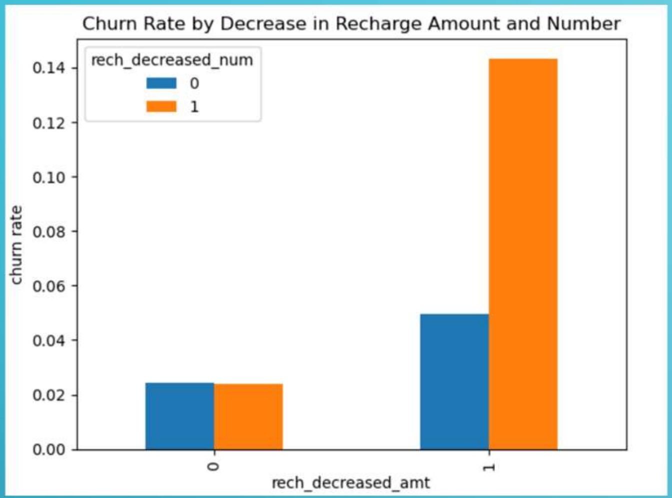
The analysis indicates a high churn rate among customers, particularly during the action phase in August. This trend is evident from the plotted graph, which shows a significant decline in minutes of usage, the number of recharges, and the total recharge amount.

Key Insights from EDA

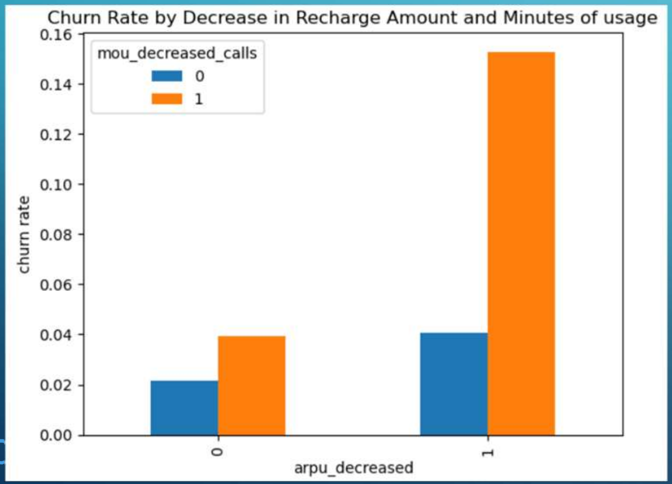
Average revenue per user (ARPU) for the churned customers is mostly densed on the 0 to 900. The higher ARPU customers are less likely to be churned.



Bivariate Analysis

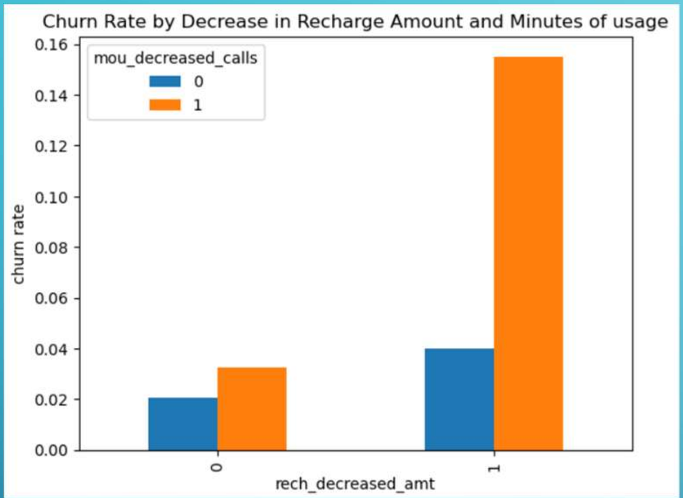


The plot indicates that customers who experienced a decrease in both their recharge amount and the number of recharges during the action phase have a higher churn rate compared to those in the good phase.



The plot indicates that customers who experienced a decrease in both their recharge amount and the Minutes of usage(MOU) during the action phase have a higher churn rate compared to those in the good phase.

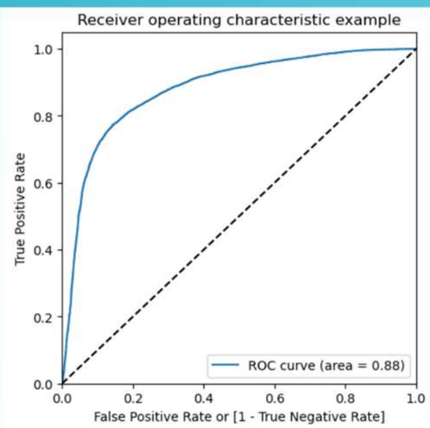
Bivariate Analysis



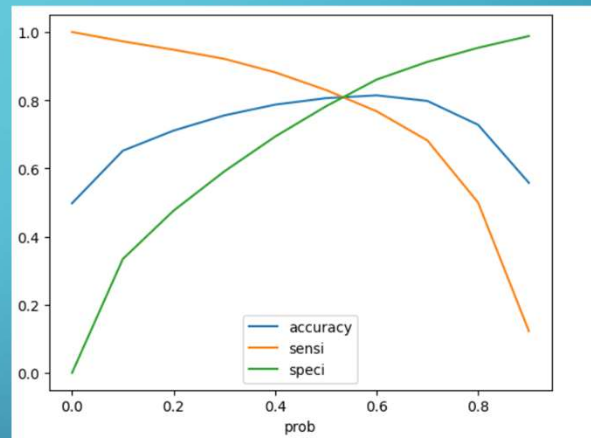
The plot indicates that customers who experienced a decrease in both their ARPU and the Minutes of usage(MOU) during the action phase have a higher churn rate compared to those in the good phase.

Model Performance

The Telecom churn model exhibited strong performance in predicting churn of customers. The ROC curve has a values of 0.88 indicating a good predictive model.



Generally ROC curve should be a value close to 1 to be a good model. Here we are getting a good value of 0.88 indicating a good predictive model.



The cutoff rate coming is 0.55 but we will take 0.40 to achieve higher sensitivity

Model Performance

1. Accuracy Score:
Training Data: 78.67%, Test Data: 72.03%
There is a difference between the training and test accuracy. The model performs better on the training set compared to the test set. This gap between training and test performance suggests the model might be overfitting the training data.

2. Sensitivity:
Training Data: 88%, Test Data: 87%
Sensitivity (or Recall) measures the proportion of actual positives (in this case, customers who churn) that your model correctly identifies. The sensitivity on the train and test data is approximately same. This suggests that the model generalizes well and performs better at identifying churners.

3. Specificity:
Training Data: 69%, Test Data: 70%
Specificity measures the proportion of actual negatives (non-churners) that are correctly identified by the model. It reflects the model's ability to avoid false positives—incorrectly classifying non-churners as churners. This suggests that the model's performance is consistent when it comes to identifying non-churners across both datasets.

Recommendation

1. Improve Customer Support: Ensure prompt and efficient customer service through multiple channels (phone, chat, social media). Training staff to handle issues empathetically can significantly boost customer satisfaction.
2. Personalized Communication: Tailor communication and offers based on customer preferences and usage patterns. Personalized outreach can strengthen customer relationships.
3. Monitor Usage Patterns: Use data analytics to track customer behavior and identify early warning signs of potential churn, such as a sudden drop in usage or complaints.
4. Proactive Outreach: Contact customers showing signs of disengagement and offer tailored solutions or incentives to re-engage them.
5. Adaptable Plans: Provide flexible pricing plans that cater to different usage patterns and budgets. Offering a variety of plans can help customers find the best fit for their needs.
6. Trial Periods for Upgrades: Allow customers to try out higher-tier plans for a limited time without commitment. This can entice them to upgrade if they find the new features beneficial.
7. Invest in Infrastructure: Ensure that network quality is consistent and coverage is extensive. Regularly assess and upgrade network capabilities to reduce service disruptions.
8. Collect Feedback: Actively seek feedback from customers regarding network performance and address issues promptly.