Telecom Churn Case Study

DS C68 Batch

Created by: Gowtham M, Tamal Sen, Raj Vala
Date: 07-Jan-2024

Upgrad Assignment

- **Problem Statement**
- **Steps followed**
- **Univariate Analysis**
- **Bivariate Analysis**
- **Baseline Models**
- **Model Building and Evaluation**
- **Conclusion**

Problem Statement

Problem Statement:

The goal of this project is to predict high-value telecom customers who are likely to churn, based on their usage patterns over a specified period. By identifying these customers early, the telecom company can take proactive measures to retain them, such as offering personalized discounts or improving service quality.

Objective:

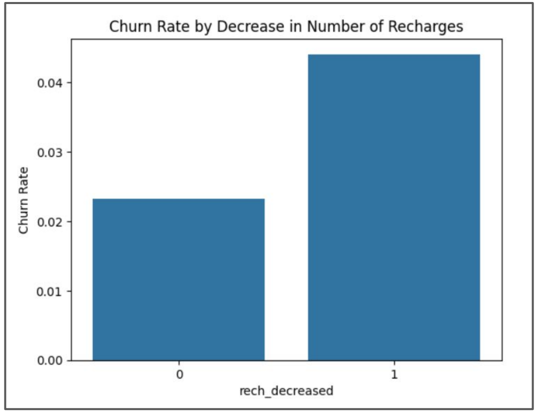
The objective is to build a predictive model that identifies high-value telecom customers likely to churn. By analyzing customer behavior, such as changes in usage, recharge amounts, and ARPU, the model aims to predict churn. The goal is to enable proactive retention strategies to reduce churn, enhance customer retention, and improve long-term business outcomes for the telecom company.

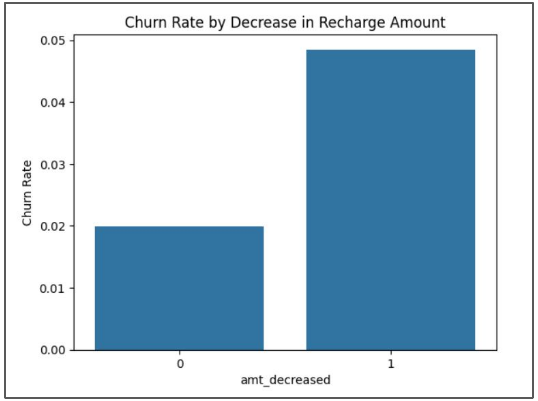
Steps followed

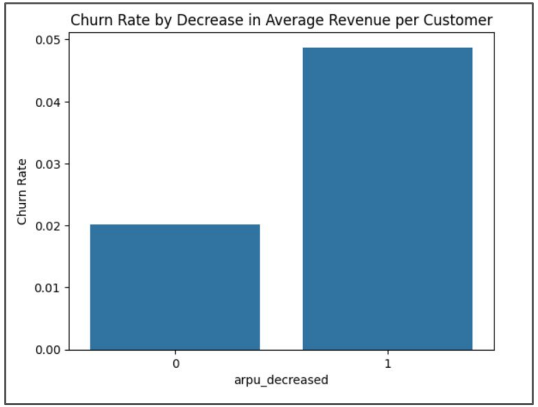
Steps followed for Analysis and Model Building

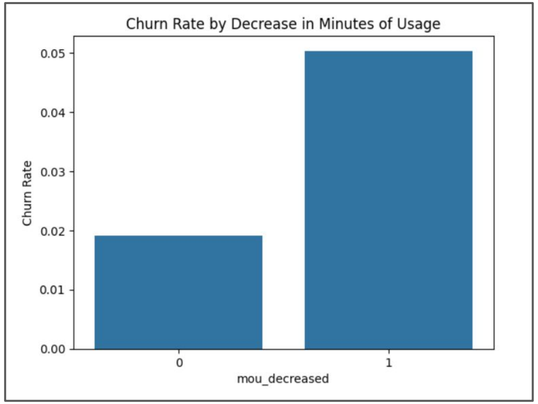
- Data ingestion, reading, understanding and visualising the data
- Preparing the data for modelling
  - Data cleaning
  - Missing values imputation
  - Deriving new variables
  - Scaling and transforming the data
  - Exploratory Data Analysis
- Building the model
  - Build the model using Logistic Regression
  - Build the model using Random Forest
- Evaluate the model
  - Evaluate against train data set
  - Evaluate against test data set
- Conclusion

Univariate Analysis





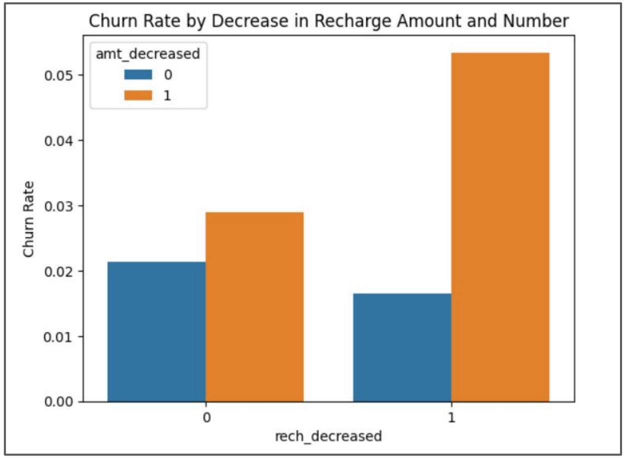


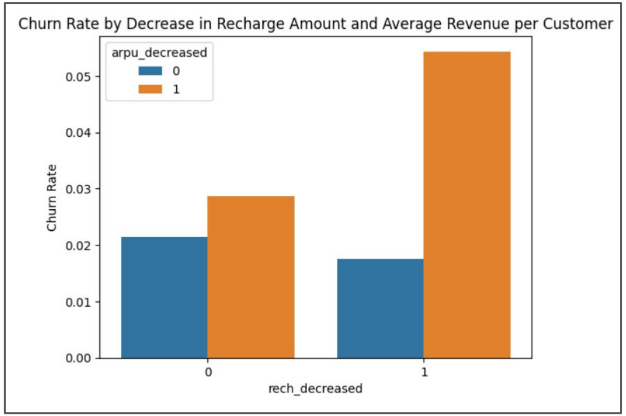


Findings:

- Customers who decreased their Minutes of Usage (MOU) in the action month have a higher churn rate.
- A decrease in the number of recharges made by customers correlates with an increase in churn rate.
- Customers who reduced their recharge amount in the action month tend to have a higher churn rate.
- A decrease in Average Revenue Per User (ARPU) leads to a higher churn rate among customers.

Bivariate Analysis





Findings:

- Churn Rate by Decrease in Recharge Amount and Number of Recharges:
  - Customers who experienced a decrease in both recharge amount and the number of recharges have a significantly higher churn rate compared to other groups.
- Churn Rate by Decrease in Recharge Amount and Average Revenue Per Customer (ARPU):
  - The churn rate is notably higher for customers whose recharge amount and Average Revenue Per User (ARPU) have both decreased.

Baseline Models

Logistic Regression:

We had created a baseline model which is performing decent, with good accuracy rate and precision, recall values. However, this model is only for baseline. We haven't done any feature engineering or feature selection till this point.

Random Forest:

The random forest model also has higher accuracy compared to regression model, but again the precision and recall values are very low for the test set. Also, we decided to continue with logistic regression and random forest models as the decision tree is the underlying structure of random forest.

Model Building and Evaluation

Model Building & Evaluation:

- The data was scaled before feeding into the model using Standard Scaler
- Created synthetic samples by doing upsampling using SMOTE(Synthetic Minority Oversampling Technique)
- For Linear regression we applied Recursive Feature Elimination to select the best columns
- Used the Random Search CV to hypertune the parameters and get the best hyper parameters to be selected for the Random Forest model
- After the model building is completed, we followed the below steps for model evaluation:
    - Evaluating the model against the test data set
    - Created the probability of prediction
    - Checked the accuracy, recall and precision, specificity and sensitivity metrics to check the model
    - Removed the columns with multicollinearity using the RFE and VIF elimination methods
    - Made the ROC-AUC curve to find the area under the curve to check model performance

Conclusion

Observations/Recommendations:

- 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. That means for the customers, whose roaming outgoing minutes of usage is increasing are more likely to churn.

Thank You!