TELECOM CHURN

BY-HARSHIT GOSAIN

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.
- In this project, you will analyze customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn.

STEPS OF THE PROJECT

The project consists of the following sections:

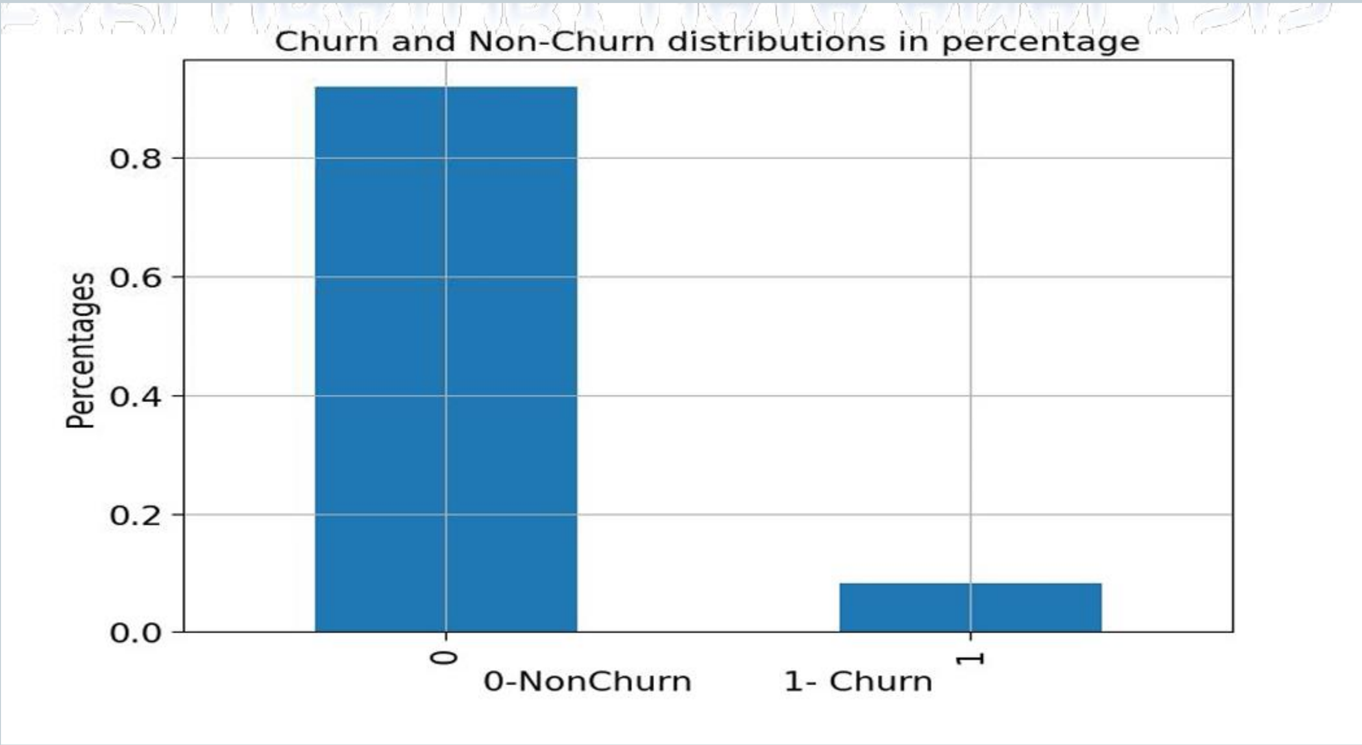
1. Data Reading
2. Exploratory Data Analysis and Data Cleaning
3. Data Visualization
4. Feature Importance
5. Feature Engineering
6. Setting a baseline
7. Splitting the data in training and testing sets
8. Assessing multiple algorithms
9. Algorithm selected: Gradient Boosting
10. Hyper parameter tuning
11. Performance of the model
12. Drawing conclusions – Summary

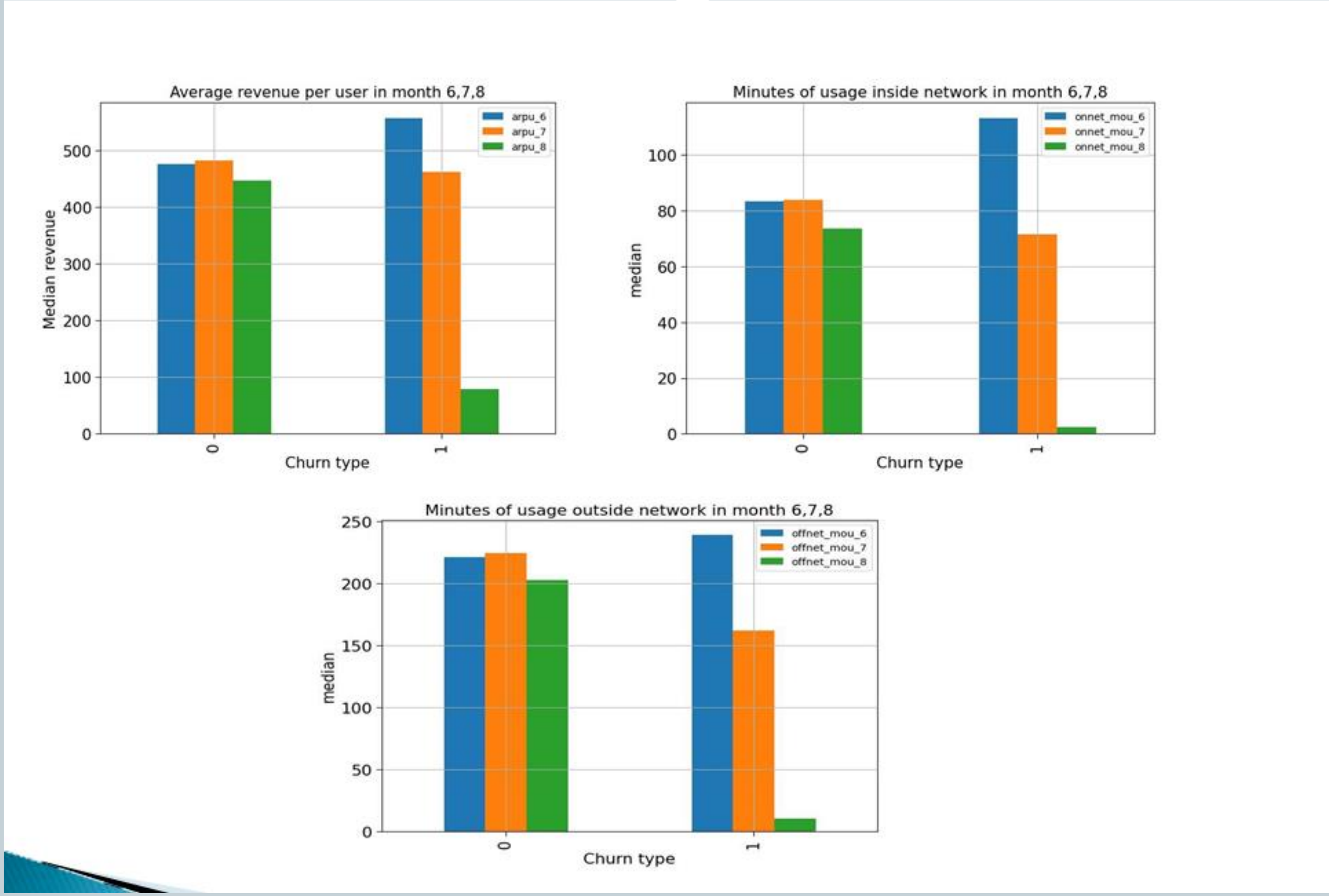
Data Reading and Data Cleaning

- Dataset contains 99999 no of rows.
- 226 no of columns.
- Number of Ïloat data type - 179
- Number of int data type - 35
- Number of object data type- 12
- we are left with 30,001 rows of records and 141 columns are available to explore after data cleaning.

Exploratory Data Analysis

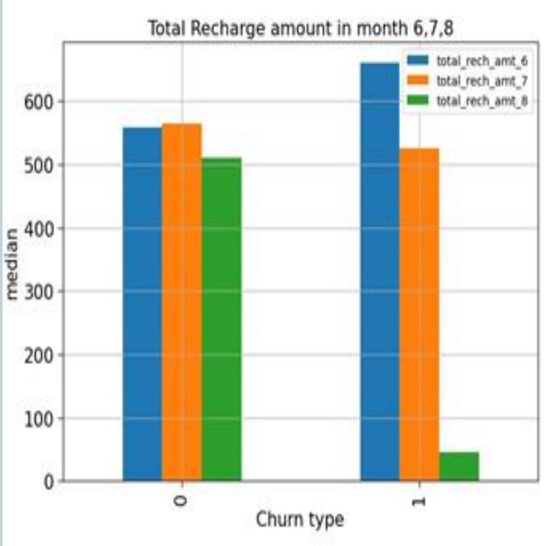
We have 92% customers belong non-churn and 8% customers belong to Churn type.

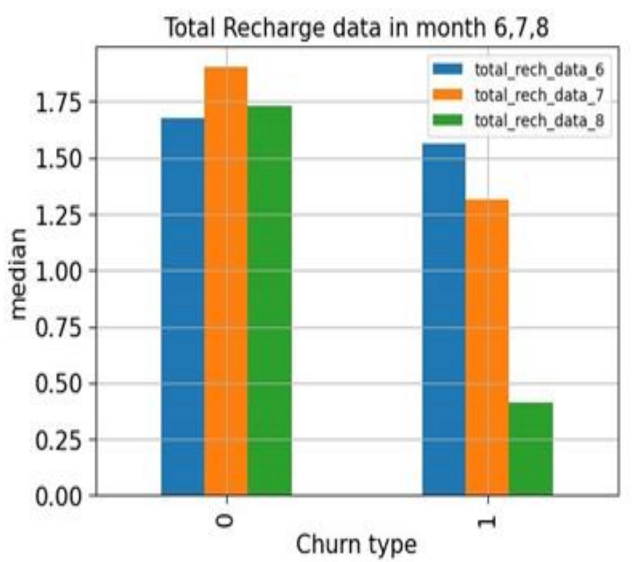




Average revenue per user in month 6,7,8
Median revenue
Churn type
arpu_6
arpu_7
arpu_8
Minutes of usage inside network in month 6,7,8
median
Churn type
onnet_mou_6
onnet_mou_7
onnet_mou_8
Minutes of usage outside network in month 6,7,8
median
Churn type
offnet_mou_6
offnet_mou_7
offnet_mou_8

- A higher average revenue per user in the sixth month suggests that if they are dissatisfied, these users are more inclined to churn.
- Users exhibiting increased minutes of usage in the sixth month are more prone to churn.
- Users displaying a significant disparity in call duration to other networks between the sixth and seventh months are likely candidates for churn.
- A greater disparity in total recharge amounts indicates a higher likelihood of churn.
- Users who haven't recharged in the sixth, seventh, and eighth months may or may not churn, as there isn't substantial evidence from the data.





MODEL BUILDING

- We will explore below models.

•Logistic regression

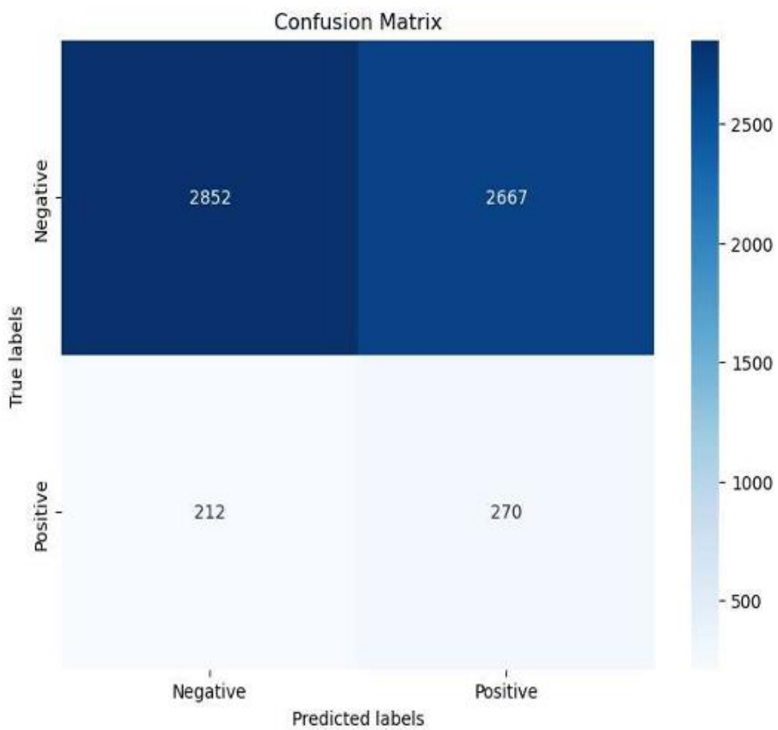
•Decision tree

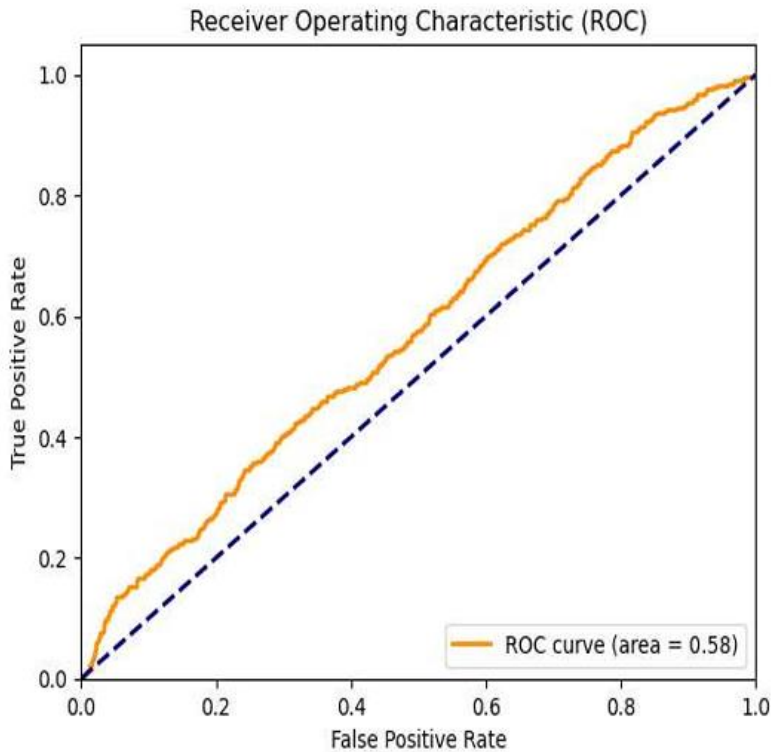
•Randomforest

•Gradientboosting

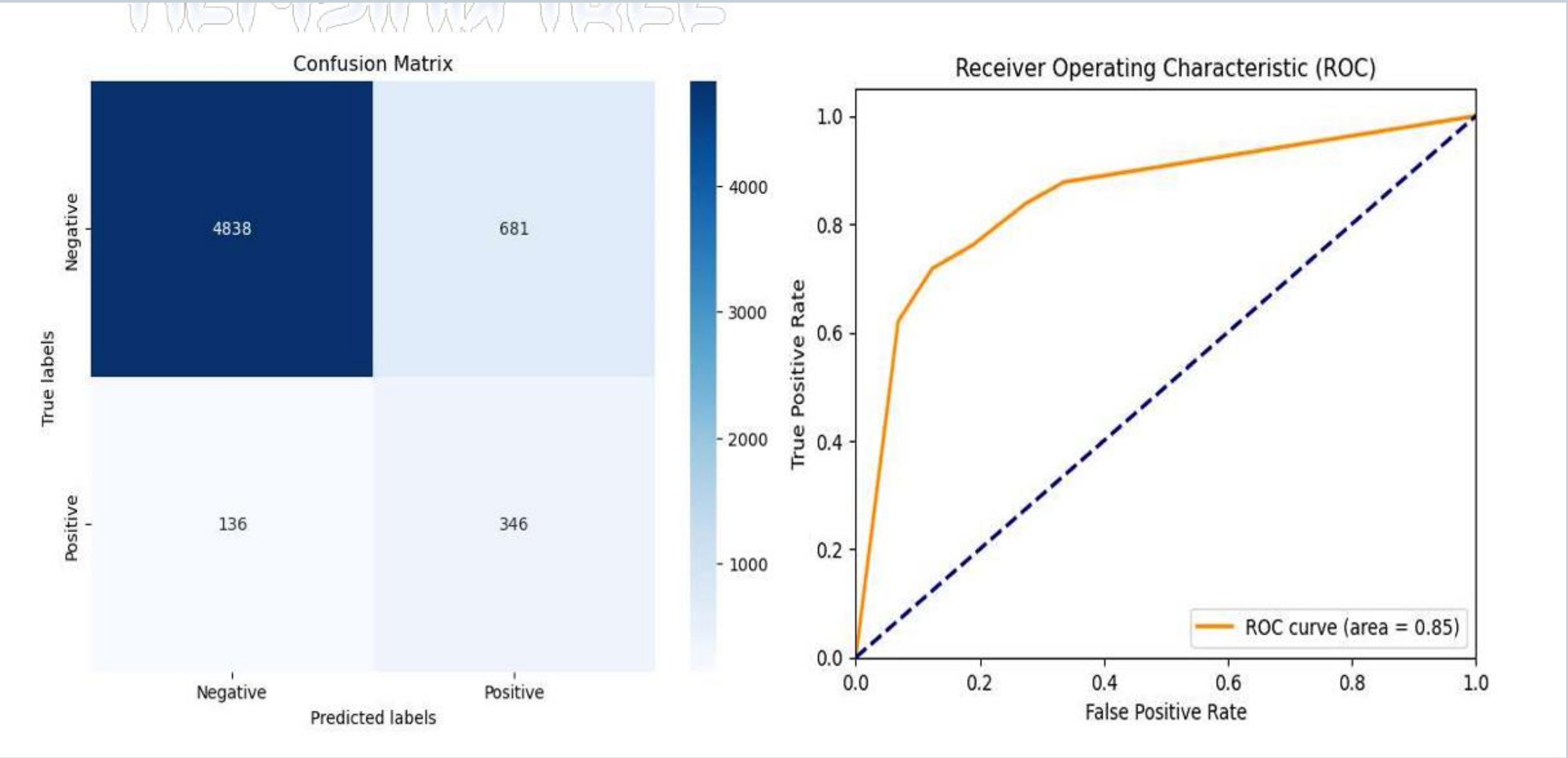
•XGboost

Logistic regression

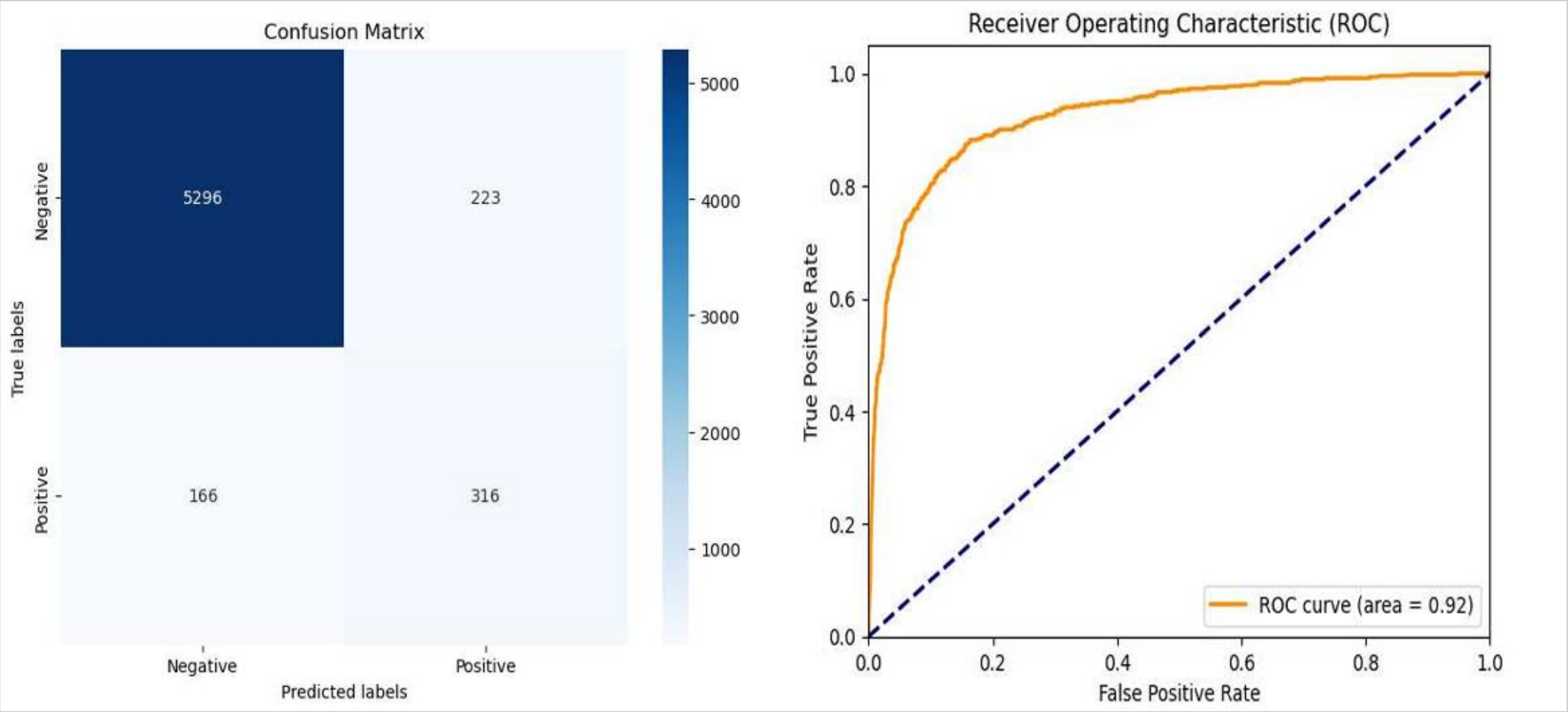




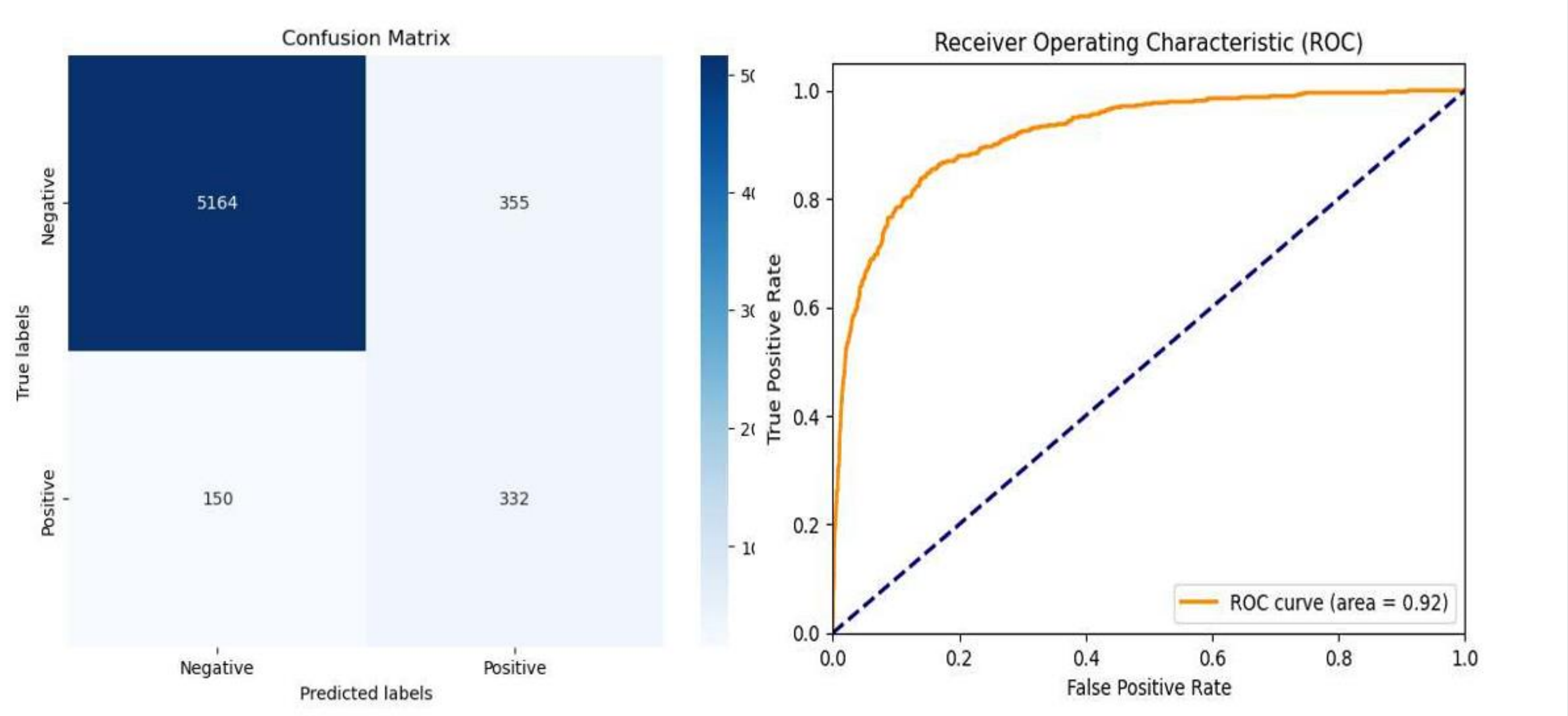
Decision tree



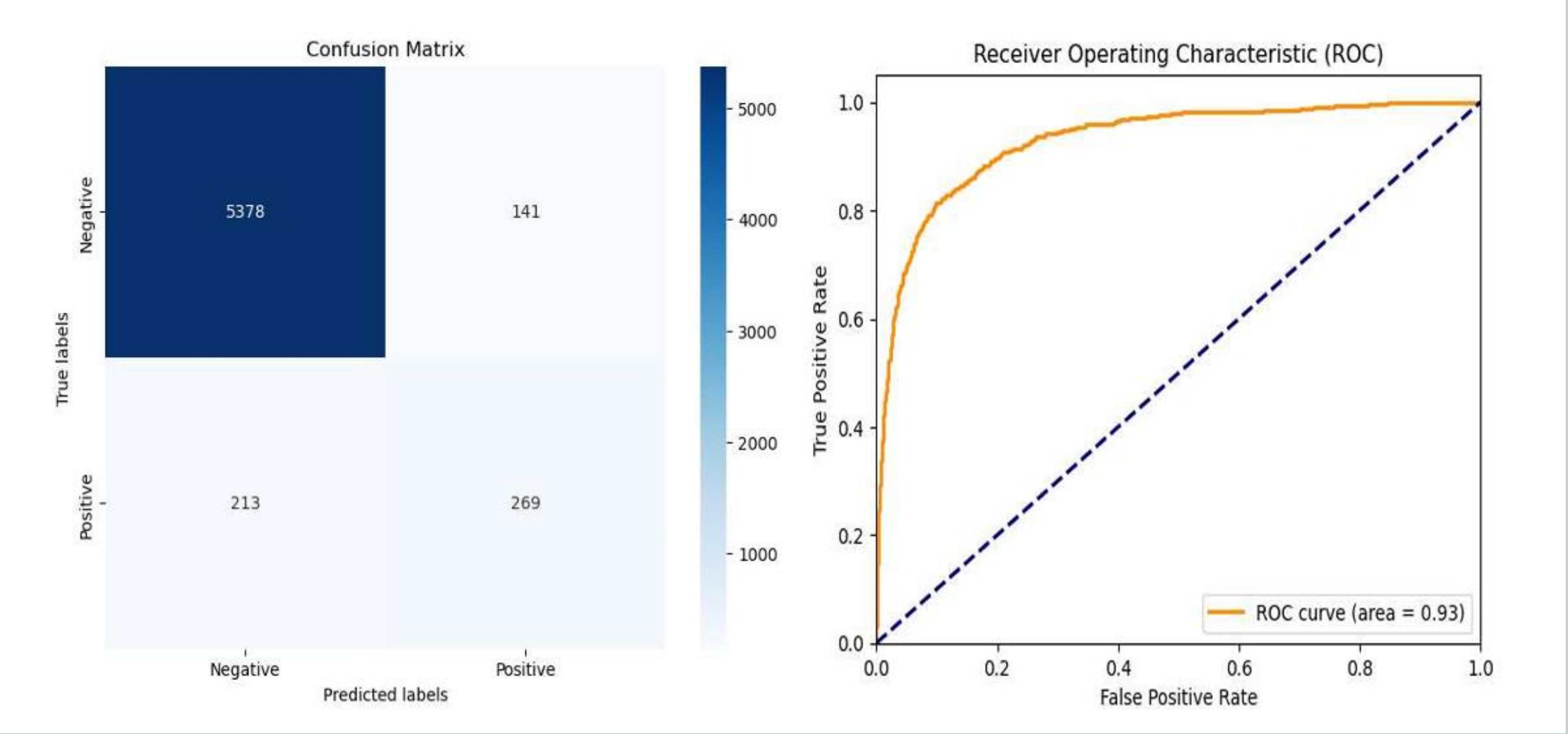
Randomforest



Gradientboosting



XGboost

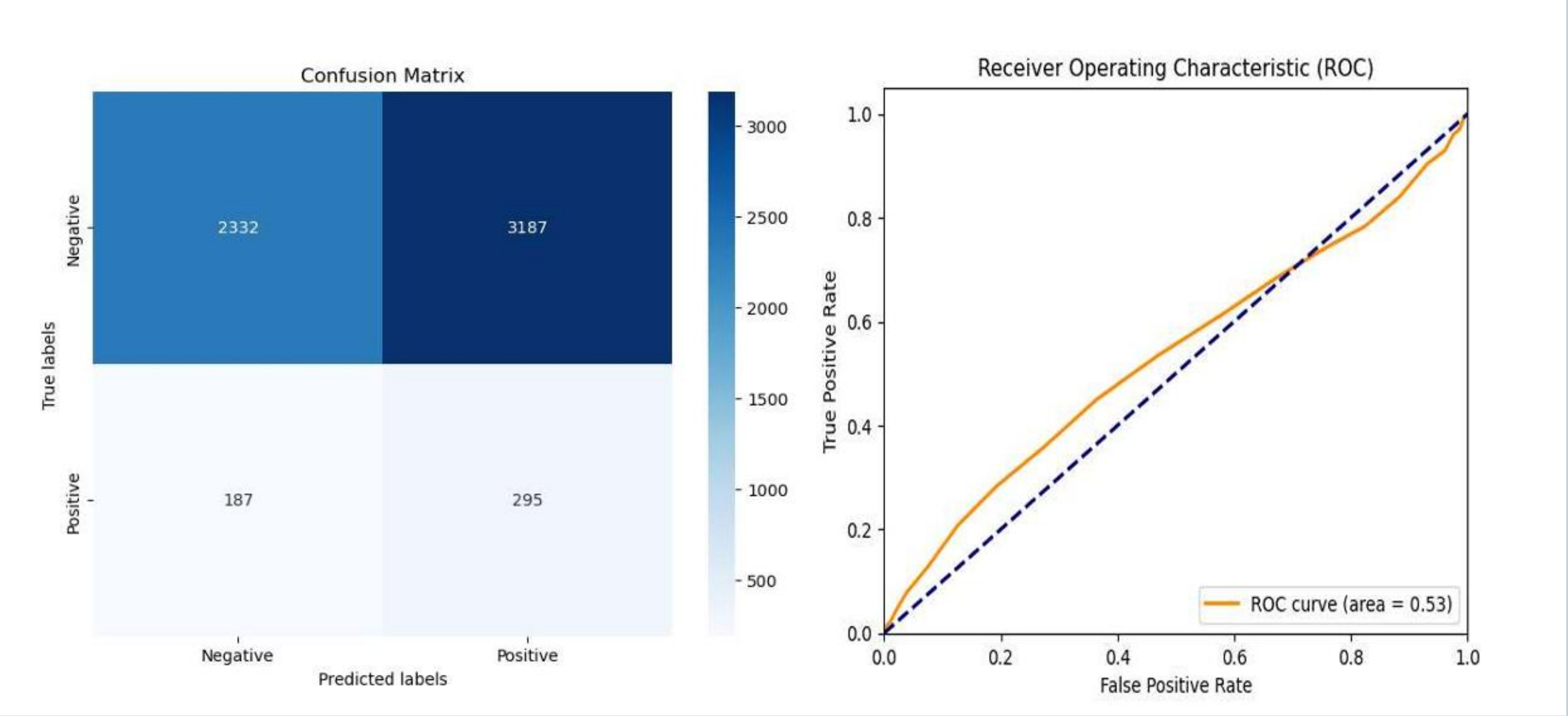


Confusion Matrix
True labels
Negative
Positive
5378
141
213
269
Negative
Positive
Predicted labels
5000
4000
3000
2000
1000
Receiver Operating Characteristic (ROC)
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.93)

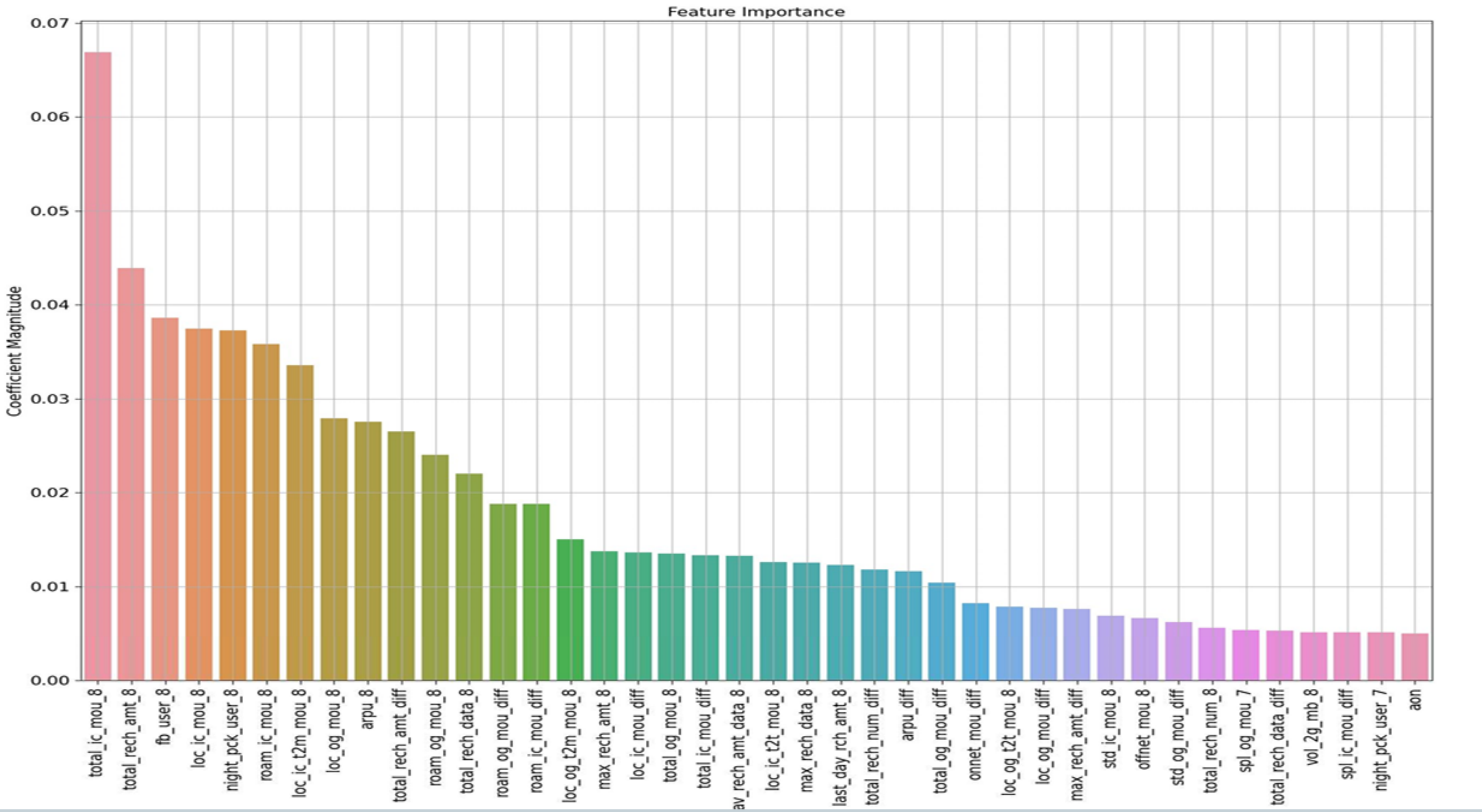
- - Random forest performed effectively in predicting churn on this dataset, achieving precision near 59%, recall near 65%, and an f1_score close to 61%.
- - Logistic regression was employed with Principal Component Analysis (PCA) in this context.
- - Notably, in this scenario, the model performed well without utilizing PCA.

	Model	precision	recall	f1_score	roc_auc
0	LogisticRegression	0.091931	0.560166	0.157941	0.575711
0	DecisionTree	0.336904	0.717842	0.458582	0.851062
0	RandomForest	0.586271	0.655602	0.619001	0.924402
0	GradientBoosting	0.483261	0.688797	0.568007	0.919795
0	XGBoost	0.656098	0.558091	0.603139	0.929580

FEATURE IMPORTANCE AND MODEL INTERPRETATION



FEATURE IMPORTANCE



CONCLUSION

- The graph above highlights the most significant features.
- Higher average revenue per user suggests a likelihood of churn, particularly if users are dissatisfied with the network.
- The duration of local calls usage also influences churn.
- Notably, a substantial difference in recharge amounts between the sixth and seventh months impacts churn.
- Users exhibiting increased Roaming in both outgoing and incoming calls are also prone to churn, warranting the company's attention.