Customer Churn Prediction in the Telecom Industry

- PRESENTED BY:

  - Vidhi Shukla
  - Unni Krishnan
  - Vamshi Krishna Vallabhaneni

- DATE: 10 FEBRUARY 2025

Introduction

Problem Statement:

Customer churn is a critical challenge in the telecom industry. Predicting churn helps businesses retain high-value customers.

Objective:

Build a machine learning model to predict churn and provide insights to reduce churn rates.

Data Understanding

Dataset:

The dataset contains customer data with metrics like ARPU, data usage, and recharge history.

Sample features include:

1. Total Recharge Amount
2. Data Usage
3. Call Usage
4. Recharge Counts
5. Customer Last month Recharge.
6. Total Monthly Charges
7. VBC(Volume based count)
8. Average Revenue per User & Total Data Usage

Exploratory Data Analysis (EDA)
Churn Distribution & Recharge Amount Distribution

Churn Distribution:

Show the distribution of churners vs non-churners.

Key Insights:

1. High ARPU customers tend to churn less.
2. Higher recharge drop correlates with churn.

Feature Engineering

High-Value Customer Filter: Filtered customers based on recharge amount.

Derived Features:

1. Recharge Drop
2. Call Drop
3. Data Usage Drop

These features help in identifying customer churn patterns.

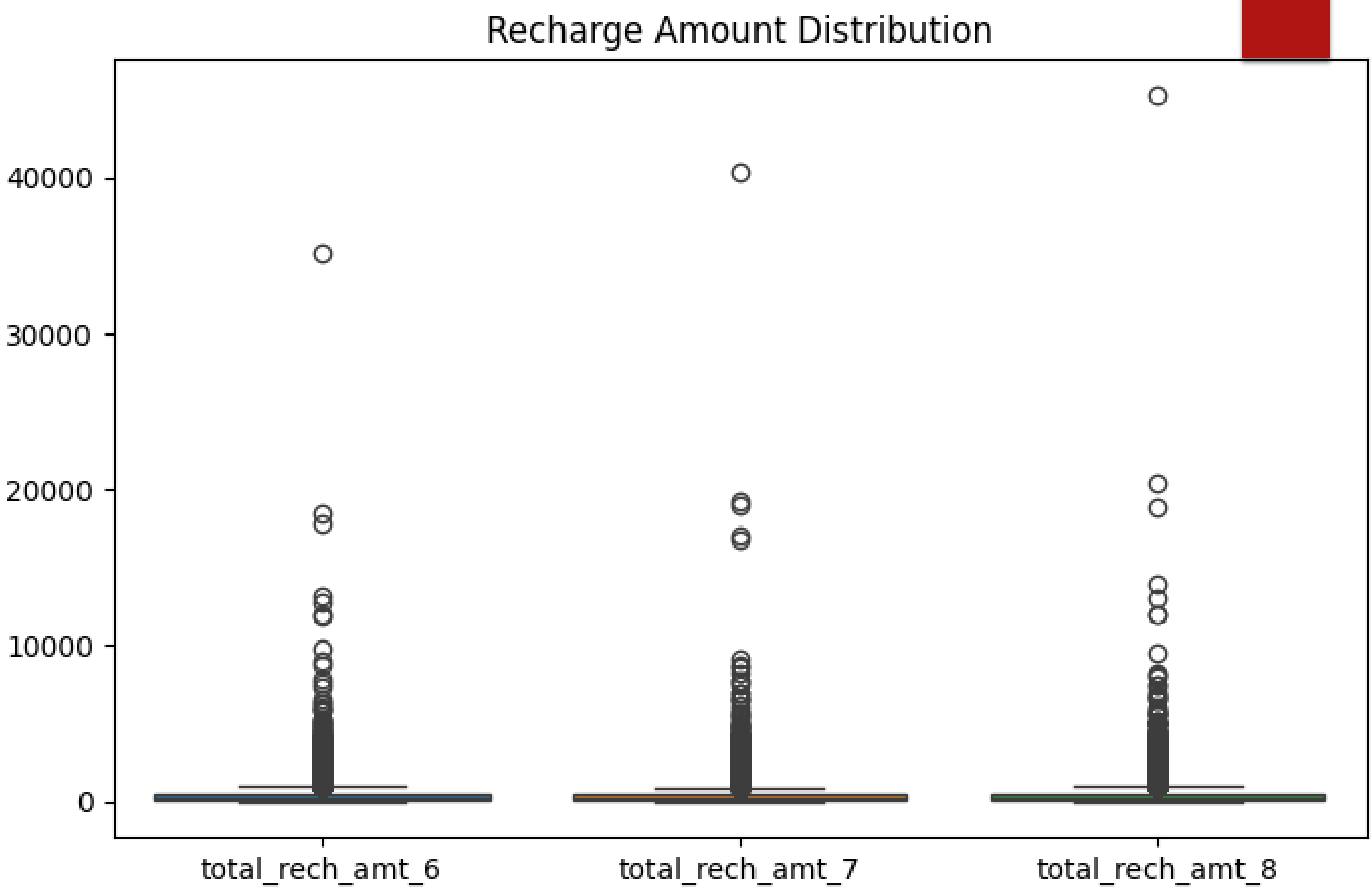
Like: High Value Customers

New Features :

Recharge Drop

Call Drop

Data usage Drop



Recharge Amount Distribution
40000
30000
20000
10000
0
total_rech_amt_6
total_rech_amt_7
total_rech_amt_8

Modeling Process

Models Used:

1. Logistic Regression
2. Random Forest
3. XGBoost

Train-Test Split: 80-20

Handling Imbalance: Applied SMOTE to balance the data.

Evaluating Logistic Regression Model

Model Performance:

	precision	recall	f1-score	support
0	0.97	0.90	0.94	5484
1	0.41	0.73	0.52	519
accuracy			0.89	6003
macro avg	0.69	0.81	0.73	6003
weighted avg	0.92	0.89	0.90	6003

ROC-AUC Score: 0.8895785813766866

Confusion Matrix:

```
 [[4940  544]
 [ 141  378]]
```

Evaluating Random Forest Model

Model Performance:

	precision	recall	f1-score	support
0	0.97	0.97	0.97	5484
1	0.63	0.63	0.63	519
accuracy			0.94	6003
macro avg	0.80	0.80	0.80	6003
weighted avg	0.94	0.94	0.94	6003

ROC-AUC Score: 0.9322082878340072

Confusion Matrix:

```
 [[5293  191]
 [ 190  329]]
```

Evaluating Random Forest Model

Model Performance:

	precision	recall	f1-score	support
0	0.97	0.97	0.97	5484
1	0.63	0.63	0.63	519
accuracy			0.94	6003
macro avg	0.80	0.80	0.80	6003
weighted avg	0.94	0.94	0.94	6003

ROC-AUC Score: 0.9322082878340072

Confusion Matrix:

```
 [[5293  191]
 [ 190  329]]
```

Model Evaluation

- Metrics Evaluated:
  - 1. Accuracy
  - 2. Precision
  - 3. Recall
  - 4. F1-Score

- Random Forest Model performed the best with the highest ROC-AUC score.

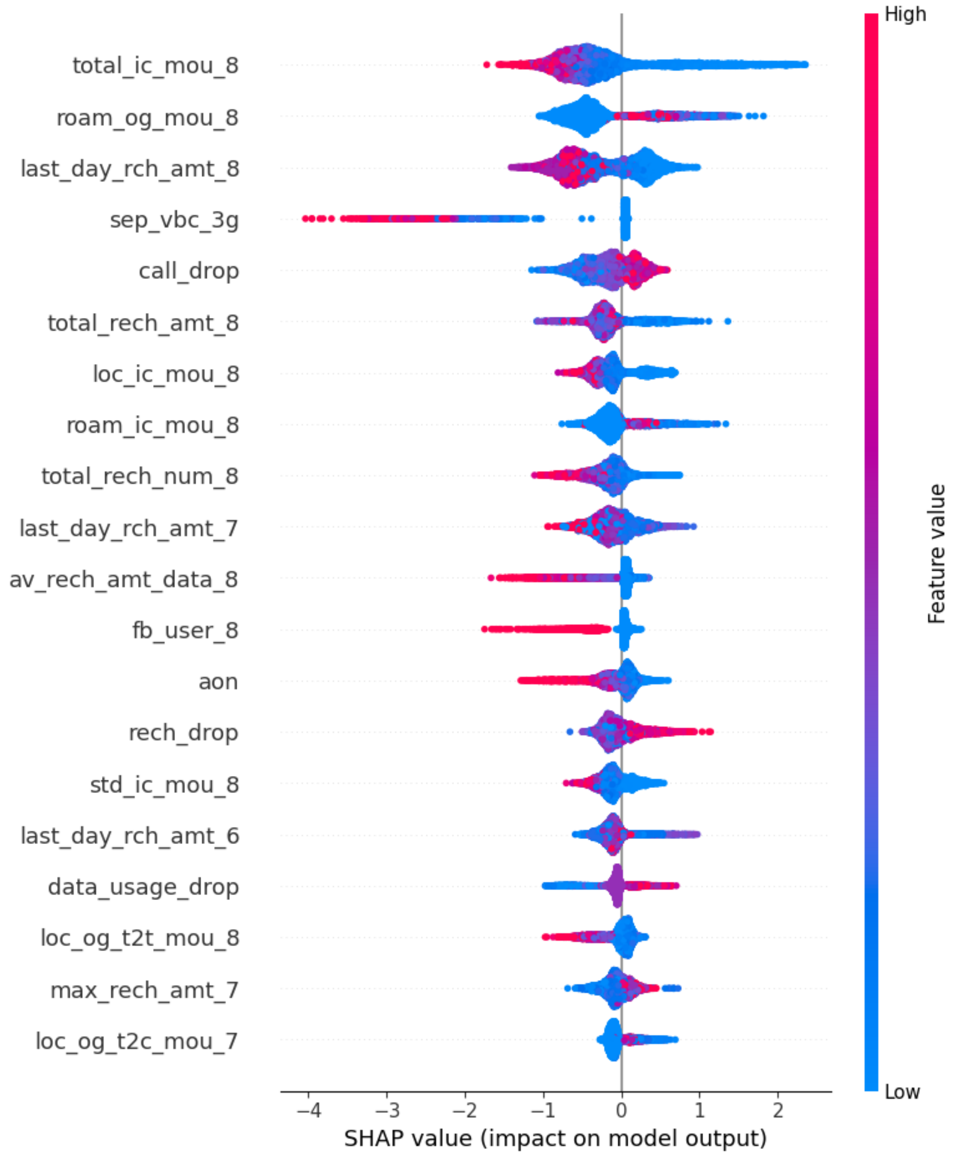
```
Best Parameters for Random Forest: {'max_depth': 30, 'min_samples_split': 2, 'n_estimators': 200}
```

```
Final Model Performance:
              precision    recall  f1-score   support

           0       0.97      0.96      0.96      5484
           1       0.62      0.63      0.63       519

    accuracy                           0.94      6003
   macro avg       0.79      0.80      0.80      6003
weighted avg       0.94      0.94      0.94      6003

ROC-AUC Score: 0.9327339016708617
Confusion Matrix:
 [[5285  199]
 [ 191  328]]
```



total_ic_mou_8
roam_og_mou_8
last_day_rch_amt_8
sep_vbc_3g
call_drop
total_rech_amt_8
loc_ic_mou_8
roam_ic_mou_8
total_rech_num_8
last_day_rch_amt_7
av_rech_amt_data_8
fb_user_8
aon
rech_drop
std_ic_mou_8
last_day_rch_amt_6
data_usage_drop
loc_og_t2t_mou_8
max_rech_amt_7
loc_og_t2c_mou_7
−4
−3
−2
−1
0
1
2
SHAP value (impact on model output)
High
Low
Feature value

Feature Importance & Recommendations

Key Features Impacting Churn:

1. Recharge Drop
2. Data Usage Drop
3. AON (Age on Network)

Business Recommendations:

1. Focus on high-value customers with declining data and recharge usage.
2. Implement targeted retention campaigns.

Conclusion

- Conclusion:

The model successfully predicts customer churn with key indicators such as recharge drop and data usage decline. Implementing retention strategies based on these insights can reduce churn.