



# Telecom Churn Project - Churn Prediction Analysis

# Project Overview

## **Objective**

To identify high-value customers at risk of churn and provide actionable insights to reduce churn rates. The business objective is to predict the churn in the last (i.e. the ninth) month using the data (features) from the first three months.

## **Dataset**

The dataset contains customer-level information for a span of four consecutive months - June, July, August and September.

## **Key Steps**

- Data preprocessing and feature engineering.
- Model development using Logistic Regression.
- Evaluation and recommendation of retention strategies.

# Data Overview and Preparation

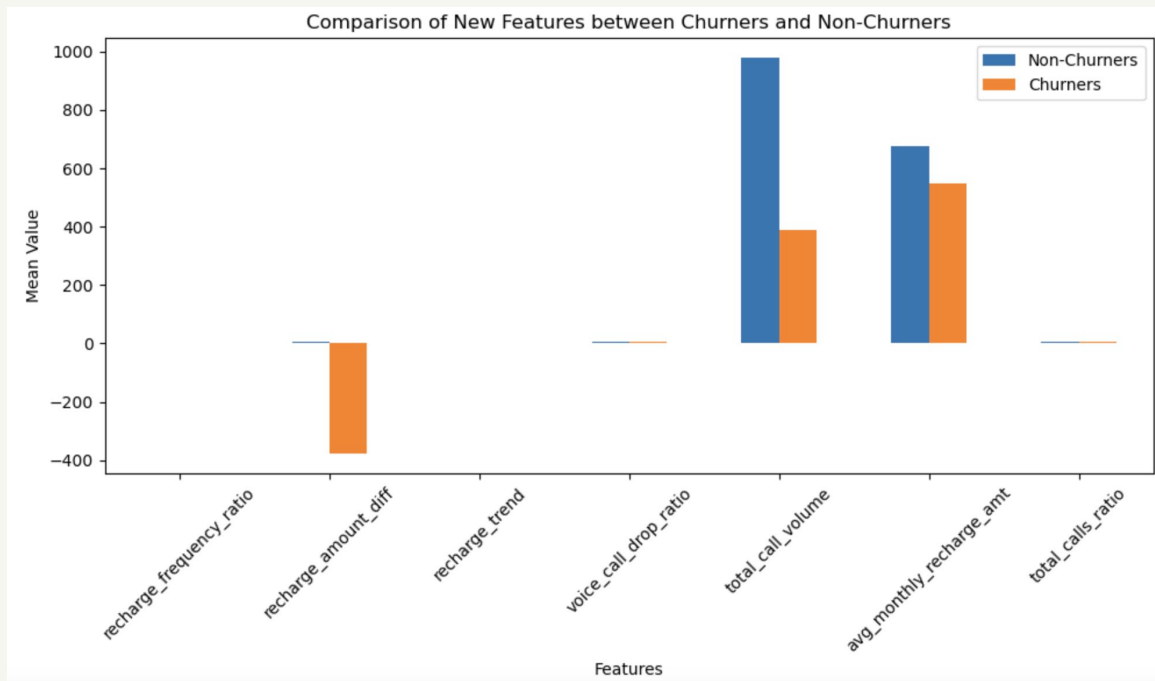
## Dataset Details

- Initial shape: (99999 rows, 226 columns)
- Final shape (after cleaning, feature engineering): (30024 rows, 136 columns)
- Key features included:
  - Recharge Amount, call duration, data usage.
  - Mobile Number, circle ID, Date columns (removed).
- Outliers handled and imputation performed for missing values

## Feature Engineering

- Added features like recharge frequency ratio, recharge amount difference, recharge trend, total call volume, average monthly recharge amount & total calls ratio.

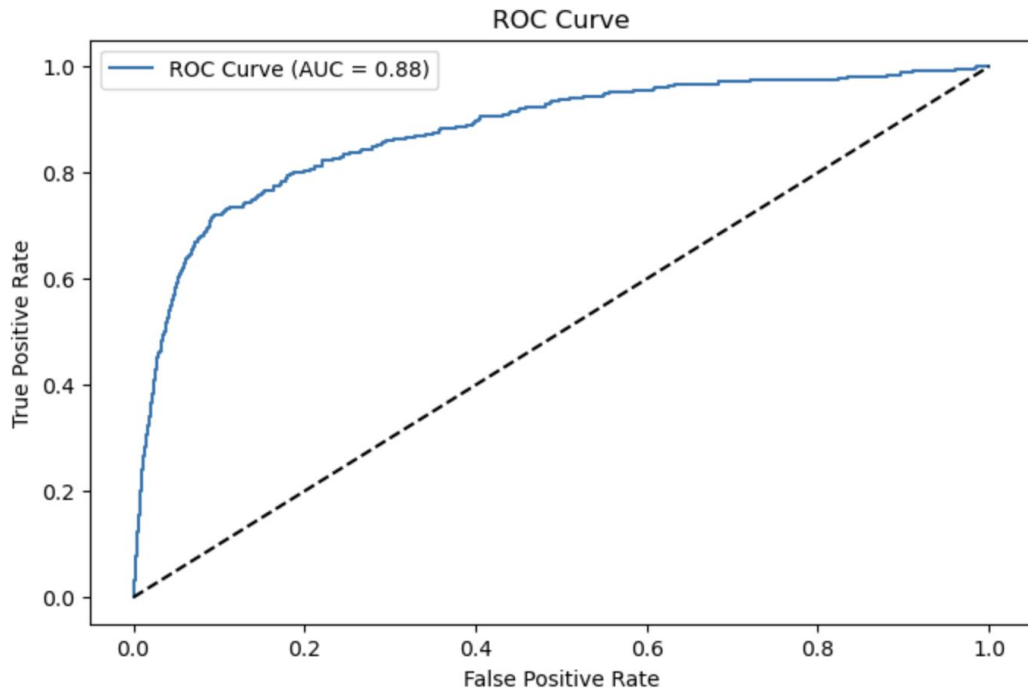
# EDA Analysis



The comparison of new features reveals significant differences between churners and non-churners. Churners exhibit lower total call volumes and average monthly recharge amounts, while showing distinct patterns in recharge frequency and call drop ratios, indicating their potential as predictive factors for churn.

# Model Development

ROC-AUC Score: 0.88



**Algorithm:** Logistic Regression

**Key Metrics for Evaluation:**

- ROC-AUC Score: 0.88
- Precision-Recall Curve:  
Focus on churners to  
assess model utility for  
targeted retention  
campaigns.

# Model Performance

## Classification Report:

	precision	recall	f1-score	support
0	0.98	0.84	0.90	5174
1	0.24	0.77	0.37	353
accuracy			0.83	5527
macro avg	0.61	0.80	0.64	5527
weighted avg	0.93	0.83	0.87	5527

## Confusion Matrix:

```
[[4335  839]
 [  82 271]]
```

# Model Evaluation on Train and Test Data

Train Set Evaluation:

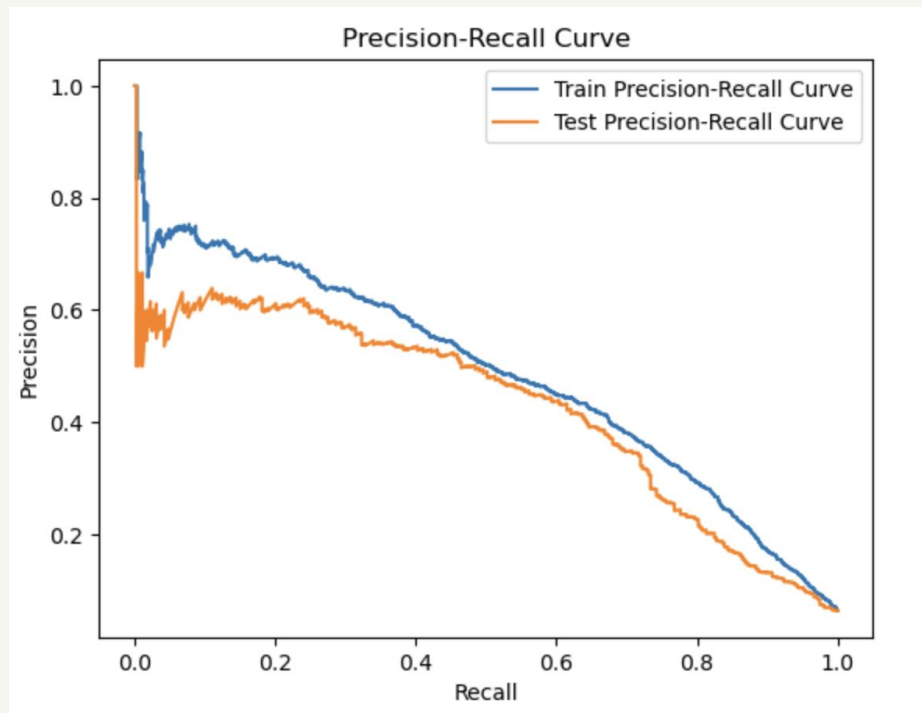
	precision	recall	f1-score	support
0	0.99	0.84	0.91	20691
1	0.27	0.83	0.40	1413
accuracy			0.84	22104
macro avg	0.63	0.84	0.66	22104
weighted avg	0.94	0.84	0.88	22104

Train ROC-AUC Score: 0.90

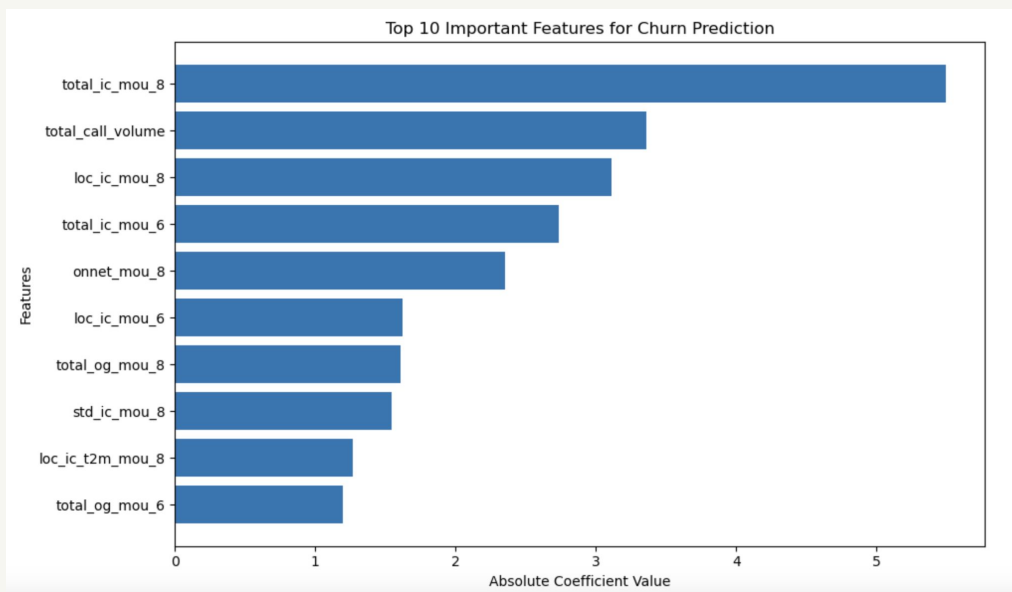
Test Set Evaluation:

	precision	recall	f1-score	support
0	0.98	0.84	0.90	5174
1	0.24	0.77	0.37	353
accuracy			0.83	5527
macro avg	0.61	0.80	0.64	5527
weighted avg	0.93	0.83	0.87	5527

Test ROC-AUC Score: 0.88



# Feature Importance



Top 3 features contributing to churn prediction:

- Total Incoming Voice Calls - Minutes\_August
- Total Call Volume
- Local Incoming Calls - Minutes\_August



# Recommendations

## Strategies to Manage Churn:

- **Targeted Retention Campaigns:** Focus on customers with high churn probability.
- **Optimize Recharge Plans:** Tailor plans for customers with declining recharge trends.
- **Improve Call Quality:** Address dropped calls for churn-prone users.
- **Promote Data Usage:** Incentivize low-data users to engage more.

## Next Steps

- Test other algorithms like Random Forest or XGBoost for comparison.