

# TELECOM CHURN – DOMAIN- ORIENTED CASE STUDY (DS-69)

To predict the churn in the ninth month using the data (features) from the first three months

## **Group Members:**

Mohan Tallapragada

Smita Gaikwad

# PROBLEM STATEMENT

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.

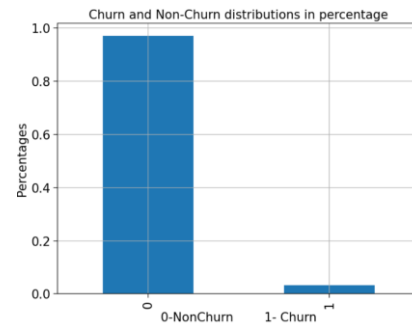
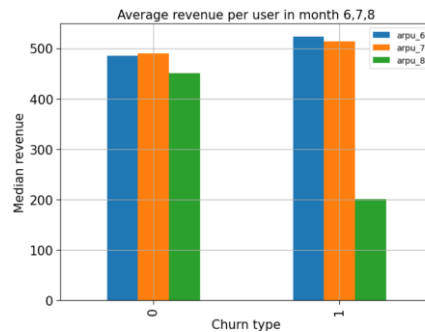
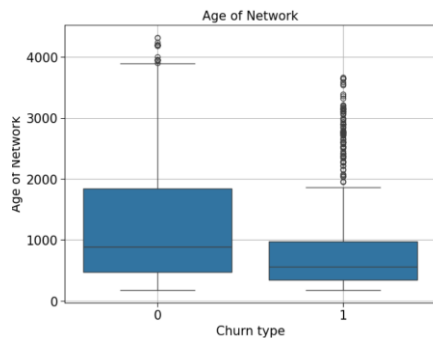
# APPROACH

## Technical Aspects

- Data pre-processing techniques such as handling missing values, feature scaling, and feature engineering
- Exploratory data analysis to understand the distribution of various features and their impact on the target variable
- Building and fine-tuning various classification models to predict churn
- Evaluating model performance using various metrics such as accuracy, precision, recall, and AUC
- Dealing with imbalanced classes using techniques such as oversampling, under sampling, and SMOTE

## II. Business Aspects

- Understanding the key factors Identifying the most profitable customer segments
- Factors that contribute to customer churn
- Analysing the impact of various marketing and promotional campaigns on customer churn
- Developing customer loyalty programs to retain customers
- Identifying areas of improvement in customer service and support to increase customer satisfaction and reduce churn
- Understanding the competitive landscape and devising strategies to stay ahead of competitors



EDA -PLOT TO CHECK PERCETANGES OF CHURN  
AND NON CHURN DATA

# KEY OBSERVATIONS

- Logistic Regression:

- Accuracy 0.8438162985149132

- Precision 0.8232809199718376

- Recall 0.8755771870710096

- f1score 0.8486241306319927

- Decision Tree:

- Accuracy 0.7928469241773963

- Precision 0.037138927097661624

- Recall 0.5294117647058824

- f1score 0.06940874035989718

- Top Features:

- ['std\_og\_t2f\_mou\_8', 'og\_others\_7',

- 'og\_others\_8', 'loc\_ic\_mou\_6',

- 'spl\_ic\_mou\_7', 'spl\_ic\_mou\_8',

- 'total\_rech\_num\_7', 'total\_rech\_num\_8',

- 'last\_day\_rch\_amt\_8', 'count\_rech\_2g\_8',

- 'sachet\_2g\_8', 'sep\_vbc\_3g',

- 'tenure', 'loc\_ic\_mou\_diff', 'spl\_ic\_mou\_diff'

# CONCLUSION

- Last month reach amount is playing crucial indicator of Churn.
- Roam outgoing, std outgoing & last recharge for 8th Month are the most important columns to predict churn..
- ARPU difference is a crucial indicator for churn
- customers with tenure less than 4 yrs are more likely to churn.
- Max Recharge Amount is a strong feature to predict churn.
- Random Forest produced the best results followed by Logistic Regression

# RECOMMENDATIONS

Overall, the analysis provides valuable insights into the factors that lead to customer churn and offers actionable recommendations that can be used to reduce churn and improve customer retention. The analysis also showed that the random forest classifier model performed the best in predicting churn with an accuracy of ~78%. The use of imbalanced data techniques such as SMOTE helped improve the performance of the model by addressing the class imbalance problem.

- Provide incentives to customers who have been with the company for a longer period to encourage them to stay with the company.
- Offer discounts or other benefits to customers who sign up for long-term contracts.
- Encourage customers to use payment methods other than electronic check by providing incentives or simplifying the payment process.
- Provide better value for money by offering packages that provide services at a lower cost for customers with high monthly charges.

