# MSDS-124

# Predicting Telecom Customer Churn

## Outline

1. Problem Definition  
2. Questions to Address  
3. Data Overview  
4. Data Wrangling  
5. Exploratory Data Analysis (EDA)  
6. Predictive Analysis  
7. Hyperparameter Tuning  
8. Feature Importance  
9. Key Insights  
10. Conclusion

## 1. Problem Definition

**Type:**

Structured Binary Classification Problem

**Objective:**  
To predict customer churn for a telecom subscription-based service and identify key factors driving churn to improve customer retention strategies.  
  
**Significance:**  
Reducing churn impacts profitability directly, and insights enable targeted retention campaigns.  
  
**Potential Impact:**  
Improved customer satisfaction, loyalty, and optimized marketing efforts.

## 2. Questions to Address

1. What are the key factors influencing customer churn?  
2. Can we accurately predict churn using historical data?  
3. What actionable insights can reduce churn rates?

## 3. Data Overview

**Dataset Source:**

Kaggle   
Telco\_customer\_churn.xlsx containing 7,043 records with 33 features.  
  
**Key Features:**  
**Demographics:** Gender, Senior Citizen, Partner, Dependents.  
**Services:** Internet Service, Tech Support, Streaming Services.  
**Account Info:** Contract Type, Monthly Charges, Total Charges.  
**Target Variable:** Churn Label (Yes/No).

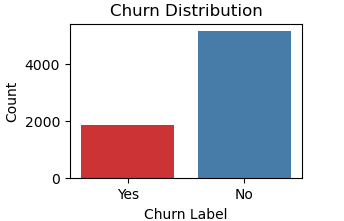
## 4. Data Wrangling

Steps Taken:  
**1. Missing Values:** 'Total Charges' converted to numeric and rows with missing values removed.  
**2. Dropped Irrelevant Columns:** CustomerID, Lat Long, State, City, and Zip Code.  
**3. Categorical Encoding:** Converted categorical features using LabelEncoder.  
**4. Scaling:** Scaled numerical features (Monthly Charges, Total Charges, Tenure Months) using StandardScaler.

## 5. Exploratory Data Analysis (EDA)

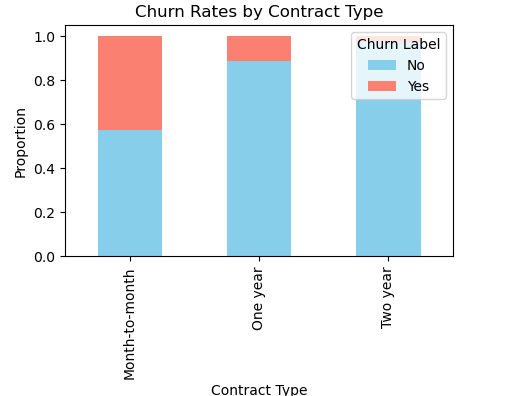
### 5.1 Churn Distribution

Finding:  
There is a class imbalance; most customers have not churned.



### 5.2 Churn Rates by Key Features

Contract Type:  
Customers on 'Month-to-month' churn rates compared to 'One year' or 'Two year' contracts.



### 5.3 Correlation Analysis

Key Correlations:  
-Tenure Months is negatively correlated with churn (-0.35).

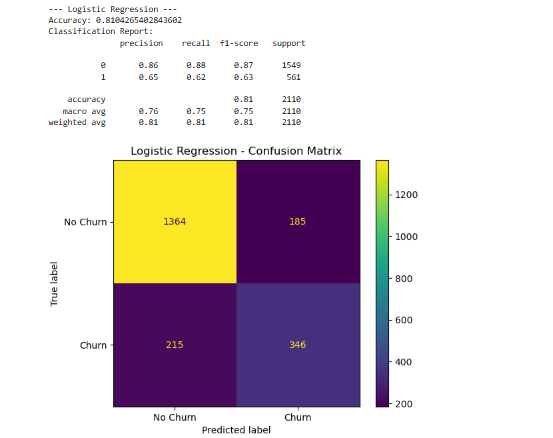
  
 Monthly Charges has a positive correlation with churn (0.19).



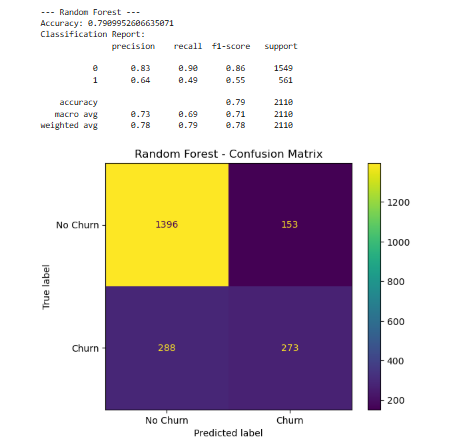
## 6. Predictive Analysis

The following models were trained and evaluated:  
**1. Logistic Regression**

**Model Evaluation Results:**



**2. Random Forest**



## 7. Hyperparameter Tuning

Tuned Model: Random Forest  
Best Parameters:  
- n\_estimators: 300  
- max\_depth: 20  
- min\_samples\_split: 5  
  
Tuned Accuracy: 81

## 8. Feature Importance

Top Features:  
1. Tenure Months  
2. Contract Type  
3. Monthly Charges  
4. Total Charges

A graph of a bar graph

Description automatically generated with medium confidence

## 9. Key Insights

1. Tenure Management:  
Engage with newer customers early to increase tenure and reduce churn.

2. Pricing Optimization:  
Monitor customers with high monthly charges, as they are more likely to churn.

## 10. Conclusion

The analysis provided a clear understanding of churn behavior and key drivers.  
Logistic Regression emerged as the best model with an accuracy of 81%, and insights derived can directly aid customer retention strategies.