Customer Churn Analysis for Telecom Industry

**Anushree P  
Sai Vidya Institute of Technology  
Department of Computer Science (Data Science)**

# Introduction

In a highly competitive telecom industry, retaining existing customers is more cost-effective than acquiring new ones. Customer churn refers to the phenomenon where users discontinue their subscription. The goal of this project is to analyze patterns that lead to churn and develop a predictive model to identify at-risk users early.

# Abstract

This project utilizes a real-world telecom dataset to perform exploratory data analysis, apply machine learning techniques for churn prediction, and interpret model results to gain insights. Customer segmentation is also performed to categorize users into Loyal, At Risk, and Uncertain groups for targeted interventions.

# Tools Used

• Python (Pandas, Seaborn, Matplotlib, Scikit-learn)  
• Jupyter Notebook  
• Machine Learning (Random Forest Classifier)  
• Data Preprocessing and Visualization Libraries

# Steps Involved in Building the Project

1. Data loading and preprocessing: Cleaned null values, encoded categorical features.  
2. Exploratory Data Analysis (EDA): Visualized class balance, feature distribution, and correlation.  
3. Model building: Trained a Random Forest Classifier to predict customer churn.  
4. Evaluation: Measured accuracy, confusion matrix, and classification report.  
5. Feature Importance: Identified key drivers of churn.  
6. Segmentation: Classified users as Loyal, At Risk, or Uncertain based on predictions.

# Conclusion

The Random Forest model achieved good accuracy and clearly highlighted important features contributing to churn. Segmenting customers helped in identifying those who are likely to churn, enabling telecom companies to implement proactive retention strategies.