## SCTR’s Pune Institute of Computer Technology Dhankawadi, Pune

**A Data Science and Big Data Report ON**

(Customer churn Analysis)

## SUBMITTED BY

Name : Karan Mundhada Name : Atharva Pardeshi

Class : TE-2 Class : TE-2

Roll no : 31254 Roll no : 31257

# Under the guidance of

(Dr. S. S. Sonawane)



## DEPARTMENT OF COMPUTER ENGINEERING

ACADEMIC YEAR 2023-24

**Contents**

1. [**Introduction**](#_heading=h.30j0zll) 3
2. [**Problem Statement**](#_heading=h.1fob9te) 4
3. [**Objectives and Scope**](#_heading=h.3znysh7) 5
4. [**Methodological Details**](#_heading=h.2et92p0) 6
5. [**Modern engineering tools used**](#_heading=h.tyjcwt) 7

# Introduction

* The telecom industry is highly competitive, with companies constantly vying for customer loyalty.
* Customer churn, or the rate at which customers leave a service provider, is a critical metric in this industry.
* Analyzing customer churn helps telecom companies understand why customers leave and identify strategies to retain them.
* In this project, we aim to analyze customer churn in the telecom industry using data science techniques.

# Problem Statement

* Telecom companies face the challenge of retaining customers in a saturated market with numerous service providers.
* Understanding the factors influencing customer churn is crucial for developing effective retention strategies.
* The primary goal of this project is to identify patterns and predictors of customer churn and provide actionable insights for reducing churn rates.
* By leveraging data science methodologies, we aim to develop predictive models to forecast customer churn and recommend targeted interventions.

# Objectives and Scope

* **Objective 1**: Analyze historical customer data to identify patterns and trends associated with churn behavior.
* **Objective 2**: Develop predictive models to forecast customer churn based on demographic, usage, and behavioral factors.
* **Objective 3**: Evaluate the effectiveness of various features in predicting churn and prioritize those with the highest impact.
* **Objective 4**: Provide actionable insights and recommendations to telecom companies for reducing churn rates and improving customer retention.

**Scope :**

* This project will focus on analyzing customer churn within a specific telecom company's dataset.
* The analysis will encompass demographic information, usage patterns, customer service interactions, and other relevant data points.
* The predictive models developed will be based on machine learning algorithms such as logistic regression, decision trees, and neural networks.
* Recommendations will be tailored to address specific churn drivers identified in the dataset and may include targeted marketing campaigns, personalized offers, and improved customer service strategies.

# Methodological Details

* Data Collection: Obtain historical customer data from the telecom company's database, including demographic information, service usage, billing records, and customer interactions.
* Data Preprocessing: Cleanse the data, handle missing values, and encode categorical variables for analysis.
* Exploratory Data Analysis (EDA): Explore the dataset to understand the distribution of variables, identify correlations, and visualize patterns related to churn behavior.
* Feature Engineering: Extract relevant features from the dataset and create new variables to improve the predictive power of the models.
* Model Development: Utilize machine learning algorithms such as logistic regression, decision trees, random forests, and gradient boosting machines to develop predictive models for churn prediction.
* Model Evaluation: Assess the performance of the models using metrics such as accuracy, precision, recall, and F1-score. Conduct cross-validation to ensure robustness and generalizability.
* Model Interpretation: Interpret the results of the models to understand the factors driving churn and their relative importance.
* Recommendation Generation: Based on the insights gained from the analysis, generate actionable recommendations for reducing churn rates and improving customer retention.
* Documentation and Reporting: Document the entire process, including data preprocessing steps, model development, evaluation metrics, and recommendations, in a comprehensive report.

# Modern engineering tools used

* Python Programming Language: Utilized for data manipulation, analysis, and modeling using libraries such as pandas, NumPy, and scikit-learn.
* Jupyter Notebooks: Used for interactive data analysis and code documentation.
* Data Visualization Libraries: Employed tools like Matplotlib and Seaborn for creating informative visualizations to aid in data exploration and presentation.
* Machine Learning Libraries: Leveraged sci-kit-learn for developing and evaluating predictive models.
* Version Control Systems: Git/GitHub for managing code repositories and facilitating collaboration among team members.