

**Project's Goal:** The goal of the project is to analyze the customer behaviour from the dataset and predict the customer behaviour. The project predicts whether a customer is more likely to churn(leave) or stay. This type of analysis is important in designing the customer retention programs.

**Customer and the importance to the customer:**

Our customers will be the telecom companies. These companies will greatly benefit by the insights that are obtained from the project which will make their customer retention programs more effective

**Data to use:** Telco Customer Churn dataset.

**Algorithms:**

- 1) Random forest
- 2) Logistic regression
- 3) Decision Tree
- 4) Ada Boost
- 5) Naive Bayes

**Learning outcomes:**

- 1) The factors that influence customer churn (Eg: price, tenure etc)
- 2) Evaluation of various prediction models and their effectiveness.
- 3) Ability to give recommendations and suggestions to reduce the customer churn.

ChatGPT:

**Project Title: Customer Churn Prediction for Telecom Companies**

**Project Goal:**

The primary objective of this project is to analyze customer behavior within the telecom industry and predict the likelihood of customer churn, where churn refers to customers leaving the service provider. The goal is to provide insights that help telecom companies design more effective customer retention programs by understanding the factors contributing to customer churn. By predicting which customers are at risk of churning, companies can take targeted actions to retain them, ultimately reducing revenue losses and improving long-term customer loyalty.

**Customer and Importance:**

Our customers are telecom companies who face the ongoing challenge of retaining their subscriber base in a highly competitive market. The insights generated by this project will enable telecom providers to identify the key factors influencing customer churn, such as pricing, service tenure, and customer satisfaction. These insights are critical for developing proactive strategies, such as personalized offers and loyalty programs, to prevent churn. By leveraging the findings from the model, telecom companies can reduce churn rates, improve customer retention, and increase overall profitability.

**Data:**

The project will use the **Telco Customer Churn dataset**, which includes various customer attributes such as demographic details, contract information, service usage, and tenure. This dataset provides a comprehensive view of the factors that may contribute to a customer's decision to churn.

**Algorithms:**

The following machine learning algorithms will be applied to predict customer churn:

1. **Random Forest:** A powerful ensemble method that combines multiple decision trees to improve prediction accuracy.
2. **Logistic Regression:** A widely-used statistical model for binary classification problems like churn prediction.
3. **Decision Tree:** A simple, interpretable model that can visualize how decisions are made based on customer data.
4. **AdaBoost:** A boosting algorithm that helps improve model performance by focusing on difficult-to-predict cases.
5. **Naive Bayes:** A probabilistic model that will help in understanding the underlying patterns in customer behavior.

**Learning Outcomes:**

Through this project, the following insights and skills will be gained:

- **Churn Factors:** Understanding the key factors influencing customer churn, such as contract length, pricing models, and customer support interactions.
- **Model Evaluation:** Comparing the performance of different machine learning algorithms in predicting churn and identifying the most effective model for this use case.
- **Recommendations for Retention:** Based on the analysis, actionable recommendations will be made to telecom companies to reduce customer churn, such as offering targeted discounts or improving service quality.

This project will deliver a valuable framework for telecom companies to anticipate customer churn and take preemptive measures, leading to enhanced customer satisfaction and retention.

**Humanized version of above([link](#)):****Project Goal:**

This project primarily pertains to the prediction of customer churn - that is, whether a telecom customer will leave or stay on with the service. This shall outline an understanding of customer behavior and the identification of key factors that influence churn to enable insight for informed decisions in the telecom company for the retention of customers. The project further provides recommendations on reducing churn through early identification of at-risk customers.

**Importance and Customer:**

The project is of importance for every telecom company since customer churn represents a big loss in revenue and market share. Thus, the accurate prediction of churning will enable the telecom providers to adopt proactive measures such as offering personalized retention programs, improvement in customer service, or pricing model adjustments. The insights from this project would provide information on how to enhance customer satisfaction and loyalty, thereby directly impacting profitability and long-term customer retention.

#### Data:

The two major datasets that will be engaged in this work are discussed below.

**Cell2Cell Customer Churn Dataset:** This is a rich, real-world dataset that contains data about more than 71,000 US-based telecom customers. The features include customer demographics, service plans, and usage patterns.

**IBM Telco Customer Churn Dataset:** Although much smaller, it is one of the most widely used publicly available datasets, comprising around 7,000 customer records in total. The information involves tenure, contract type, and monthly charges.

These datasets, with their rich set of features, can be used to delve deep into the analysis of customer behavior and usage in terms of their contracts and hence provide a very strong foundation for churn prediction.

#### Algorithms:

For predicting customer churn, the following machine learning algorithms will be used:

**Random Forest:** powerful ensemble method that does exceptionally well when there is a large dataset and complex interactions among features.

**Logistic Regression:** the classic statistical model used for binary classification problems like churn prediction.

**Decision Tree:** The simple and interpretable model to understand the reason for decisions based on the features of the customers.

**AdaBoost** is a boosting algorithm that focuses on hard-to-predict cases and improves overall prediction accuracy.

**Naive Bayes:** This is a probabilistic model that does prediction faster and efficiently. The algorithm can be applied for knowledge discovery to find out the hidden pattern in customer data.

#### Expected Learning Outcomes:

From this project, we would like to learn:

**Key Churn Factors:** Understand the most important driving factors to customer churn, which include contract length, pricing, and service usage patterns. **Model Evaluation:** Understand how different algorithms perform and which models work best in predicting churn. **Practical Insights:** From the findings, actionable recommendations will be made for a telecom company on targeting customers at higher risk of churning and suggesting retention strategies.

We hope to have a high-value model for churn prediction and at least some practical insight into real-world customer retention programs as the outcome of this project.

Version 1:

## Telecom Customer Churn Prediction

Shiva Dhanush Konuru

Manikanta Sai Kommireddy

Mayuresh Bhangale

Chetan Rahane

sxk230027

mxk220132

mxb240017

crr220000

---

### Objective

The main goal of this project involves analyzing the behavior of customers in the telecom sector and establishing the possibility of churning, a situation in which customers leave the service provider. It aims to bring useful insights that may enable the telecom firms to create effective customer retention programs through the recognition of various variables that drive customers to leave. Predicting which customers churn, on the other hand, enables companies to take corrective actions in order to retain them in a very focused manner and thereby help decrease revenue losses, enhancing the quality of long-term customer loyalty.

### Customer and Importance

Our customers are telecom companies facing the problem of retaining their subscriber base continuously within a highly competitive market. The model should provide the telecom providers with insights into the most influential factors that cause customer churn, including tariff plans, service duration, and customer satisfaction. This will become crucial for proactive measures in terms of customized offers and loyalty programs to help prevent churn. Based on the output from the model, the telecom companies are able to reduce churn, retain customers, and consequently gain higher profitability.

### Data

The project will make use of the Telco Customer Churn dataset, containing various customer attributes, including demographic details, contract information, service usage, and tenure. This gives a broad basis on which such customers can actually leave. In other words, this is a complete dataset that encompasses almost all possible causes for churning.

### Algorithms

Following are some of the machine learning algorithms that will be applied for the prediction of customer churn:

- **Random Forest:** A very powerful ensemble method in which, to improve upon the results of prediction, several decision trees are combined.
- **Logistic Regression:** This is one of the most used statistical models that is used to predict binary classification problems, such as churn prediction.
- **Decision Trees:** To give insight into how decisions are made on customer data in a very interpretable way.
- **AdaBoost:** To further improve model performance by focusing on hard-to-predict examples.
- **Naive Bayes:** A probabilistic model which will help in understanding the underlying patterns in customer behavior.

### Learning Outcomes

The following insights and skills are obtained with this project:

- **Churn Factors:** Identify the most important factors causing changes in customer churn, such as contract duration, pricing models, and interaction with customer support.
- **Model Evaluation:** Compare the performance of different ML algorithms in predicting churn and point out the most potent model in this use case.

The outcome of this project will give a great framework to the telecom companies in predicting customer churn and undertaking necessary steps to counter the situation, hence improving their customer satisfaction by reducing churn.