**Telecom Churn prediction using machine Learning**

Telecom markets are one of the most competitive markets in the world, the presence of a huge number of providers in each country makes these companies the obligation to search for new methods to keep their customers. Facing this fierce competition all the companies become more customer-centric by focusing on Customer Relationship Management to retain existing customers. In its research for Bain and Company, Frederick Reichheld stated that the cost of acquiring a new customer could be higher than that of retaining a customer by as much as 700% and that increasing customer retention rates by a mere 5% could increase profits by 25% to 95%.

Telecom is considered one of the most businesses that generate a huge amount of data because to allow communication between customers it is necessary to record all the information about all the activities made by the client. All providers collect information about their clients like call detail record information, customer demography and personal details, bill, and payment details, and customer value-added services details. The huge volume of data encourages these companies to use data-driven techniques such as machine learning and deep learning to get insights and to analyze the behavior of customers to predict their future comportments and have time to work on fixing issues beforehand.

Churn prediction is the process of determining whether customers can leave in a certain period. This prediction and quantification of the risk of losing customers can be done globally or individually and is mainly used in areas where the product or service is marketed on a subscription basis. The prediction of churners is generally based on analyzing and studying the customer's behavior. It involves the use of modeling and machine learning techniques that can sometimes use a considerable amount of data.

These behaviors can be:

* Variations in consumption or user behavior.
* A change to inactive client status or a drop in service usage.
* The formulation of a claim (number, frequency, and types of claims).
* An increase in consumption leading to a sharp rise in the bill.

The data used in this project can be download from the web site of Kaagle:

<https://www.kaggle.com/datasets/jpacse/datasets-for-churn-telecom?resource=download>

Below are guidlines on how to write-up your report for the final project.

Tasks to be implemented in python:

1. Preprocessing task :
2. Summary and basic information of the data
3. Replace null values and perform exploratory data analysis
4. Transform skewed variables, categorical and binary features
5. Selection features: use three methods for feature selection:
6. Remove collinear features
7. Remove features with greater than a threshold percentage of missing values
8. Keep only the most relevant features using feature importance from a model
9. Make prediction using machine learning models: Logistic Regression, SVM, k-NN, CNN, RF, XGboost, Neural Network …etc.

Tasks to be discussed:

1. What conclusions do the results support about the strengths and weaknesses of each method?
2. How can the results be explained in terms of the underlying properties of the algorithm and/or the data.