General presentation of the project

Name of the selected project

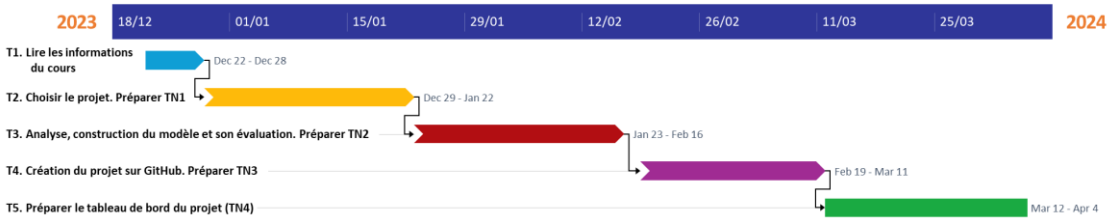
Telecom-churn

Definition of the project

Orange Telecom needs to minimize the number of subscribers who churn from their services, or leave the network for another one. The data provided by the company is that of about 3300 subscribers, some of whom have churned. To minimize the number of churns is to minimize the cost of increasing the number of subscribers and increasing revenues.

The objective is to build a prediction model that can alert managers at Orange Telecom when one or more subscribers are at risk of churning. These managers will have a plan to execute to retain the maximum number of these subscribers.

Project Planning



Risks:

- *T1: Need more time to understand all the information.*
- *T2: Not finding a suitable project within the planned period.*
- *T3: Encounter problems while preparing or executing code.*
- *T4: Encountering issues when using Git/GitHub.*
- *T5: Need more time to finalize the project.*

Evaluation of the performance of the implementation

- *Variance between planned schedule and actual production schedule: The target is to finish on the same scheduled date. The baseline situation is a one-week gap. The result: a two-week gap more than planned.*

- *Gap between the number of hours worked and planned: the target is zero hours and the baseline situation is a gap of 10 hours more than planned. The result: a gap of -3 hours (less than the scheduled hours)*

Explanation of variances in delivery from planning

Reminder of the results actually achieved with the initial objectives.

This project built a model to predict which subscribers have a high chance of churning from network services. The model is based on 19 attributes for each subscriber. At the end, observations are summarized that are also based on correlations and data visualizations.

Justification for the choices made during the implementation:

Approach or methods

Supervised learning is used because the response variable exists in the dataset (called 'Churn').

Applied techniques

The following four algorithms are used to build the model and the results are evaluated according to a few criteria (explained in the detailed document). Then, we choose the algorithm that gives the best prediction results.

- KNeighborsClassifier
- DecisionTreeClassifier
- RandomForestClassifier
- XGBClassifier

Tools used

- *Git and GitHub: project versioning and sharing*
- *Visual Studio Code: Running Code Locally*
- *Google Colab: Online Code Execution*
- *Replit: Running the code online (another option)*

Statement of Problems

- *Checking the correlation of the attributes, we see that all the attributes have a low correlation with the "Churn" response attribute. A few attributes have a strong correlation (about 1) with other attributes. Only one attribute of each pair was kept during the modelling.*
- *There are attributes that have non-numeric values. We transformed those with 2 unique values so that they have the values 0 or 1, which is very useful for statistical calculations.*
- *The tools: it was difficult to execute the code on Jupyter Notebook. Google Colab was a much simpler alternative.*

Conclusion and recommendations

What would be done differently if I had to do it all over again?

- *Use the Google Colab and/or VS Code tool from the beginning for code execution.*
- *Plan the work on the project well and follow the execution plan.*

Most significant learning

- *See how to compare the results of different learning algorithms.*
- *Understand how to use the different tools together: VS Code, Google Colab, and Git/GitHub.*