

Scoring Part 2

TSE – M2 DSSS – 2024-2025 – Raquel ALCARAZ

Project

Context:

With the provided dataset, your task is to predict customer behavior in order to enhance retention for an anonymous company. By analyzing all relevant customer information, you will be able to develop targeted strategies aimed at improving customer retention.

Data:

Each row in the dataset represents an individual customer, and each column contains specific customer attributes as outlined in the column Metadata. The dataset includes the following information:

- **Churn:** indicates whether or not the customer left within the last month (target variable).
- **Services:** Data about the services each customer has subscribed to, such as phone, multiple lines, internet, online security, online backup, device protection, technical support, and streaming TV and movies.
- **Customer account details:** Information such as the length of time a customer has been with the company, contract type, payment method, paperless billing status, monthly charges, and total charges.
- **Demographic data:** Customer demographic details such as gender, age range, and whether they have partners or dependents.

The raw data contains 7043 rows (customers) and 21 columns (features).

This information will allow you to analyze patterns related to **customer churn**, predict if a customer is going to churn and develop strategies for retention.

Expected work:

You should prepare a presentation of 15 minutes maximum (PowerPoint shape, in ppt or pdf format) for the steering committee of the marketing department on the **26/11/2024**.

The outline of the presentation is imposed as follows:

- 1- Context & objectives of the meeting
- 2- Presentation of the data analysis results
- 3- Machine Learning methodology used
- 4- Results and recommended marketing strategies
- 5- Conclusion

You will be graded on content and form. Be professional!

I expect marketing recommendations that focus on developing targeted propositions to retain clients at high risk of churn. The aim is to identify specific value propositions or engagement strategies that will effectively address their needs and motivations, encouraging them to stay with us.

Data variables

- **id** : unique ID that identifies each customer
- **gender** : Whether the customer is a man or a woman
- **customer_senior**: Whether the customer is a senior citizen (older than 65) or not (1, 0)
- **customer_partner** : Whether the customer has a partner or not (Yes, No)
- **dependent_family** : Whether the customer has dependents (children, parents, grandparents...) or not (Yes, No)
- **month_tenure** : Number of total months the customer has been with the company
- **phone_subscription** : Whether the customer subscribes to home phone service with the company (Yes, No)
- **multiple_line** : Whether the customer has multiple telephone lines or not (Yes, No, No phone service)
- **internet_type** : Indicates if the customer subscribes to Internet service with the company (DSL, Fiber optic, No)
- **online_security** : Whether the customer subscribes to an additional online security service provided by the company (Yes, No, No internet service)
- **backup_service_online**: Whether the customer the customer subscribes to an additional online backup service (Yes, No, No internet service)
- **device_protection_plan**: Whether the customer subscribes to an additional device protection plan for their Internet equipment (Yes, No, No internet service)
- **technical_support** : Whether the customer has technical support or not (Yes, No, No internet service)
- **streaming_tv** : Whether the customer uses their Internet service to stream television programming from a third party provider. The company does not charge an additional fee for this service (Yes, No, No internet service)
- **streaming_movie** : Whether the customer uses their Internet service to stream movies from a third party provider. The company does not charge an additional fee for this service (Yes, No, No internet service)
- **contract_type** : Customer's current contract type (Month-to-month, One year, Two year)
- **paperless_billing** : Whether the if the customer has chosen paperless billing or not (Yes, No)
- **payment_method_type**: Indicates how the customer pays their bill (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))
- **amount_charges_monthly** : Indicates the customer's current total monthly charge for all their services from the company
- **amount_total_charges** : The total amount charged to the customer
- **Churn** : Whether the customer churned or not (Yes or No)

Practical work guidelines

Organization:

Over the next two practical sessions (and continuing at home), you will organize into **4 groups of 4 students and 1 group of 3 students** for the project that will be presented during the oral examination on 26/11/2024. You will email me the composition of each group before November 8.

Email address: raquel.alcaraz-serna@ut-capitole.fr

Each group organizes itself as it sees fit for the project and can parallelize the different components of the project (Python code, data analysis, PPT presentation, overall organization) but each member must be comfortable with all the steps. On the day of the presentation, all group members must participate in presenting.

Collaboration between groups is completely fine, but any copied and pasted sections will be easily identifiable.

The **deliverables** (both the presentation in PPT format and the notebook code) must be submitted **no later than November 25 at midnight**.

I expect each team member to focus on developing a solid understanding of each component of the project, aiming for clarity and coherence over complexity. Rather than creating overly complicated solutions, please prioritize a clear approach that you fully understand and can confidently explain. Each part should logically connect to the overall objectives and be presented with clear, consistent reasoning.