Project Future

Data & Decision Sciences

21 May 2020

You are hired as a data scientist for a large bank. Your goal throughout this exercise is to familiarize yourself with the domain and use this knowledge to produce value for the bank, through the analysis of the data.

1. Exploratory Data Analysis

Perform an initial Exploratory Data Analysis on the given dataset. The goal of this task is to familiarize you with the dataset. In this step you should:

- Understand what each table/column represents.
- Recognize if there are missing/wrong values in the data.
- Check for outliers in the data.
- Obtain insights through descriptive statistics.
- Identify key relationships within the data.

Hint: visualizations are key for this step!

2. Initial preprocessing

The goal of this task is to bring your data in the proper form to be further analyzed in the next steps. In this step you should:

- Clean the data (handling missing values, fixing errors, outliers, etc.)
- Encode the variables.
- Perform any feature engineering step you see fit (group features, define KPIs, etc.)

3. Classification

This bank's Board of Directors is planning an aggressive marketing campaign to increase their number of paid subscription plans. It has tasked you to build a model that can predict if a user is likely to switch to a paid plan.

In this task you should:

- Build a model that correctly classifies users based on their plan, i.e. "Standard" vs "Paid" (all subscription-based plans should be grouped under the "paid" label)
- You can use any technique you want (heuristic, statistical, machine learning. etc.)
- The goal of this task is to produce a pipeline for predicting if a user has a "paid" or "standard" plan.
- Be careful of what features you use for this task. Will all of the features be available to you in a real-world scenario?
- During the final days of the program, you might be asked to produce predictions on number of users, previously undisclosed to you!

4. User modeling

At this point, you aim to group the users based on their most descriptive characteristics (i.e perform some kind of clustering).
For this task try to:

- Select a subset of the most useful features
- Enrich it with new, more meaningful features
- Combine information from different tables
- Present the results in a suitable format and elaborate on their descriptive characteristics

5. Unengaged and Churned Users

One of the prevalent tasks that subscription services have to deal with is that of "churn prediction". The issue with this task is that there are usually no actual labels; they heavily rely on the definition of "churn".

- Define a target metric to measure user engagement. How would you define an engaged vs. unengaged user?
- Using your logic from above, build a model (heuristic/statistical/ML) to classify engaged and unengaged users.
- Let's assume an unengaged user is a churned user. Now suppose we use your model to identify unengaged users and implement some business actions try to convert them to engaged users (commonly known as reducing churn).
- How would you set up a test/experiment to check whether we are actually reducing churn?
- What metrics and techniques would you use to assess the impact of the business action?

General guidelines:

- Each team has been invited to their own, exclusive Teams Private Channel, along with their instructors. You can enjoy every advantage of Microsoft Teams there, like:
  - Communicating with each other as a team
    The most efficient way is to create a "thread" with a specific topic, where people can reply to, rather than writing each message as standalone and making the chat difficult to navigate.
  - Creating, scheduling and hosting your own meetings, sharing your screens, etc.
  - Sharing files and links
  - Tagging specific people (by typing "@" and then their name)
  - Ask for your instructors help and contribution

- Each team has been invited to their own, exclusive Private GitHub Repository, along with their instructors. You are advised to share code and collaborate with each other through this platform.

- This Team Project will conclude with a short presentation of your work and your findings. This should be considered as your final task, I.e. to summarize a project that required multiple hours to do in just a few minutes, while still remaining engaging. Someone might appreciate your work and reward you for it!