Technical test

Data Science challenge

At Modjo, we gather and process all audio and video meetings between Sales / Customer Success teams and their prospects/clients to provide insights and analytics on how companies interact with them.

In this test, we will focus on how Modjo software is used by… Modjo! Indeed, we are diligent users of our own product and we use it extensively to continuously improve our sales and customer care processes.

Problem to solve

Audio and video calls with our prospects/clients are of different types, and they can be labeled depending on these types using **tags** among the following:

Cold Call - 1st Call - Demo - Prospect Follow Up - Set up - Onboarding Team - Onboarding Managers - Client Follow Up - Unscheduled Follow up - Other
(some details on these can be found at the end of this document)

Your task will be to **predict which tag** should be attached to calls based on some of their characteristics.

You will have to deliver a fully functional **python notebook** answering the following questions. You may write some parts of your code in separate python files and import them into the notebook if you want to. Feel free to use any python library you find useful.

You must not hesitate to take relevant initiatives and explain your choices.

📊 **Part I - Problem understanding and data exploration**

- What is the value of using a prediction algorithm in a product such as Modjo?
- What can you tell from the dataset features? Are all the columns relevant for the task at hand? You may use various plots to illustrate your findings and set the stage for the ML task.

⚙️ **Part 2 - Machine Learning**

- Build a model which must be able to predict the **tag** from the other characteristics of a call. Explain your choices and potential feature engineering steps.
- How do you evaluate its performance? How could it be improved?
- Can you think of some next steps which would be needed to bring your algorithm into production? What could go wrong in a production environment?

💯 **We will especially assess:**

- Problem framing / data exploration
- ML approach
- Interpretation of results
- Clean/understandable code

Dataset

You should have received a csv file with the dataset for the challenge.
It contains real data from our database, where the tags were manually assigned to each call. These calls are described by several characteristics explained at the end of this document.
Each line corresponds to a tag attributed to one call.

NB: It's rare but possible for one call to have 2 (or more) different tags (if a demo is done during a 1rst call for example), in which case it will be represented by as many lines in the dataset. You can consider that we only need to predict one tag for each call.

Description of the tags

Cold Call: Call on the phone a potential prospect which may have been contacted by email before but did not answer, and may never have heard of Modjo. The objective is to book a meeting.

1st Call: First video call to make the prospect talk about their needs and sometimes briefly show them the product.

Demo: Walk prospective customers through the product to show them what it does.

Prospect Follow Up: Get some updates on where the prospect stands and whether or not they made a decision about buying Modjo.

Set up: Call with the administrators of new clients to connect to all their other tools, give relevant access to new users etc.

Onboarding Team: The prospect has just become a client and we onboard the users and show them how to use Modjo.

Onboarding Managers: The prospect has just become a client and we onboard the team managers and show them how to use Modjo.

Client Follow Up: Get some updates on how things are going for a client.

Unscheduled Follow up: Quick update / news check.

Other: The call was tagged as “Other” (different from not being tagged at all)

Description of the dataset columns

Column	Description
id	Unique id of the call
duration	Duration of the call in seconds
date	Datetime of when the call took place
userId	Unique id of the modjo user (sales or customer care) performing the call
modifiedById	Unique id of the modjo user who edited the call lastly
phoneProvider	Phoning tool with which the call was made and recorded
direction	Outbound: Modjo called the prospect/client Inbound: the prospect/client called Modjo
mediaType	Whether the call was a video or an audio call
dealId	Id representing the deal, which can be seen as a proxy of the client
userTalkRatio	Percentage of the time where the Modjo user speaks
longestContactMonologue	Longest continuous period of time during which the prospect/client spoke without interruption, in seconds
patience	Average time it takes for the modjo user to answer after the prospect/sales is done talking during the call
interactionSpeed	Average number of speaker changes in one minute
role	Whether the modjo user is admin or regular user
teams	Team(s) to which the modjo user belong
contacts	How many non-modjo users attend the call