Data Science Challenge

Dear candidate, as part of the recruitment process, we would like to evaluate your technical skills with a challenge:

Problem Statement

'Znailla Bank', a subsidiary within the 'Znailla Group', which is a multinational insurance group with a presence in the banking industry, offers a term deposit product.

In the last 2 years, 'Znailla Bank' has invested massively in reaching out to its portfolio of existing customers - those with already a current account- to offer the product through a call center campaign. In fact, 'Znailla Bank' has been running a process which selects a subset of their customer portfolio every week, and the list of those customers is sent to the call center to be actioned.

The campaign has a positive impact, but given the significant cost of the call (~8€) and Znailla's commitment to process optimization, the management has decided to investigate how to use Analytics to help us to improve the costs-efficiency of the process.

The head of Marketing has asked whether it would be possible to reduce the number of calls by predicting the "wasted calls" (calls to non-converting customers). Specifically, he asks how many calls can be saved, given that he wants to lose as little business as possible.

As the data scientist responsible for this project, your role is to build a Machine Learning pipeline to identify customers that should be called within the weekly cohort and provide recommendations on how to reduce the number of calls.

To do so, you are given access to a dataset (https://archive.ics.uci.edu/dataset/222/bank+marketing - please use bank-additional-full.csv), containing the history of calls and conversion. The head of Marketing is particularly interested in:

- How you frame the problem
- Which the drivers of conversion are
- What your recommendation would be to cut the number of calls supposing that you get a "test/prediction set" of new customers every week.

The head of Marketing adds the following pieces of information:

- Each new contract will generate ~80€ profits over its lifetime
- No new contract would have been signed without the call (it's the only advertising way). Also, we can assume that if a customer didn't buy with the call, he/she will not buy at any time.

Requirements

The application must be written in Python and non-code deliveries must be written in the English language. Please ensure that your code is understandable.

Expected delivery

- We expect you to provide a link to a code repository containing your proposed solution to this challenge. Alternatively, you may also submit an attached file. Please keep in mind the software design and clean code principles – moreover the scalability of the whole ML product - while implementing your model.
- We also would like to get a rough description about the end to end (from implementation till serving/monitoring) design of such a machine learning model going into production (just a description, not in the code level). This ML System Design will also be discussed along with your submitted code (coming from the previous step) during the interview.