GoHealth: Data Science Exercise

Background Information

A small coffee shop chain in Illinois is asking for some assistance in building a more targeted marketing campaign. They have a list of leads that they have previously reached out to along with some basic information such as where they live, their beverage preferences, etc. They have been able to track these individual users to determine whether or not they made a purchase at the coffee shop chain. This data is stored in a csv file, *train.csv*.

Task 1

a. Create a model to predict if a lead will purchase coffee. This will be indicated by the "bought_coffee" column in the data set.
b. Provide a probability and prediction for the leads in the *predictions.csv* dataset.

Task 2

The owner of the coffee shop takes your results from Task 1 and sends coupons through the mail to the 15 leads with the highest probability score. Only 1 of them end up purchasing coffee from the coffee shops. The owner would like an explanation as to why more leads didn't purchase coffee. How would you respond to them? Could you provide a method of evaluation for your model's performance?

Task 3

The owner of the coffee shop wants you to set up an application in which they provide information about a lead and receive a prediction in near-real time. Describe in a few sentences or bullet points how you would develop a solution that meets the owner's request.

Submission Instructions

1. Include the following:
    a. Any code used to evaluate the data and/or build a model.
    b. A csv containing the results of Task 1b. Make sure to include at least the "ID", "probability", and "prediction" columns.
    c. A written response to Task 2.
    d. A written response to Task 3.
2. Zip all of the files together.
3. Email the zipped files back to the GoHealth recruiter with whom you have been in contact.

Important Notice

- The data used in this evaluation is entirely fictitious.
- The data and details of this exercise are property of GoHealth. Do not distribute any of the included or referenced content.