Awake Labs - Take Home Challenge

README

We have supplied you with a publicly available dataset that includes certain patient attributes and medical data. Your task is to predict the likelihood of a patient getting a heart attack.

We expect this challenge to take between 5-7 hours to complete. We are giving you until November 21st at 5pm ET to share your submission.

What we expect to see in your submission

We expect to receive a link to a Github repo from you along with brief instructions on how to set up and run the code on our machines. You should complete this challenge as though you were writing production-ready code. You are free to choose your preferred coding language and framework to complete this challenge. Please include the following in your notebook:

1. **Exploratory Data Analysis (EDA) and Data Cleaning**
   Perform EDA to gain an understanding of the dataset and what challenges or issues it may contain. This is a good opportunity to make use of visuals to help you describe the data and their relationships. We also expect you to summarize your findings in a few brief bullet points.

2. **Modeling**
   Your task is to predict the likelihood of a patient getting a heart attack. As you complete this task, apply a model of your choosing to the data. Explain which parameters are used in your model and why. Also explain how you decided which model to use, what are the assumptions you made as well as the tradeoffs of your chosen model.

3. **Questions**
   After completing the first two steps above, please answer the following questions:

   a. Please describe how you would put your model into production, including how you would test and monitor the performance of your model.
      i. Think about the challenges you foresee, as well as the resources & help from your colleagues that you may need.

   b. Say you were given 10 minutes of continuous ECG data for each of the patients in the dataset - How would that change your approach?
      i. E.g. What else would you do in your EDA? Would you use a different model?

How you will be evaluated

1. **Communication**
   As the first person to be hired as a Data Scientist at Awake Labs, you will inevitably have to explain your work to your teammates and other stakeholders. We will be paying close attention to the comments you leave in your code and your use of visuals. We will also look at the summary of your EDA, your justification of your model choice, as well as the answer to the questions listed in bullet point 3.

2. **Creativity**
   We do not expect you to come in to this challenge with a deep understanding of the factors that may lead to a heart attack. We will be paying close attention to the insights you can draw from the data, and how you make use of online and offline resources to guide your approach.

3. **Code Quality**
   We need someone who is confident in their abilities and who can be relied upon to complete their work effectively. We will be looking at the quality of your code and whether your code can run on our machines.

Data Dictionary

Attribute	Description
age	Age of the patient
sex	Sex of the patient
cp	Chest pain type ~ 0 = Typical Angina, 1 = Atypical Angina, 2 = Non-anginal Pain, 3 = Asymptomatic
trtbps	Resting blood pressure (in mm Hg)
chol	Cholestoral in mg/dl fetched via BMI sensor
fbs	(fasting blood sugar > 120 mg/dl) ~ 1 = True, 0 = False
restecg	Resting electrocardiographic results ~ 0 = Normal, 1 = ST-T wave normality, 2 = Left ventricular hypertrophy
thalachh	Maximum heart rate achieved
oldpeak	Previous peak
slp	Slope
caa	Number of major vessels
thall	Thalium Stress Test result ~ (0,3)
exng	Exercise induced angina ~ 1 = Yes, 0 = No
o2Saturation	Blood oxygen saturation (%)
output	Target variable