

GoHealth: Data Science Exercise

Background Information

Your company, Acme Co., sources candidates for companies hiring new employees. Recently, a number of our clients have complained that candidates have not been showing up to interviews. Your boss has provided you with the attached data set in hopes that you can find some way of identifying candidates at risk of not attending scheduled interviews

Task 1

- a. Create a model predicting if a candidate will attend an interview. This will be indicated by the "Observed Attendance" column in the data set. Create the model only using the records where this column is not null
- b. Provide a probability and a prediction for the candidates where the "Observed Attendance" column is null.

Task 2

A client has scored a candidate with your model and it gave the candidate a 30% chance of attending the interview. However, the candidate did come to the interview. The client would like to know why there is this apparent discrepancy in your model. How would you explain this occurrence? Could you provide a better way for the client to evaluate your model's performance

Task 3

Describe how you would create a simple application or RESTful API that returns back predictions given the required parameters. A few sentences or bullet-points will suffice.

Submission Instructions

1. Include the following in your analysis packet:
 1. The code used to create your model in either a text document or jupyter notebook
 2. A csv containing the predictions from Task 1b. Make sure to include the "Name(Cand ID)" column, the probability of attendance, and the predicted attendance (yes or no)
 3. A few paragraphs in response to Task 2,
 4. Program files, unit tests, build scripts, etc. that are related to Task 3.
2. Put all the files in one folder then ZIP it up into a single file.
3. Send the ZIP file back to your contact.

Good luck and have fun!

Notice

1. The company names and individuals referenced above are all fictitious.
2. The content is GoHealth's property. Do not distribute.

Column Descriptions

Data Set*

*Original data set is property of Kaggle user VishnuRaghavan. It can be accessed at <https://www.kaggle.com/vishnusraghavan/the-interview-attendance-problem/data>