Data scientist - Technical check-up

Thanks for your interest in the data scientist position! You will find here some guidelines with the data attached. You will have 5 days to do this technical test. We will plan a 45 minutes session to review it together.

Take-home challenge: free pizza!

The objective is to make this a good and fun way to discover how it would look like working together in real-life.

At CybelAngel, we apply machine learning tools on textual data we find on forums, or other sources where the syntax is approximate. To get a sense of how you would deal with such data, we'd like you to work on this Kaggle Competition :
https://www.kaggle.com/c/random-acts-of-pizza .
You are given a dataset where each row corresponds to a request of a Reddit user, that has triggered a free pizza action (or not). Given all the information provided, you are asked to predict if a certain request is likely to receive a free pizza or not.

How we'll measure success
Free pizza, but no free money. Imagine you are the pizza vendor launching this awesome marketing campaign for people to get to know your brand. You want to minimize the cost of this campaign. Hence, you want to minimize the number of pizzas given to people who did not deserve it.

Guidelines

- Make use of jupyter notebooks, and of any python library you want.
- Prepare a presentation of 10 minutes of your work. Explain clearly the hypotheses you may have made. Provide supporting code. You can use the notebooks as a support only, or prepare slides. Please note that the presentation will also be evaluated if you choose to provide one (if you make some, please send them along with the code).
- Keep the focus on the business value of your work. Being "right" is not the end goal, it's how much value you add to CybelAngel !
- Remember that what we are interested in is not the highest performing model you can achieve, but rather your approach on the problem.

Important note:
Use only the json dataset attached to this email for your study. You do not need to upload your results on Kaggle.