Thank you for applying to our Data Science position! As the next step in our interview process, we have a fun project for you to work on, to help us gauge your skills and thought pattern.

**Problem**

Using a dataset of past advertisements on the Internet, can we accurately predict what image will be an advertisement based on attributes of that image?

**Project**

The features encode the geometry of the image (if available) as well as phrases occurring in the URL,

the image's URL and alt text, the anchor text, and words occurring near the anchor text.

* Number of Instances: 3,279 (2,821 non ads, 458 ads)
* Number of Attributes: 1,558 (3 continous; others binary)
* 28% of instances are missing some of the continuous attributes.
* Missing values should be interpreted as "unknown"
* Class Distribution- number of instances per class: 2,821 non ads, 458 ads.

The task is to predict whether an image is an advertisement ("ad") or not ("non ad").

**Deliverables**

Please send us the following:

1. Code, and associated files, used for the project. You can send us a zipfile, or upload the project to a public github repo.
2. The algorithm you developed to make your predictions
3. How we can run the algorithm on a test data set
4. The process you used to analyze the data and came to your conclusions