pager

Task and Logistics

You will have 7 days to complete this task, but you don't need to spend more than 4 hours on it. Walk us through your assumptions and thought process, as well as your modeling and tool selection decisions. We suggest using a jupyter notebook to present your results.

Pager's Honor Code

By submitting this exercise, I agree that I will submit my own work and only my own work. I will not submit the work of any other person. I will not engage in any activity that would falsify or misrepresent my results or the results of others.

Data

Download the Dataset[1]

Questions

1. **[ML Concepts]** You'll notice that a few hundred beer recipes have no Style assigned to them. Can you use ML to predict these missing values? Please describe the approach/algorithm you'd take and justify your decision.
2. **[ML Concepts - Bonus]** *IndieBrew* wants to use this data to analyze their users' preference in beer recipes and build a recommender system to help users explore new ones. Briefly describe the ML approach you would take to recommend new recipes to a user based on:
    a. Their own submitted recipes
    b. The current recipe they are viewing
3. **[ML Application]** Select one of the problems you've discussed above, and train a model to solve it. Due to the time constraint of this test, we do not expect the best performance out of this model. However, please do include your thoughts on your model's performance and how that could be improved.
4. **[Systems Design]** *IndieBrew* loved your Style predictor's performance and is excited to deploy this model in production. Assume that there is a monitoring service in the backend which will send alerts when it detects anomalous events such as a new recipe submitted without a Style. Please describe the possible approaches you can take to integrate your ML model into this existing architecture. Which one is the most appropriate in this situation? Why?
5. **[ML Engineering]** Implement a barebone *working* version of the service you described in #4. NOTE: if you have trained one of the recommender models for #3 instead, simply replace your model with a dummy one that follows scikit-learn's API.

[1] A link to the data can be found here: https://www.kaggle.com/jtrofe/beer-recipes/data