Machine Learning Engineer Assignment

Every month, there are thousands of events active on TicketSwap websites. These events range from music festivals to concerts to sporting events, stand up comedy shows and more. To help our users find the events they would like to attend, we try to extract the event lineup, as well as any additional information, from the event titles.

Part 1

Attached with this assignment is a list of event titles, ***event_titles.txt***, and a list of artists from our database, ***artists.txt***. For the sake of this assignment, we do not differentiate between artists, stand up comedians, football clubs, or festival names. They all are listed together in the same *artists.txt* file.

We would like you to build a solution, given an event title, it would return the potential lineup (artist names) or equivalent (football teams, festival names, etc).

Your solution should consider the following:

- Not all artists are listed in our database. Think of ways to extract ones not in our database yet.
- The same artists can have different spellings, think Pink vs P!NK.
- The same name can refer to multiple things, think Arsenal (football club) vs Arsenal (band)

Bonus question:

- What additional information can you extract from the event titles? What additional business value do you think this information can provide?
- Can you differentiate between your confidence levels for when different artists are extracted?

Part 2

Once your solution is ready, other parts of our product will need to communicate with it. What is the best way to make your predictions available for the rest of the product to consume? Will your solution differ in case the predictions are needed in real-time vs offline batch jobs? What are the different deployment options you can think of?

Please build a very simple prototype for making your predictions available on demand to our different microservices?

Part 3

Your prototype proved to be successful and you want to push it forward. How would you approach other teams to put it in production, and make sure they consume your predictions? How will you convince the different business stakeholders about your solution? What additional documentation will you need to communicate with the corresponding technical teams? Will you need to collect additional data for your future iterations on your solution?

Deliverables

Please submit your code for parts 1 and 2, in addition to a clear writeup for all three parts. In the writeup, you can explain your reasoning and decisions made. You can also explain your future directions, for example if you have advanced ideas that you did not have time to implement now but would like to explain here. We would like to get a better idea for how you think and your communication skills.

Best of luck to you