Data Scientist, Conversational AI- Written Exercise

Instructions

This is very open ended exercise and there is no correct answer, so please don't spend too much time deciding key aspects on it. For example we are happy to use the data of your choice, an algorithm, and to use any REST API you want.

Please find the details below, try not to spend more than 4 hours on this project

Practical Exercise

The goal of this assignment is to create an NLP model to classify text to one of 5-10 different classes (or intents).

- You may use a dataset of your choice (publicly available, or Kaggle data may be sources).
- Should be programmed in Python
- For each class, assemble a set of training data and apply an algorithm of your choice to determine the best match for some new text input.
- Output results from some evaluation to demonstrate the performance of your model*
- Implement a simple REST API to enter new input, and have a predicted class returned.
- Upload your solution, and all documentation to a Github repository, and invite us as collaborators (Github: Cionad) Include instructions how to run your solution.

Example flow: Classifier to detect topic: movies, games, books, … etc.
Input: "Spyro the Dragon"
Output: {"class" : "games", "confidence" : 0.53}

*Note model performance is not the key goal here, we want to see how you approach a problem, engineer a solution, validate and document the process.

You may be asked to present your work and answer questions in a follow up interview.