Interview Exercise: NLP

Task Overview

In this exercise, you would be given a dataset of natural language text scraped from websites. Your goal is to build a classifier to take text and associate it with a brand. This is an open-ended exercise, and you are free to implement the method of your choice.

Programming Language

You can use a programming language and any open-source library.

Dataset

JSON file with brands and text from webpages.

1. 'raw_text' is your input text to the classifier
2. 'brand' is the output label from the classifier

Output

1. Build a natural language classification model
2. Build a server on top of the classifier to return results via a curl request

Example:

```
Server Call:

curl -d '{"text": "Netflix  English Ukraine Swedish Sing In Unlimited
films TV programmes and more Watch anywhere Cancel at any time Ready to
watch Enter your email to create or restart your membership Email
address Get Started  Enjoy on your TV Watch on smart TVs PlayStation
Xbox Chromecast Apple TV Bluray players and more Download your
programmes to watch offline Save your favourites easily and always have
something to watch Downloading Enjoy on your TV Watch on smart TVs
PlayStation Xbox Chromecast Apple TV Bluray players and more Enjoy on
your TV Watch on smart TVs PlayStation Xbox Chromecast Apple TV Bluray
players and more Frequently Asked Questions What is Netflix Netflix is
a streaming service that offers a wide variety of awardwinning TV
programmes films anime documentaries and more  on thousands of
internetconnected devices You can watch as much as you want whenever
you want without a single advert  all for one low monthly price Theres
always something new to discover and new TV programmes"}' -H "Content-
Type: application/json" -X POST http://0.0.0.0:55555/brand

Output returned:

[{"brand": "netflix", "probablity": 0.82}]
```

Deliverables

Upload the following to GitHub and provide a link:

1. Source code with MIT License at the beginning of each file: https://opensource.org/licenses/MIT
2. Output results of your test/ validation data-set
3. Short writeup of your methodology describing each step in your program.
4. README with instructions on how to run your program with a list of all dependencies (libraries etc.)

Guidelines

1. Explore the data & drive presentations/ recommendations

2. Build a model
    a. Summarize model performance results
3. Build a simple server to hold the model and return results
4. Present your plan to
    a. Deploy model to production
    b. Maintain
    c. Continuously train