README file

geoLondonSep22.json file contain geo data for London

each line is a json object; you may read line by line and process json file

**the fields of interest are:**

text and coordinates

{'\_id': 1574487483722895360, 'date': datetime.datetime(2022, 9, 26, 19, 54, 3), 'screen\_name': 'LaytonWilliams', 'qualityScore': 0.7195228758169935, 'text': "It's giving 10/10 \U0001f90c I taught them well! New Class K #BadEducation @ London, United Kingdom https://t.co/XlwiLtf6pa", 'url': 'https://t.co/AWTcLoBZdN', 'description': 'Actor • Dancer • Singer • Writer @CBGActors 🎭 @TheJoneseVoice 🗣@ProFromTheShows 🕺🏽 (He/Him)', 'listed\_count': 130, 'user\_created': datetime.datetime(2009, 8, 9, 22, 9, 35), 'geo\_enabled': True, 'favourites\_count': 13630, 'coordinates': [-0.1094, 51.5141], 'location': 'London', 'followers\_count': 49534, ' friends\_count': 49534, 'place\_name': 'City of London, London', 'place\_country': 'United Kingdom', 'country\_code': 'GB', 'place\_coordinates': [[[-0.112442, 51.5068], [-0.0733794, 51.5068], [-0.0733794, 51.522161], [-0.112442, 51.522161]]], 'hashtags': ['BadEducation'], 'mentions': [], 'source': 'Instagram'}

credModelFiles – this is for task 2

this directory contains three files

1.highQuality.json

2. lowQuality.json

These files contain data in

{"\_id":{"$oid":"623cb1491fdfc2cfea3b85c4"},"username":"miguel\_zavalaa","description":"•Mind with no sense in it• •Stay away from me ‘cause I’m dancing to quite a different drumbeat•","qualityS":0.6081996434937612,"text":"Turisteando (@ Piccadilly Circus in London, Greater London) https://t.co/ear25pmZNc https://t.co/528VOdxIeF"}

“\_id” is unique tweet id

“text” is tweet text

“qualityS” is the credibility score computed by the algorithm we discussed in the class

For the task ii; we need only the text data (you are free to exploit credibility score if you wish)

Process these text data and build a dictionary

Third file contains the background data - For legacy reasons this is in a different format

{"\_id":"1613006911946776576","text":["confidence","dont","fix","lost","need","negative","playing","right","support","well"],"score":0.6367300158372926}

“\_id” is unique tweet id

“text” is tweet text

“score” is the credibility score computed by the algorithm we discussed in the class

You can process the text field by reading into a dictionary