Step 1 : Data Extraction from Guardian APIs

To use this API, you will need to sign up for API key here:

https://open-platform.theguardian.com/access/

Step 2 : Create a model using the extracted data

create the pipeline model using model.ipynb . This file will save the model in a folder and will show its accuracy .

Step 3 :

Create a topic and then start it .

Step 4 :

Run the producer.py to create a kafka stream data

Step 5:

 Run consumer.py file to read the stream , load the created model and generate the prediction on the fly. Then send data to elasticsearch

⚠️   Please, add the saved model full path in the consumer.py file.

⚠️   you have to specify the ip address of your kafka broker in : producer.py and consumer.py

⚠️   if you didn't use any separate virtual machine all ip adresses must be set to :127.0.0.1 on producer.py and consumer.py

Extra Tips

• if you are going to use the ELK on docker as we did:

 - you have to install the docker engine on your VM/PC and go to elk-docker folder and run "docker-compose up -d " after some time the ELK stack will be ready so you have specify the ip address of the ELK stack

-Note : the default username and password of Kibana are:

*username: elastic

*password: supcom2022

If you want to change the please go to .env in the elk-docker folder

News Data classification