

Instructions to run the program

1. Download the zip file and extract it.
2. Open the terminal and navigate to the path where the folder is extracted
3. Make sure you have python 3 installed and then install the python libraries to run the program
 - a. Pip install pandas
 - b. Pip install numpy
 - c. Pip install nltk
 - d. Pip install matplotlib
 - e. Pip install keras
 - f. Pip install tensorflow
 - g. Pip install wordcloud

4. To run the Logistic Regression on Dataset1 (2 labels), execute LR_Dataset1.py with an IDE or from a terminal with the command:

“python3 LR_Dataset1.py” from inside the corresponding directory

You will need to change the line 22 in the code to the path where you are keeping Dataset1.

5. To run the Logistic Regression on Dataset2 (5 labels), execute LR_Dataset2.py with an IDE or from a terminal with the command:

“python3 LR_Dataset2.py” from inside the corresponding directory

You will need to change the line 22 in the code to the path where you are keeping Dataset2.

6. To run the CNN on Dataset1 (2 labels), execute CNN_Dataset1.py with an IDE or from a terminal with the command:

“python3 CNN_Dataset1.py” from inside the corresponding directory

You will need to change the line 22 in the code to the path where you are keeping Dataset1.

7. To run the CNN on Dataset2 (5 labels), execute CNN_Dataset2.py with an IDE or from a terminal with the command:

“python3 CNN_Dataset2.py” from inside the corresponding directory

You will need to change the line 29 in the code to the path where you are keeping Dataset2.

(Note: These models take time to train as the dataset is big and training is done for multiple epochs. You can see the accuracy of the models after each code finishes execution)

8. Python Kafka Installation guide:

<https://kafka.apache.org/downloads> download the latest stable binary from this link

Navigate to /user/user01 in VM centos machine.

Execute:

```
tar -xvzf kafka_2.11-2.1.0.tgz
```

You will see a folder kafka_2.11-2.1.0

Navigate inside that folder and start zookeeper:

```
bin/zookeeper-server-start.sh config/zookeeper.properties
```

start broker:

```
bin/kafka-server-start.sh config/server.properties
```

Command to list all topics:

```
bin/kafka-topics.sh --list --zookeeper localhost:2181
```

Command to delete topics:

First navigate to bin folder under kafka_2.11-2.1.0

```
./kafka-topics.sh --zookeeper localhost:2181 --delete --topic  
topicname
```

command to check consumer data:

```
bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic topicname  
--from-beginning
```

Blogpost we referred for kafka installation:

<https://www.sohamkamani.com/blog/2017/11/22/how-to-install-and-run-kafka/>

9.To run kafka:

You will need to change the file to update the path to the data files.

Navigate inside the extracted folder.

Execute

```
"Python3 kafka_amazon.py"
```