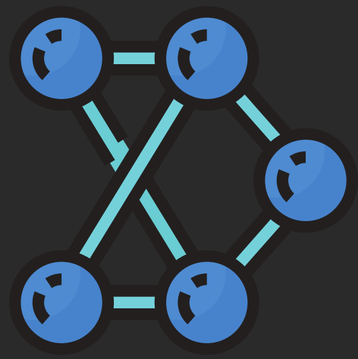
Integrative Project, December 3rd, 2021

COVID-19 tweets classification visualization and generation system

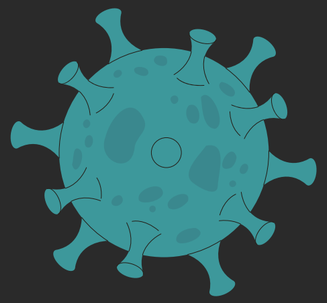
- Mariely del Rosario Nieves González | 1709102
- Ariadna Elizabeth Moo Sosa | 1709097
- Erick Ruben Jaimes Curiel | 1709069
- Nilda Amira Yah Ucan | 1709161
- LLisette Ruiz Peña | 1809150

Introduction



IMPORTANT CONCEPTS:

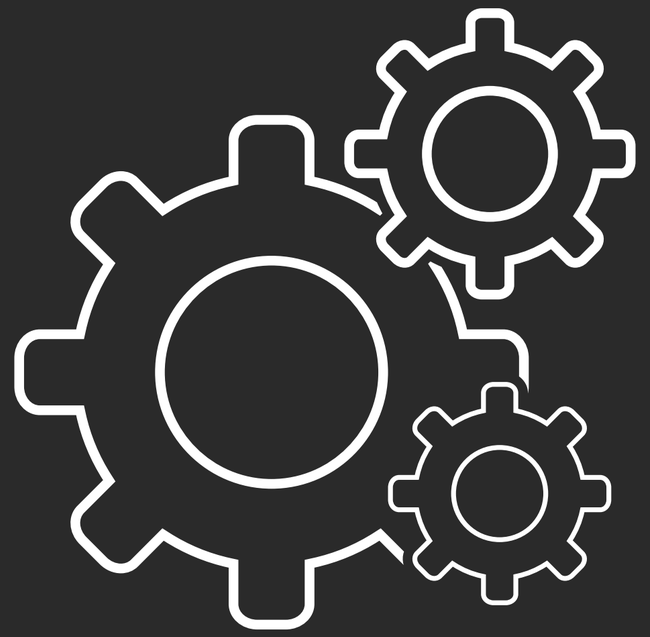
- SENTIMENT ANALYSIS
- TWEETER
- NETWORK



BUSINESS OBJECTIVES

- BUILD A SYSTEM CLASSIFICATION COMMENTS AND BE ABLE TO VISUALIZE THE OBTAINED RESULTS.
- CLASSIFY TWEET TEXT EITHER POSITIVE OR NEGATIVE.
- THE SYSTEM MUST RETURN OR CREATE A POSITIVE COMMENT GIVEN A NEGATIVE ONE.
- BUILD A DASHBOARD TO VISUALIZE IN REAL TIME THE DISTRIBUTION OF THE COMMENTS AND THEIR PREDICTION CLASSIFICATION.

System Design and Technologies



PROGRAMMING LANGUAGE:

python™

LIBRRIES

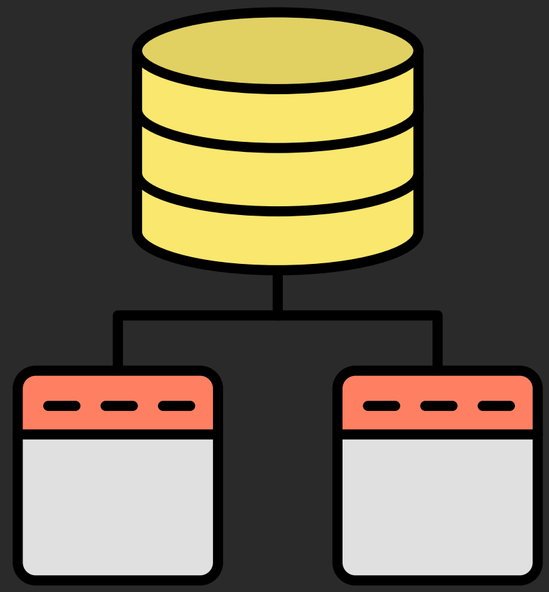
1. SCIKIT-LEARN
2. NUMPY
3. PANDAS
4. MATPLOTLIB

FRAMEWORKS:

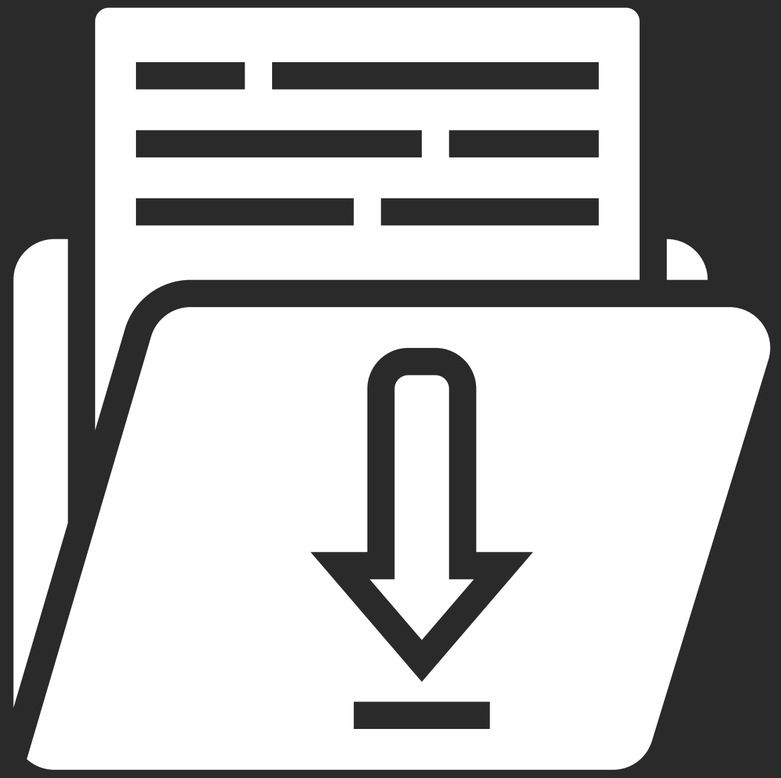
1. KERAS
2. TENSORFLOW
3. TABLEAU

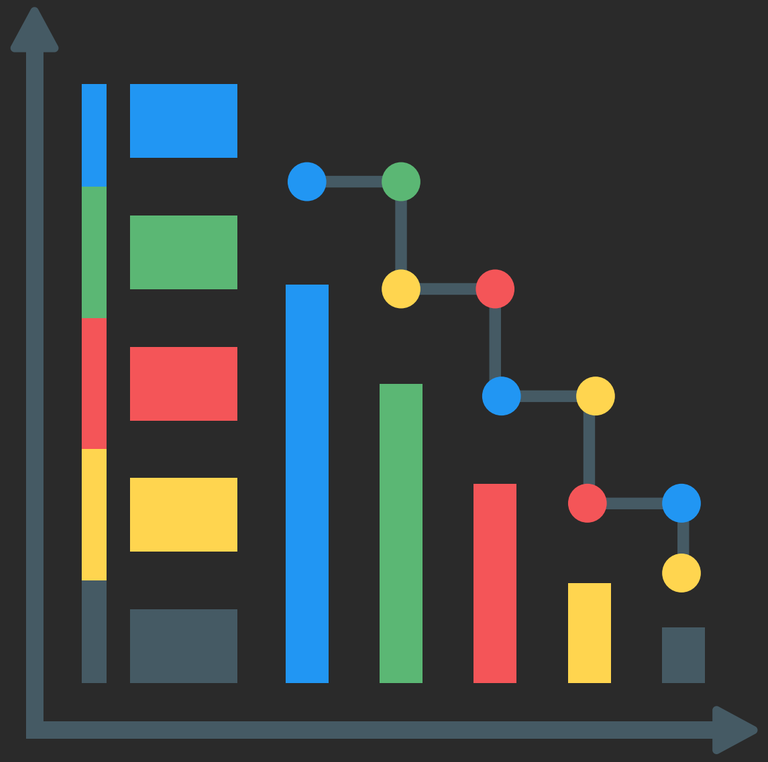
Sequence to sequence modeling

1. Create the user account on Twitter.
2. Request the APIs.
3. Download the tweets.
4. Importing-Applying model.
5. Get the data from Kaggle.
6. Training data.
7. Create the two models (classification and generation).
8. Dashboard creation- toxic to not toxic tweet generation.



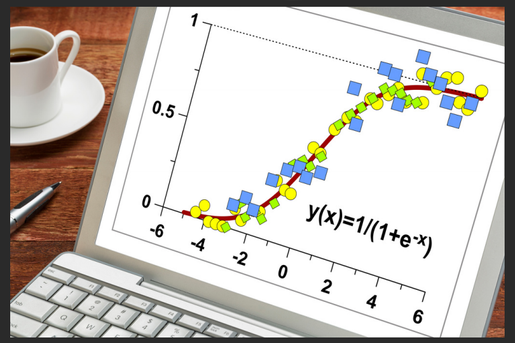
Data Understanding





Experiments models

Logistic regression




```
API_KEY = 'gEfxM0qizKupAtxiJisZyfAHc'
SECRET_KEY = 'LTK9PKTdyxtIPLxFVgguThijjwYmjAN4UkwR3panqgW9oWByjL'
BEARER_TOKEN = 'AAAAAAAAAAAAAAAAAAAAANNMWQEAAAAAseDuRG%2FqCkTwmPokq%2B
ACCESS_TOKEN = '3542954414-euXzt1IRoeg6gYZhb1WF7JgJyB1QTFFf0AFrT7m'
SECRET_TOKEN = '1PGMEabNqgYSx9fIpf0HWkdZqsSVXk3Jd0Ke3zk7oCn1A'
```

Covid tweets from the api

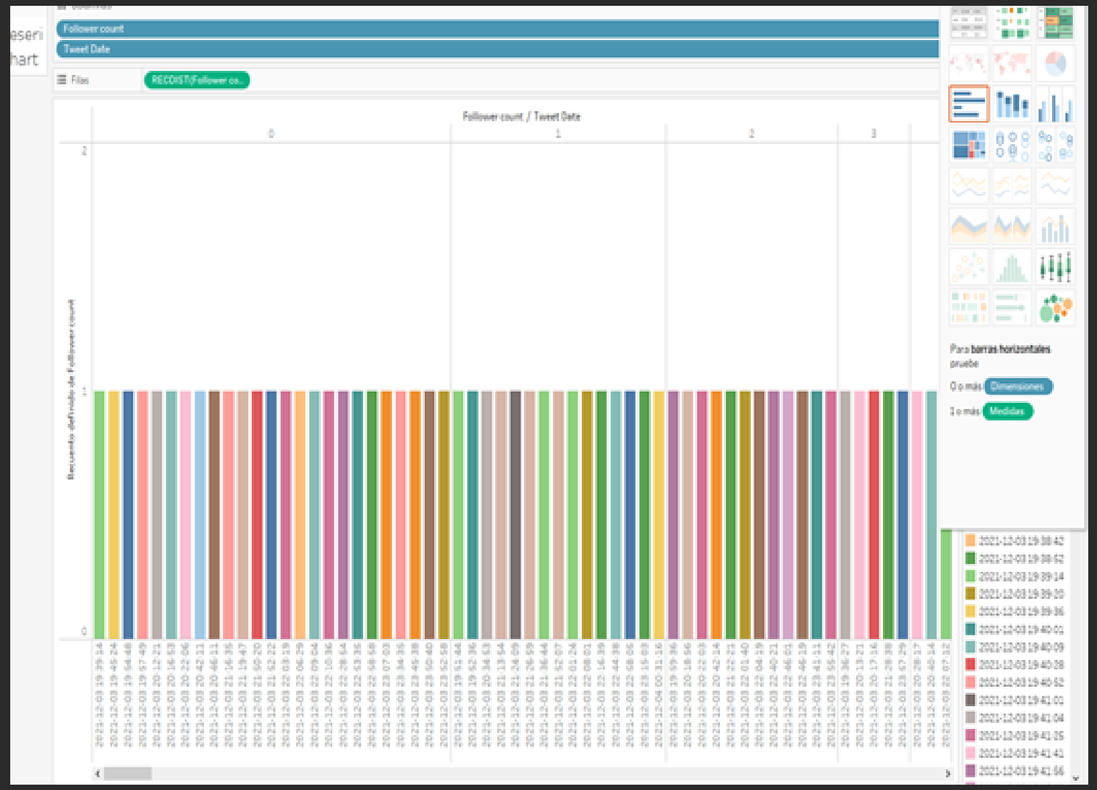
Keras

```
Tweet no. 002
=============
Generating with seed:
 abl polic thiswhat point mask covid
310

_________________________________________
 abl polic thiswhat point mask covid
310,seen amp light twant seed work provid case covid
131,lover bealth alpray best mone treet sone pree
```

Generation char model

Dashboard



Timeseries of follower account of twitter by minute

Heatmap distribution by account and prediction

Conclusions

- Words can be compared to categorical variables.
- Embedding layer enables us to convert each word into a fixed length vector of defined size.
- The resultant vector is a dense one with having real values instead of just 0’s and 1’s.
- The model to be applied, due to its input parameters and its accurancy.
- The model is important to train and save it to import it from another notebook.

REFERENCES

The report has all the references of the project

[1]"TWEET GENERATION WITH NEURAL NETWORKS: LSTM AND GPT-2", MEDIUM, 2021. [ONLINE]. AVAILABLE: HTTPS://TOWARDSDATASCIENCE.COM/TWEET-GENERATION-WITH-NEURAL-NETWORKS-LSTM-AND-GPT-2-E163BFD3FBD8. [ACCESSED: 04- DEC- 2021].

2]"PREDICTING TRUMP TWEETS WITH A RNN", MEDIUM, 2021. [ONLINE]. AVAILABLE: HTTPS://TOWARDSDATASCIENCE.COM/PREDICTING-TRUMP-TWEETS-WITH-A-RNN-95E7C398B18E. [ACCESSED: 04- DEC- 2021].

[3]G. TANNER, "GENERATING TEXT USING A RECURRENT NEURAL NETWORK", GILBERTTANNER.COM, 2021. [ONLINE]. AVAILABLE: HTTPS://GILBERTTANNER.COM/BLOG/GENERATING-TEXT-USING-A-RECURRENT-NEURALNETWORK. [ACCESSED: 04- DEC- 2021].

[4]K. TEAM, "KERAS DOCUMENTATION: CHARACTER-LEVEL TEXT GENERATION WITH LSTM", KERAS.IO, 2021. [ONLINE]. AVAILABLE: HTTPS://KERAS.IO/EXAMPLES/GENERATIVE/LSTM_CHARACTER_LEVEL_TEXT_GENERATION/. [ACCESSED: 04- DEC- 2021].