4.NLP based model to perform sentiment analysis on the coronavirus

Harsh Chandekar

Symbiosis Institute of Technology, Pune

24/08/2021

Run the project on **Jupyter Notebook**.

The packages used are

* tweet-preprocessor
* xgboost
* scikit-learn
* nltk
* numpy
* scipy
* seaborn
* matplotlib

To install these packages

pip install tweet-preprocessor

pip install xgboost

pip install scikit-learn

pip install nltk

pip install numpy

pip install scipy

pip install seaborn

pip install matplotlib

The order of execution is Top to Bottom sequentially.

The Machine Learning Algorithms used for predictions are

1. Logistic Regression
2. Random Forest Classifier
3. Naïve Bayes Classifier
4. Support Vector Machine
5. K Neighbor Classification
6. Stochastic Gradient Descent
7. Extreme Gradient Boosting

Among all these algorithms, Logistic regression has highest accuracy of 62.32%

Using optimal parameters, accuracy improves to 65.09%.

The average time to run the program on Jupyter Notebook is 40 min.