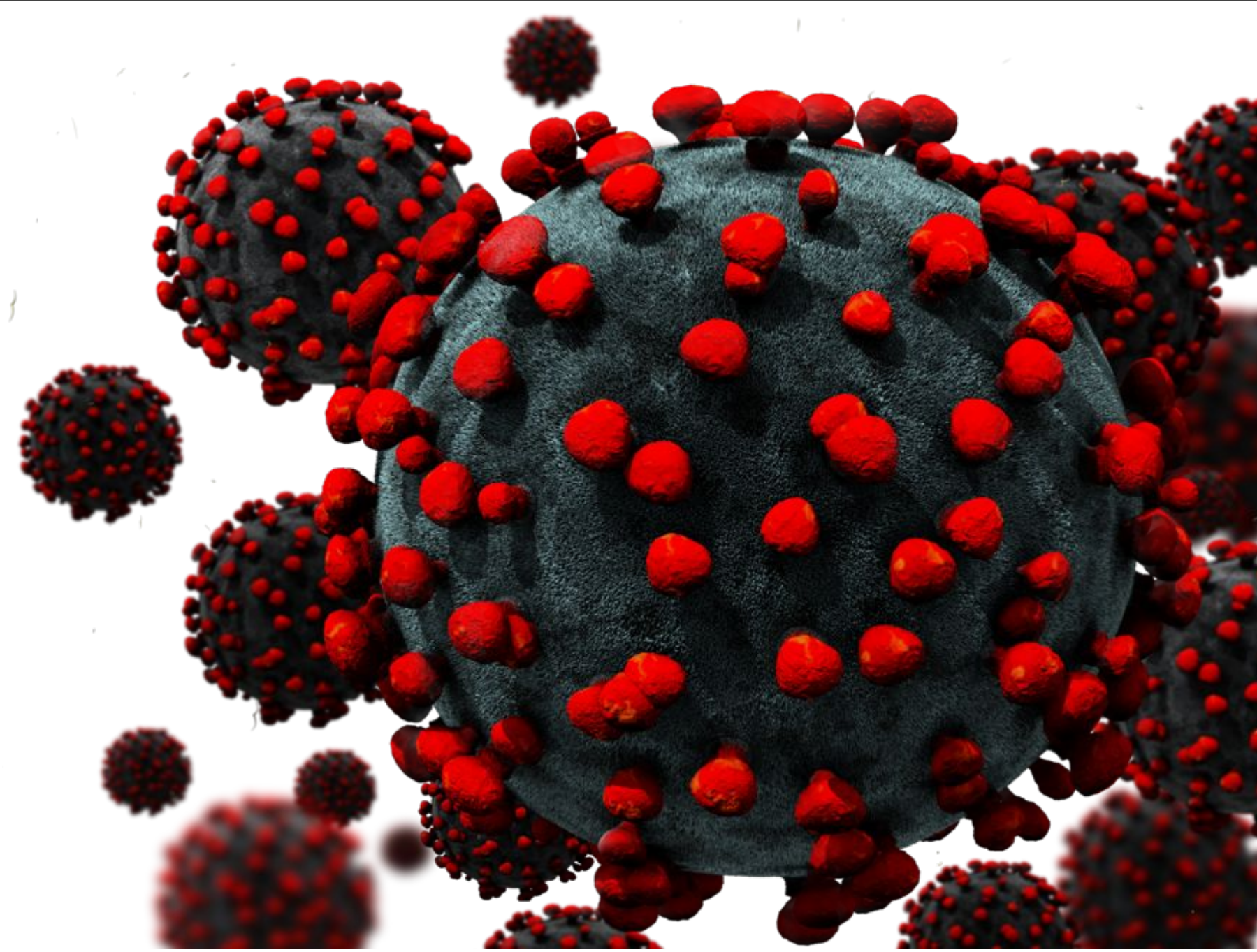# Time Series Analysis of Corona Virus cases in African Countries

Princess Allotey

DSC 305 - Data Science & Analytics Final Project

Introduction

Corona Virus (COVID-19) is an infectious disease which was first discovered in Wuhan, China in December 2019. As at early May, the virus had spread to more than 200 countries worldwide.

I analyzed COVID-19 data from African countries including data on confirmed cases (both daily and total), symptoms, deaths, government measures and school closures. My aim was to explore the trend in COVID-19 cases and what indicators affect the spread of the virus.

Materials and Methods

Datasets were mainly gathered from Our World in Data[1], the European Centre for Disease Prevention and Control (ECDC)[2,], the Centre for Systems Science and Engineering at Johns Hopkins University[3] and the World Health Organization (WHO)[4,].
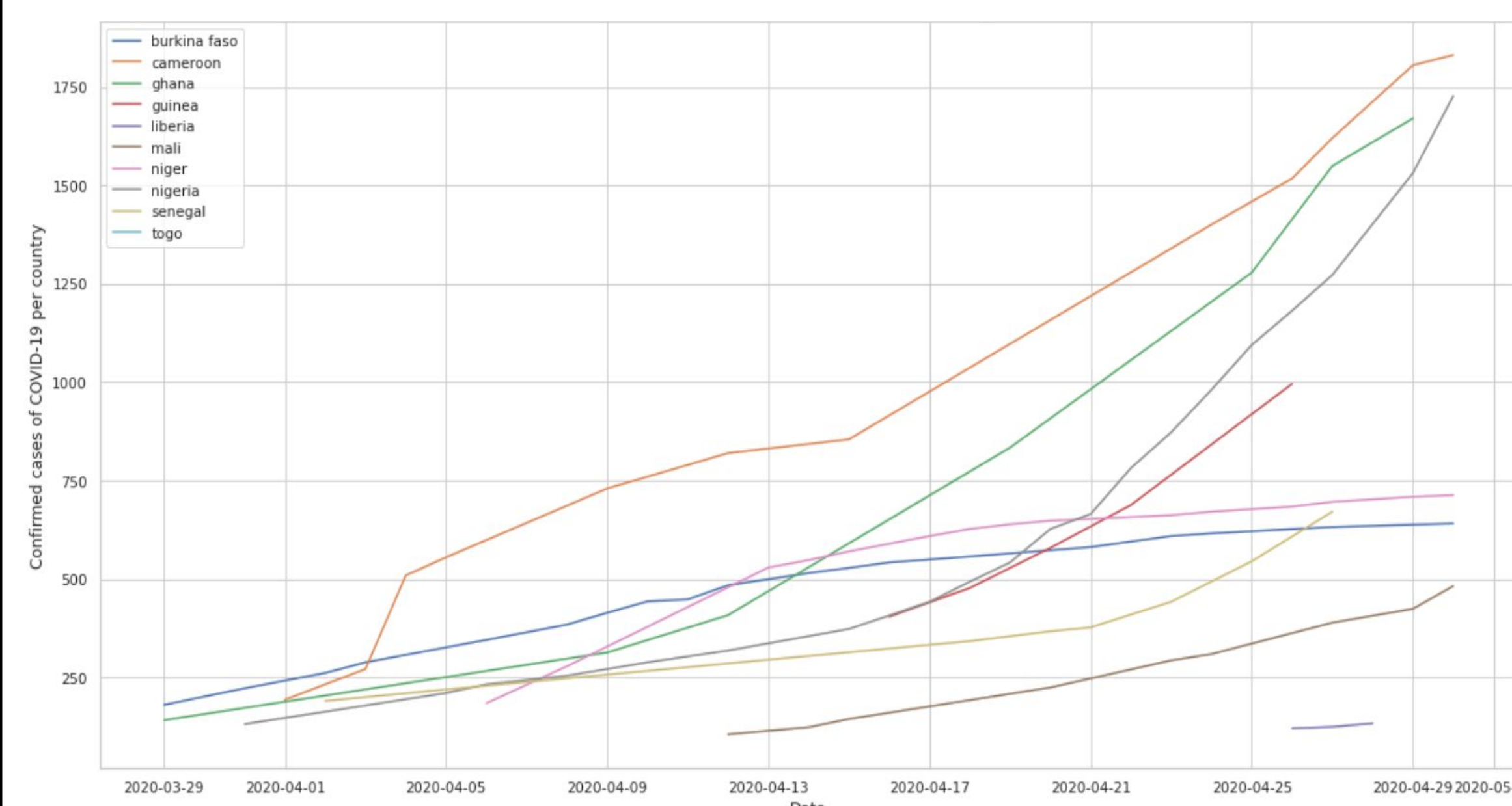
Altogether, I used:
- 19 datasets with time series data
- 8 datasets with indicators data

The machine learning tools I used included Natural Language Processing, Random Forests and Logistic Regression.

Results

1. Confirmed cases of COVID-19 in selected African countries as at May 1st, 2020

2. Using Natural Language Toolkit (NLTK), for text processing, to predict whether a COVID-19 patient will eventually die or survive the pandemic

World
Model Training:
Training set: 331
Testing set: 111
NLTK function:
Ensemble Method Accuracy: 90.09%
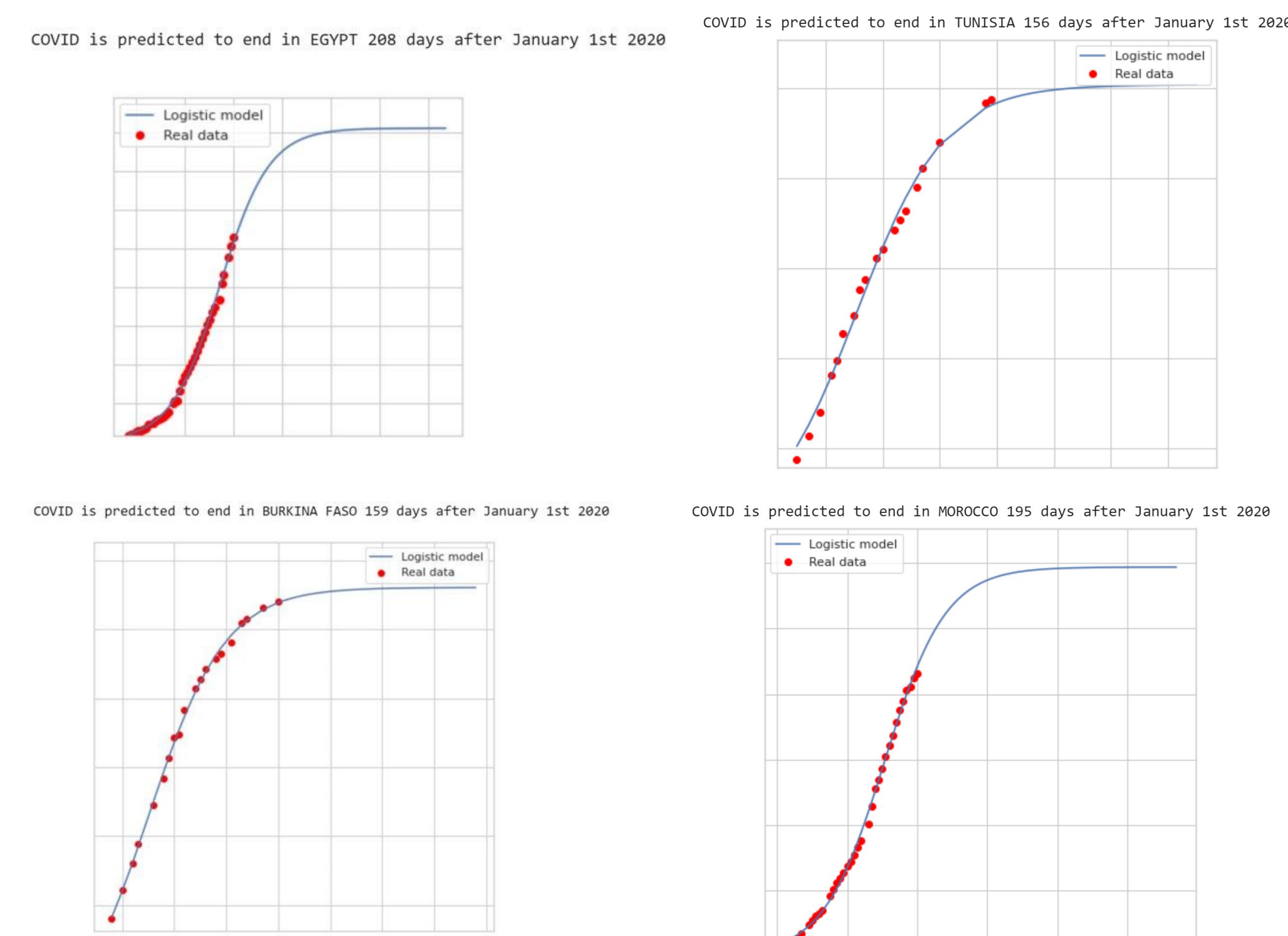
Selected African countries
Model Training:
Training set: 30
Testing set: 10
NLTK function:
Ensemble Method Accuracy: 90.00%

3. Using logistic models to predict when the COVID-19 curve will flatten in selected African countries

Conclusion

I was able to prepare a logistic model, for some African countries, which determined approximate dates for when the curve will flatten. I was also able to determine the outcome of a COVID-19 patient dependent on text data.

Future Work

- Prepare an SIR Model for different African countries.
- Explore and learn from case studies
- Consider more indicators

Acknowledgements

- Professor Thomas Allen
- Centre College

References

1. Our World in Data, https://ourworldindata.org. Accessed 30th March 2020.
2. European Centre for Disease Prevention and Control, https://www.ecdc.europa.eu/en. Accessed 30th March 2020.
3. Centre for Systems Science and Engineering at Johns Hopkins University, https://github.com/CSSEGISandData. Accessed 30th March 2020.
4. World Health Organization (WHO), https://www.who.int/. Accessed 30th March 2020.