

Time Series Analysis of Corona Virus cases in African Countries (January 15th, 2020 - May 1st, 2020)

DSC 305 Final Project

Princess Allotey

princess.allotey@centre.edu

Problem and Motivation

- Pandemic by World Health Organization(WHO) on 30th January 2020
- By early May, the virus has spread to more than 200 countries worldwide
- Focus on Africa because I am from Ghana, and I want to contribute to current research on COVID on the continent
- There is a growing need to learn more about the virus including its mode of spread and its relation to indicators of interest



Data Discovery



Center for Systems Science
and Engineering



Big Idea Questions (simplified)

1. What is the trend in COVID over the days under study?
2. What indicators can predict the likelihood of occurrence of COVID? (age, sex, government measures, school closures, travel history, travel to Wuhan, chronic disease)

*In short, modeling the rate of spread of the disease and its relation to indicators

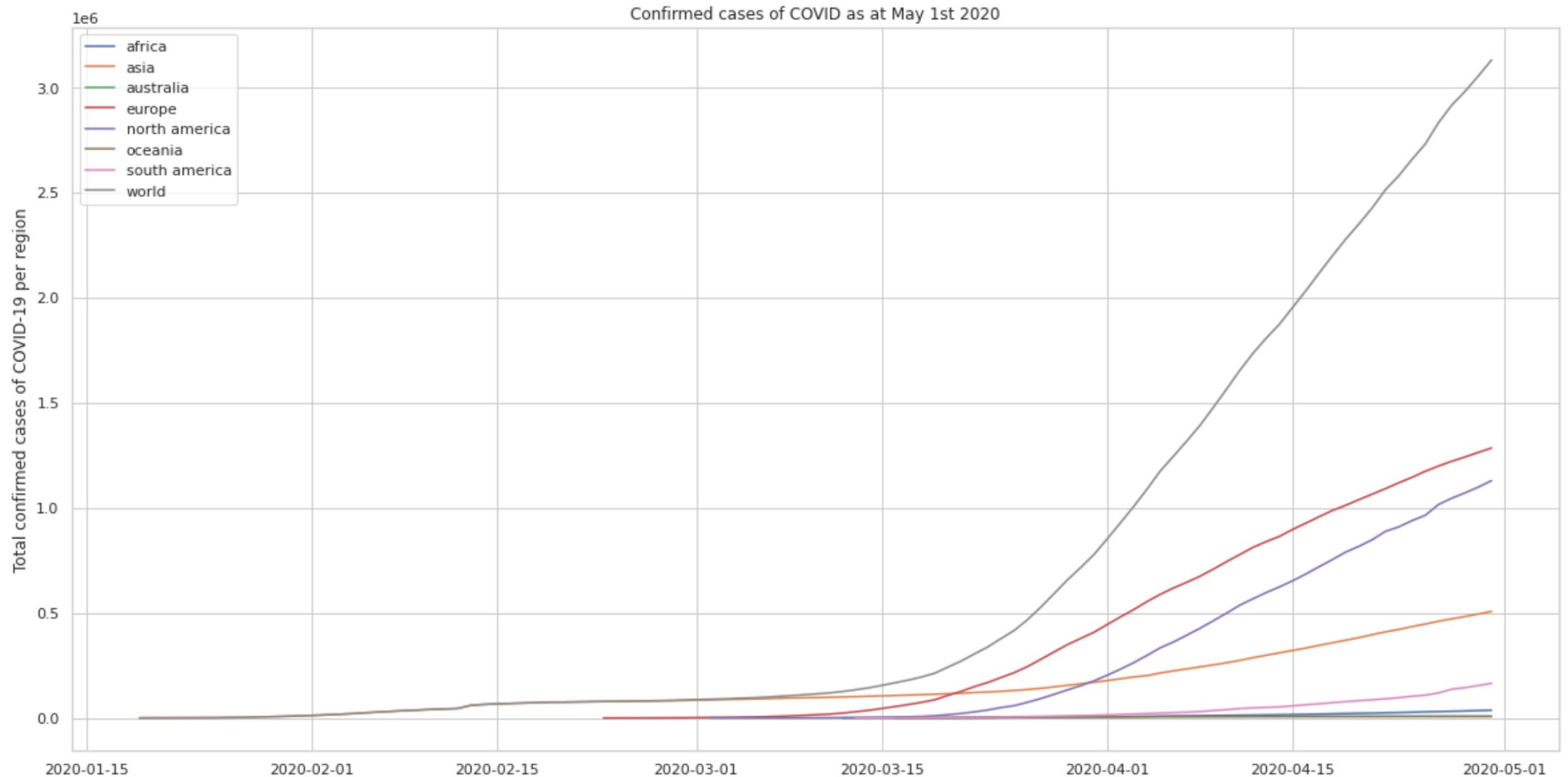
*Still a Work in Progress

- **Time Series Data: 19 datasets**
- **Indicators Data: 8 datasets**

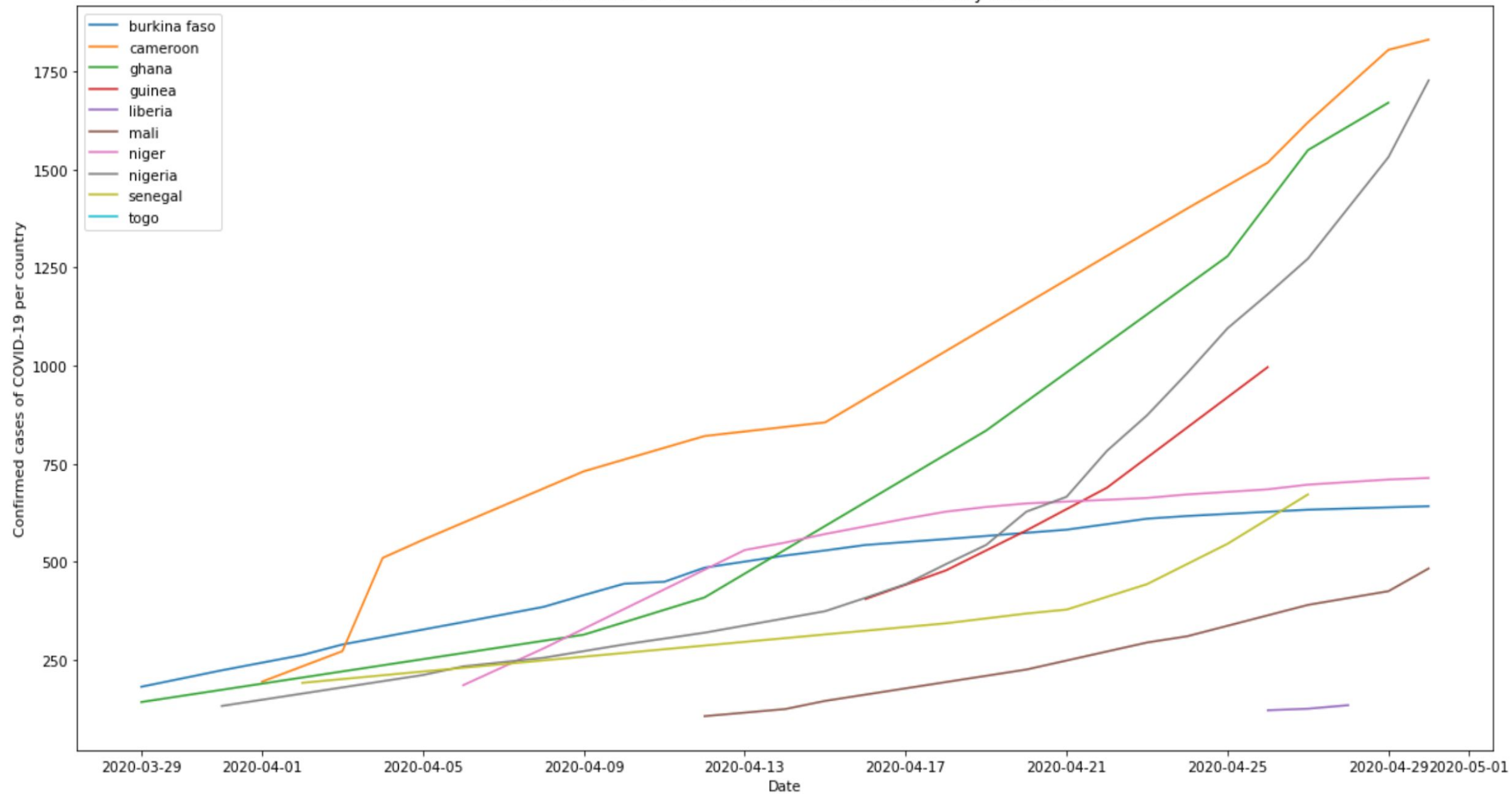
Cleaned data + extracted needed dataframes

*Did not get data for all African countries

Data Exploration with Visuals



Confirmed cases of COVID in African countries as at May 1st 2020



Models Used

- **NLTK with Ensemble method**
- RandomForestClassifier
- LogisticRegressionCV
- **Logistic Model (formula)**

*Will be presenting on only a sample of these

Using NLTK to predict whether a COVID patient will eventually die or survive the pandemic (outcome and additional_information columns)

Using NLTK model to predict whether a COVID patient will eventually die or survive the pandemic (chronic_disease_binary and outcome)

Using NLTK model to predict whether a COVID patient will eventually die or survive the pandemic (symptoms and outcome)

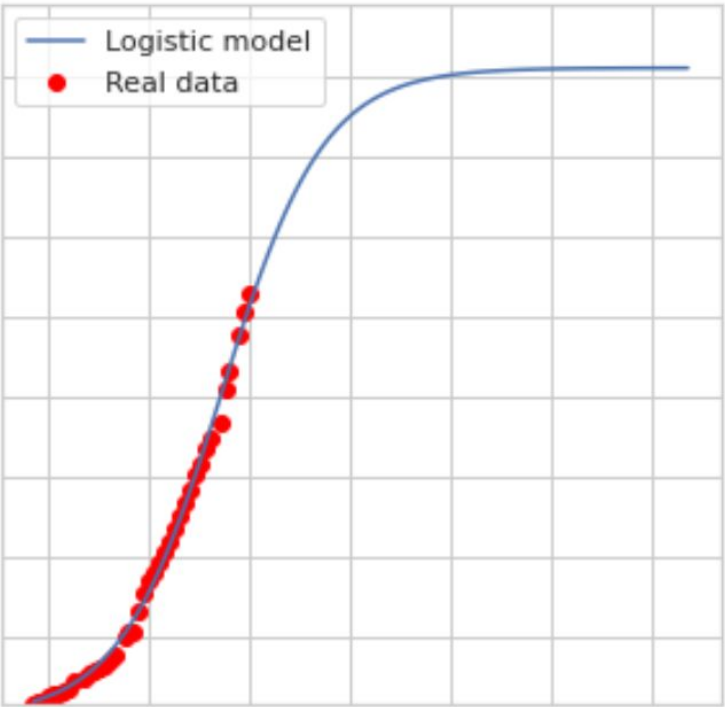
Using logistic models to predict when the COVID curve will flatten in selected African countries

Predicting the likelihood of death from COVID cases

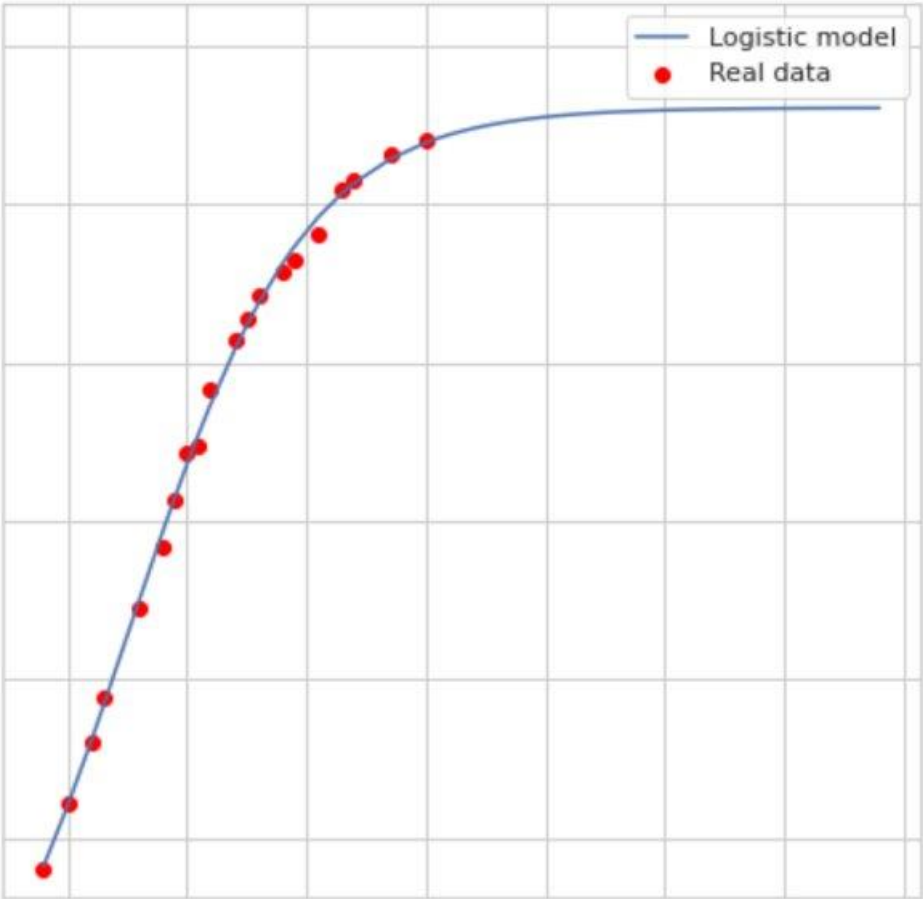
Predicting the effect of current government measures on COVID cases (World)

Predicting the effect of current government measures on COVID cases (Africa)

COVID is predicted to end in EGYPT 208 days after January 1st 2020



COVID is predicted to end in BURKINA FASO 159 days after January 1st 2020



Core of NLTK function

- remove punctuations
- replace whitespace between terms with a single space
- remove leading trailing whitespaces
- change words to lower case - Hello, HELLO, hello are all the same word!
- replace normal numbers with "numbr"
- remove stop words from additional info
- remove word stems using a Porter stemmer

Using NLTK to predict whether a COVID patient will eventually die or survive the pandemic (WORLD)

(outcome and additional information columns) (260,000 rows of data on COVID patients)

Number of words: 390

Most common words: [('numbr', 469), ('case', 276), ('phnumbr', 127), ('hospit', 56), ('china', 45)]

Training: 331

Testing: 111

K Nearest Neighbors: Accuracy: 84.68468468468468

Decision Tree: Accuracy: 86.48648648648648

Random Forest: Accuracy: 88.28828828828829

Logistic Regression: Accuracy: 86.48648648648648

SGD Classifier: Accuracy: 87.38738738738738

Naive Bayes: Accuracy: 87.38738738738738

SVN Linear: Accuracy: 86.48648648648648

Ensemble Method Accuracy: 88.28828828828829

	precision	recall	f1-score	support
0	0.96	0.86	0.91	74
1	0.77	0.92	0.84	37
accuracy			0.88	111
macro avg	0.86	0.89	0.87	111
weighted avg	0.89	0.88	0.89	111

CONFUSION MATRIX:

		predicted	
		dead	alive
actual	dead	64	10
	alive	3	34

Using NLTK to predict whether a COVID patient will eventually die or survive the pandemic (AFRICA)

```
(outcon Number of words:39
Most common words: [('citizen', 26), ('travel', 15), ('gambia', 13), ('french', 13), ('bangladeshi', 11)]
Training: 30
Testing: 10
K Nearest Neighbors: Accuracy: 90.0
Decision Tree: Accuracy: 100.0
Random Forest: Accuracy: 100.0
Logistic Regression: Accuracy: 100.0
SGD Classifier: Accuracy: 100.0
Naive Bayes: Accuracy: 100.0
SVN Linear: Accuracy: 100.0
```

```
Ensemble Method Accuracy: 100.0
      precision    recall  f1-score   support

     0         1.00      1.00      1.00         2
     1         1.00      1.00      1.00         8

   accuracy          1.00          1.00          1.00          10
  macro avg          1.00          1.00          1.00          10
weighted avg          1.00          1.00          1.00          10
```

CONFUSION MATRIX:

	predicted	
	dead	alive
actual dead	2	0
actual alive	0	8

*Way more alive people
than dead

Future Explorations (this summer!)

- SIR Model (MAT 360 – Differential Equations)
- Case studies
- Consider more indicators:
 - ventilators available
 - number of tests
 - Body Mass Index
 - Immunological Naviety (for example, in North Korea)
 - population density affects the rate of spread of the virus
 - pollution

Operationalize Results

- Governmental or non-governmental institutions in Ghana which might be interested in Data Science
- Considering minoHealth AI Labs in Ghana: democratising Quality Healthcare with innovative and cutting-edge technologies like Artificial Intelligence, Data Science and Cloud computing in Africa.

minoHealth

AI Labs

References

- Pydata, <https://pandas.pydata.org/pandas-docs/version/0.13/visualization.html>, Accessed 30th March 2020
- Toward Data Science, <https://towardsdatascience.com/covid-19-infection-in-italy-mathematical-models-and-predictions-7784b4d7dd8d/>, Accessed 30th March 2020
- Python Data Analysis Library, <https://pandas.pydata.org/>. Accessed 30th March 2020.
- NumPy, <https://numpy.org/>. Accessed 30th March 2020.
- sklearn, <https://scikit-learn.org/stable/>. Accessed 30th March 2020.
- Seaborn, <https://seaborn.pydata.org/>. Accessed 30th March 2020.
- Our World in Data, <https://ourworldindata.org>. Accessed 30th March 2020.
- European Centre for Disease Prevention and Control, <https://www.ecdc.europa.eu/en>. Accessed 30th March 2020.
- Centre for Systems Science and Engineering at Johns Hopkins University, <https://github.com/CSSEGISandData>. Accessed 30th March 2020.
- World Health Organization (WHO), <https://www.who.int/>. Accessed 30th March 2020.
- Be Outbreak Prepare, <https://github.com/beoutbreakprepared/nCoV2019>. Accessed 30th March 2020.