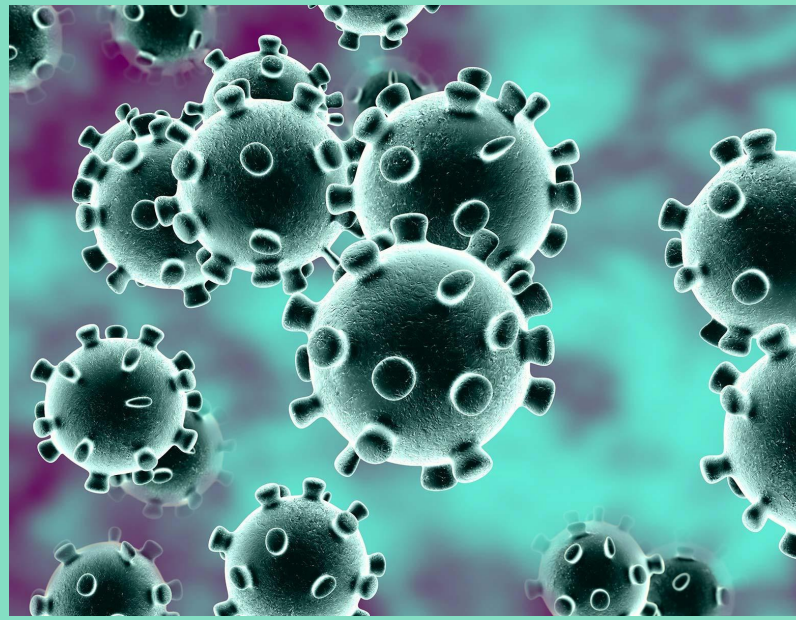
EPIDEMIOLOGICAL TRAJECTORY OF COVID-19

A FOCUS ON AFRICAN COUNTRIES

Princess Allotey

Centre College, USA



BACKGROUND INFORMATION

- The Corona Virus Disease (COVID-19) is an infectious disease which was first discovered in Wuhan, China in December 2019.
- It was declared a pandemic by the World Health Organization on 30th January 2020.
- As at late June, the virus has spread to more than 213 countries worldwide

BIG IDEA QUESTIONS

01

What is the trend in COVID cases over the days under study?

02

What are the patterns in how the indicators in the datasets relate over time?

03

Can we predict the likelihood of occurrence and the rate of occurrence of COVID, in the near future?

04

Can we predict the likelihood of occurrence of a COVID case considering information obtained from existing patients?

Our World in Data

JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING
Center for Systems Science and Engineering

DATA SOURCES

World Health Organization

GitHub

Time Series Data & Indicators Data

SAMPLE OF INDICATORS DATA

- Location
- Confirmed cases
- Deaths
- Tests
- Stringency index
- Population
- Population density
- Age
- GDP per capita
- Extreme poverty
- Death rate
- Diabetes prevalence
- Smokers
- Handwashing facilities
- Travel history location
- Reported market exposure
- Additional information
- Chronic disease binary
- Chronic disease
- Location
- Travel history binary
- Outcome
- In Intensive Care Unit currently
- On Ventilator Currently
- Testing policy
- Life expectancy
- Containment and closure policies
- Hospital beds per thousand

METHODS

01. LOGISTIC MODELS

Fit the trend in COVID-19 cases

02. NEURAL NETWORKS

Learned from time-series data to predict future COVID-19 cases

03. SIRD MODEL

Susceptible-Infected-Recovered-Dead (SIRD) model to examine trend of COVID-19

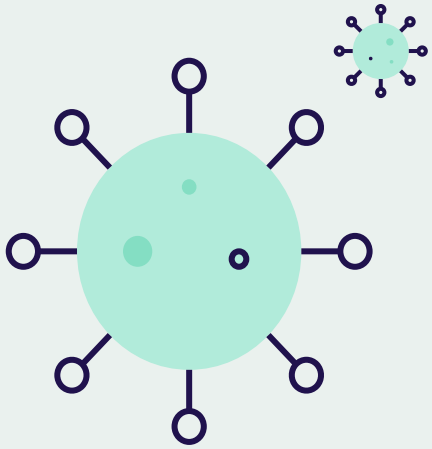
04. NATURAL LANGUAGE PROCESSING

Used text data to predict whether a COVID-19 patient will eventually die or survive the pandemic

*Achieved different levels of progress with these methods

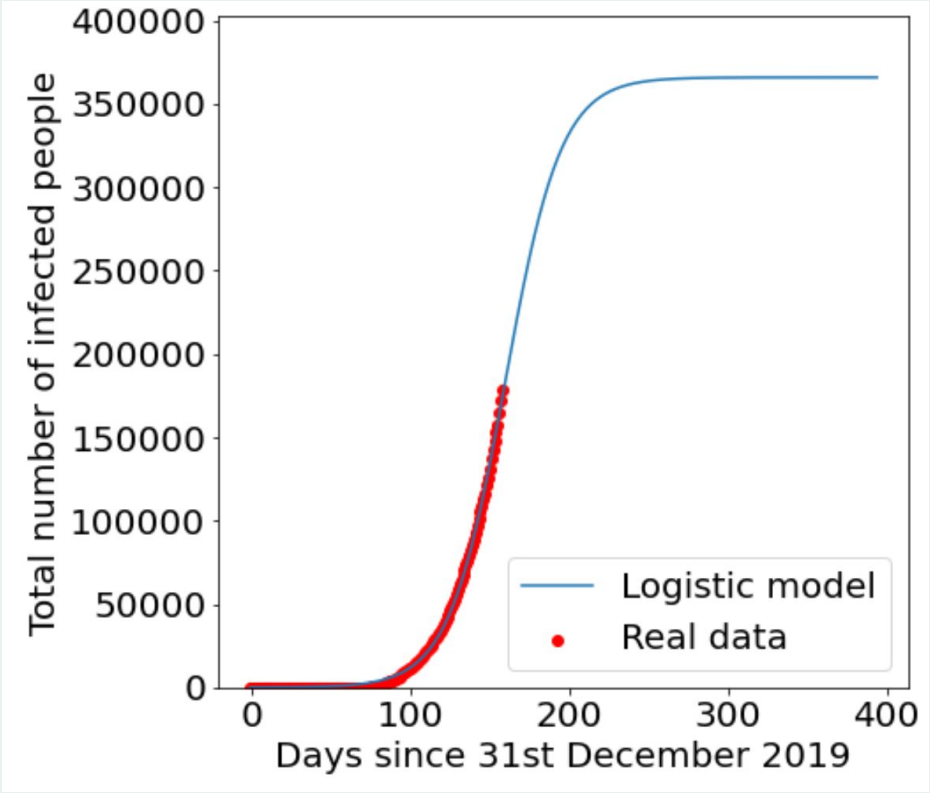
LOGISTIC MODEL

- **Assumption**: The COVID curve is a Logistic curve
- As early as December 31st, 2019 to as late as June 7th, 2020
- **Parameters**:
  - infection speed
  - day when maximum infection occurred
  - total number of recorded infected people at the infection's end
- Graphs that follow show:
  - x-axis: Number of days since first confirmed case
  - y-axis: Total number of infected people



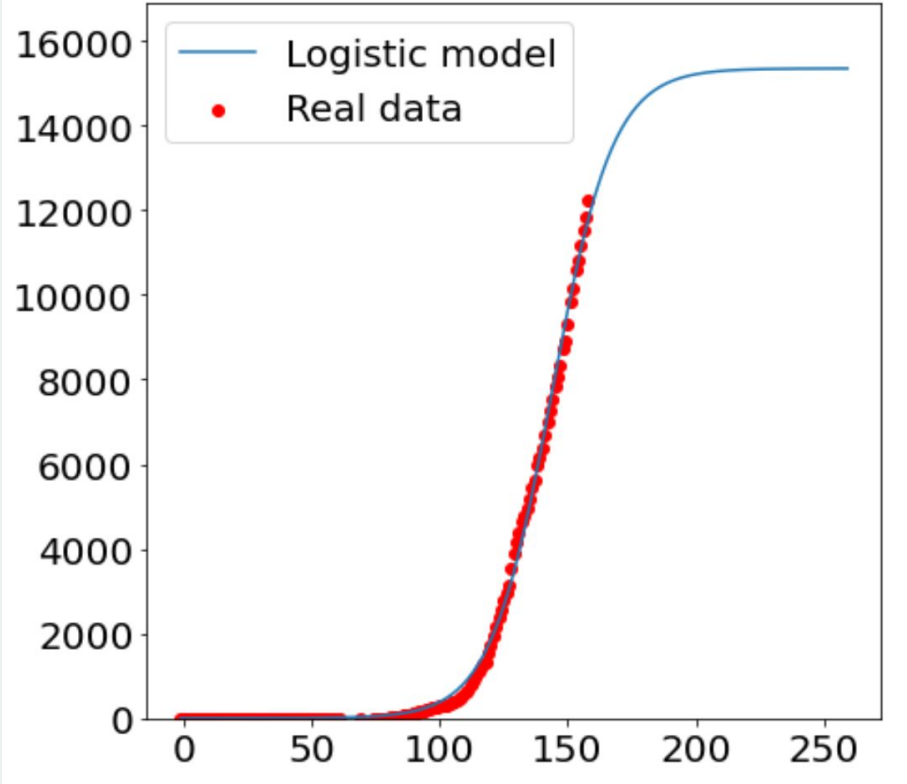
LOGISTIC MODEL FOR COVID IN AFRICA (CUMULATIVE)

- Data from 52 African countries
- First case: December 31st , 2019
- COVID is predicted to end in Africa on January 29, 2021



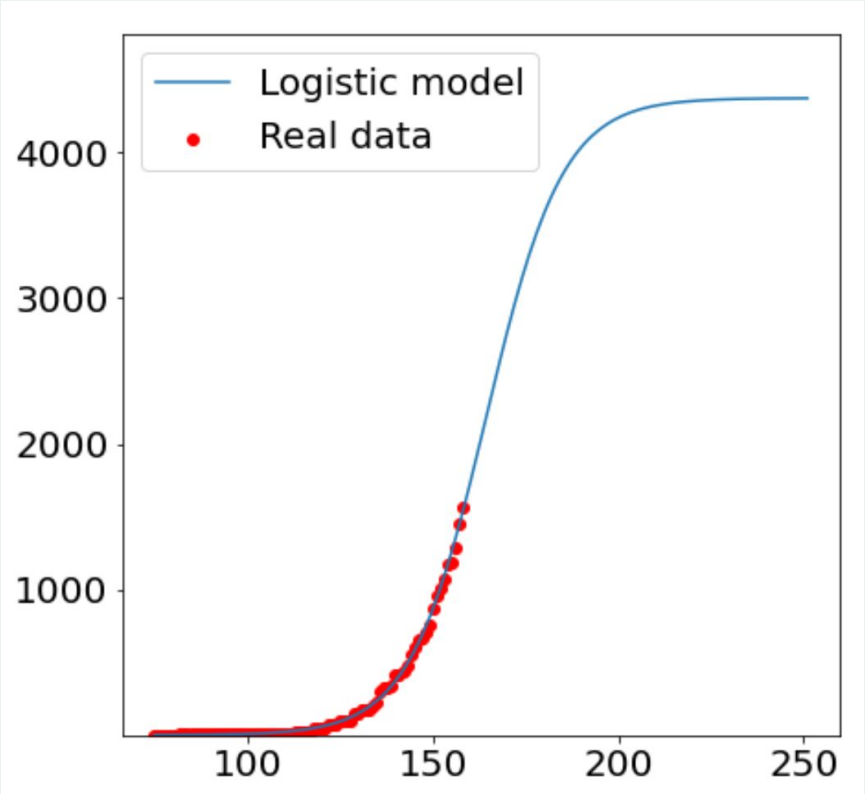
LOGISTIC MODEL FOR COVID IN NIGERIA

- First case: February 28th ,2020
- COVID is predicted to end in Nigeria on November 14th, 2020



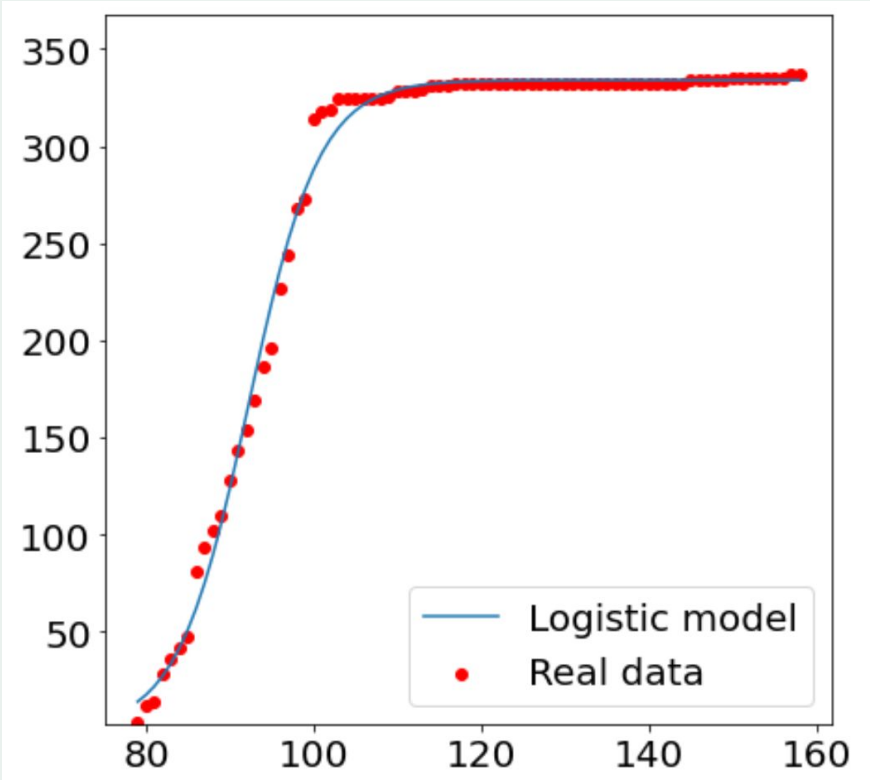
LOGISTIC MODEL FOR COVID IN CENTRAL AFRICAN REPUBLIC

- First case: March 16th, 2020
- COVID is predicted to end in Central African Republic on November 23rd, 2020



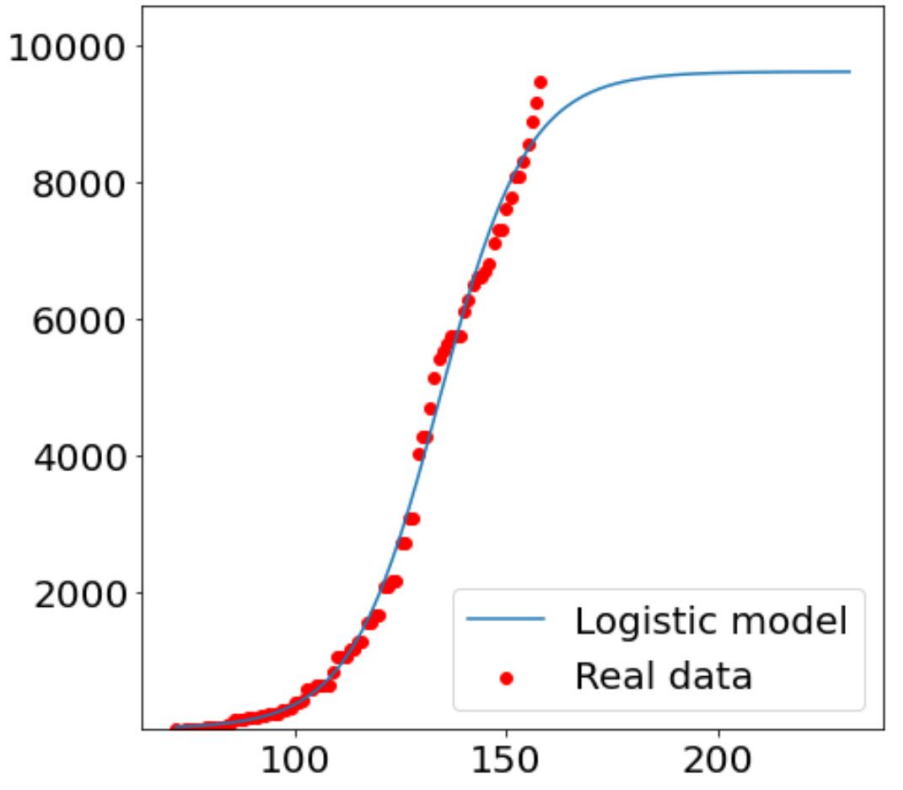
LOGISTIC MODEL FOR COVID IN MAURITIUS

- First case: March 20th, 2020
- COVID is predicted to end in Mauritius on July 15th, 2020



LOGISTIC MODEL FOR COVID IN GHANA

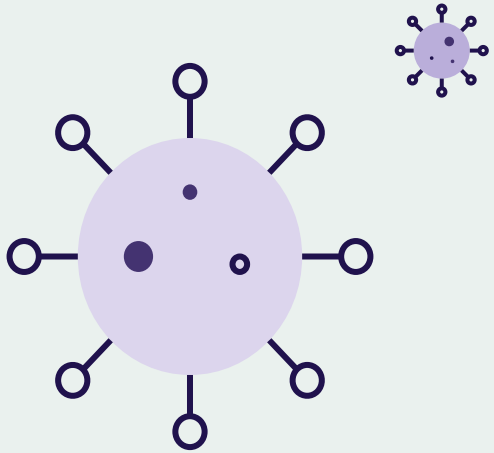
- First case: March 13th, 2020
- COVID is predicted to end in Ghana on October 31st, 2020



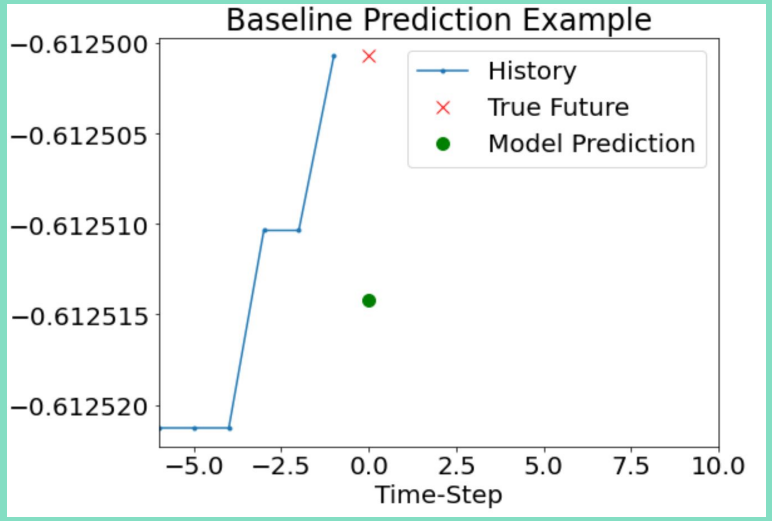
RECURRENT NEURAL NETWORK: LONG SHORT-TERM MEMORY (LSTM)

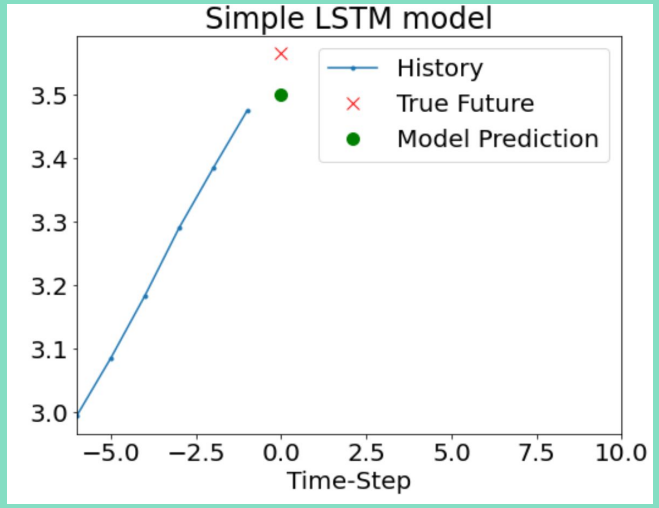
- **Assumption**: The current number of confirmed COVID cases depends on the previous number of confirmed COVID cases

- **Approach**:
  - Trains on a section of available data
  - Validates on a section of available data
  - Set a Baseline Prediction Example
  - Used LSTM to get a better prediction



RESULTS OF RECURRENT NEURAL NETWORK





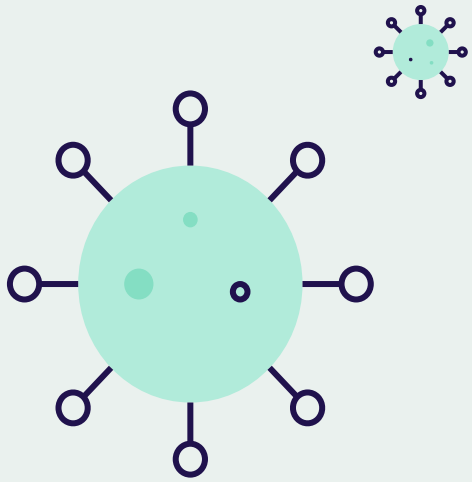
BASELINE PREDICTION

BEATING THE BASELINE PREDICTION

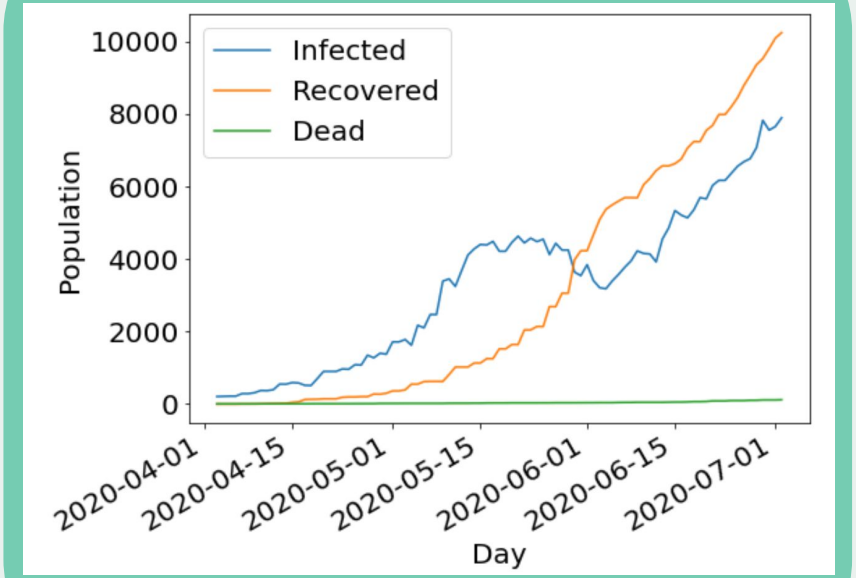
*A Work-In-Progress

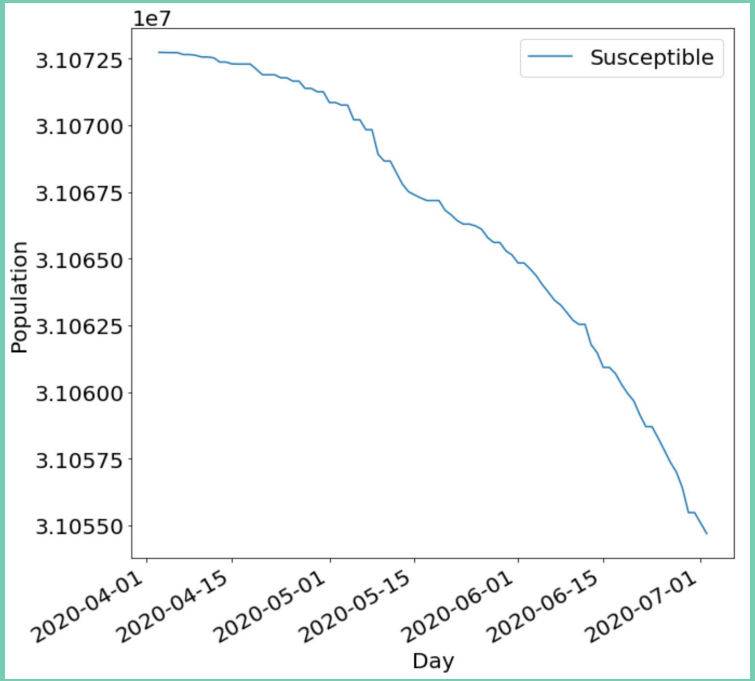
SUSCEPTIBLE-INFECTED-RECOVERED-DEAD (SIRD) MODEL

- **Assumption**:
  - At any time, a person is either in the susceptible, infected, recovered or dead compartment
  - A person can only go from susceptible to infected
  - An infected person eventually recovers or dies three weeks after getting infected with COVID
  - The model is not cyclic



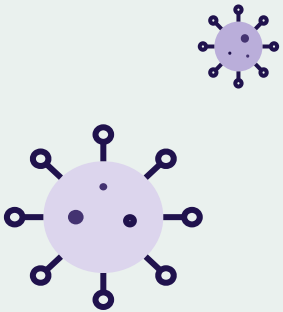
SIRD MODEL FOR COVID-19 IN GHANA





NATURAL LANGUAGE PROCESSING USING THE NATURAL LANGUAGE TOOLKIT (NLTK)

- **Goal**: To predict whether a COVID patient will eventually die or survive the pandemic
- **Dataset**:
  - More than 1,000,000 rows of data on COVID patients
  - Columns used: Outcome & Additional Information
- **Approach**:
  - NLTK function takes in text data
  - Creates a bag of words and identifies common words
  - Trains data on odd number of models and calculates each model accuracy
  - Votes on outcome of each patient (Ensemble Method)



RESULT (WORLD)

CONFUSION MATRIX:

		predicted	
		dead	alive
actual	dead	4298	6
	alive	234	85

ACCURACY: 94.81%

RESULT (AFRICAN COUNTRIES)

CONFUSION MATRIX:

		predicted	
		dead	alive
actual	dead	48	0
	alive	4	5

ACCURACY: 92.98%

CONCLUSION & FUTURE WORK

- There is a growing need to learn more about the virus including its mode of spread and indicators of interest
- Focused on African countries to contribute to current research on the continent
- Logistic Model, Recurrent Neural Network, SIRD Model, Natural Language Processing
- Future work:
  - Use more up to date datasets and develop more rigorous models
  - Expand on this project into the fall semester as my Data Science Capstone Project

REFERENCES

- https://towardsdatascience.com/covid-19-infection-in-italy-mathematical-models-and-predictions-7784b4d7dd8d
- https://www.tensorflow.org/tutorials/structured_data/time_series
- https://colab.research.google.com/github/tensorflow/docs/blob/master/site/en/tutorials/structured_data/time_series.ipynb#scrollTo=0opH9xi5Mtlk
- https://ourworldindata.org/covid-deaths
- https://github.com/beoutbreakprepared/nCoV2019
- https://coronavirus.jhu.edu/testing/individual-states/california
- https://raw.githubusercontent.com/OxCGRT/covid-policy-tracker/master/data/OxCGRT_latest.csv
- https://github.com/OxCGRT/covid-policy-tracker
- https://ourworldindata.org/coronavirus-source-data
- https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/codebook.md