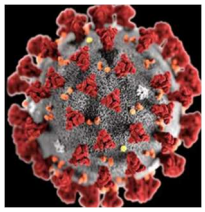

CUNY MS Data 698 – Spring 2021
Subhalaxmi Rout
02/14/2021

TIME SERIES FORECAST ANALYSIS OF COVID-19 CASES

Introduction:

Coronavirus disease is an irresistible infection caused by the respiratory disease coronavirus 2 (SARS-CoV-2). It was first found in Wuhan, China, in December 2019, and has since spread universally, causing a constant pandemic. On March 11, 2020 the World Health Organization (WHO) declared the novel coronavirus (Covid19) outbreak as a global pandemic. We will perform a time series analysis to predict the number of infected cases, and deaths in the United States starting from May 1st — June 30th.

Questions:

We will find the answer of below questions:

- Predict the amount of people infected using time series (more than 7million by July 2020)
- Predict the amount of deaths using time series (more than 1.5million by July 2020)

Other approaches:

We will use other machine learning techniques to predicts number of infection cases, and deaths, then compare the accuracy of the model.

Team Plan:

SL NO	Section	Description	Team member
1.	Research Topic	Search the topic for the project	ALL (working on covid-19 timeseries analysis)

2.	Question Formulation	What problem we are going to solve?	ALL (predict the numbers of cases, deaths etc.)
3	Data Collection and cleaning	Data extract, clean	2 -3
4	Exploratory Analysis	Statistical analysis, understand data and its pattern	1 - 2
5	Hypothesis Generation	Perform hypothesis test	1 - 2
6	Literature Review		2 - 3
7.1	Model Building	Perform time series analysis	1 - 2
7.2	Model Building	Other models to check the accuracy	1 - 2
8	Result Analysis		ALL
9	Journal Paper Presentation	Final paper and Presentation	3 – 4 or (ALL)

Here is my proposal plan for the team, I am open to new approach also. I found some useful resources to help us in our project, given link below.

- https://link.springer.com/article/10.1007/s42600-020-00105-4
- https://towardsdatascience.com/predicting-number-of-covid19-deaths-using-time-series-analysis-arima-model-4ad92c48b3ae
- https://ourworldindata.org/covid-deaths?country=~USA

Note: This proposal I have mentioned about infection cases, and death. If we have bandwidth we can work on prediction of number of hospitalizations, and number of recovered. Here I mentioned the data from May to June, because I do not know how much the data size will be, if the size is not huge, we can go for more months of data.