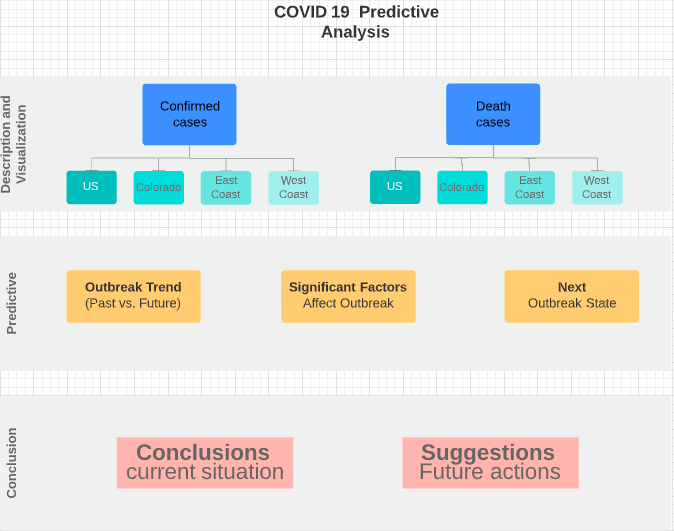
**Project for Coronavrius**

Candice Chuang,Qingdi Xiao, Wenbin Yang, Yuxiang Wang, Ziwen Huang

**ABSTRACT**

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).The disease was first identified in December 2019 in Wuhan, the capital of China's Hubei province, and has since spread globally, resulting in the ongoing 2019–20 coronavirus pandemic. Common symptoms include fever, cough and shortness of breath. Other symptoms may include fatigue, muscle pain, diarrhoea, sore throat, loss of smell and abdominal pain. The time from exposure to onset of symptoms is typically around five days, but may range from two to fourteen days. While the majority of cases result in mild symptoms, some progress to viral pneumonia and multi-organ failure. As of April 12, 2020, more than 1.8 million cases of COVID-19 have been reported in 210 countries and territories, resulting in more than 112,000 deaths. More than 415,000 people have recovered, although there may be a possibility of reinfection. Specifically, in the US, more than 549,000 cases have been reported, more than 32,000 people have recovered, while still resulting in more than 21,000 deaths. In Colorado, 6,893 cases have been reported and 274 deaths have happened.



Our plan includes three parts: (1)Descriptive Statistics with visualization (2) Predictive Models (3) Conclusion and Suggestion. As for the current situations, we plan to compare U.S circumstances based on four units : U.S, Colorado, East coast, West coast

We will summarize both the confirmed and death cases with:

1. Daily, Weekly, and Total
2. The cumulative increasing rate
3. Key indicators, based on daily, weekly, and cumulative

* Death rate
* Recovery rate
* Pos/Neg rate

1. Using Maps to compare the cumulative cases, cumulative positive cases, cumulative negative cases, and cumulative deaths cases of each region.
2. Relate to the other states’ features to get the correlations to find which factors affect the outbreak.

We will predict:

1. Future trends of the COVID-19 outbreak.
2. Significant factors affect the outbreak of COVID-19.
3. The Next possible outbreak states, which states will be the next outbreak states (confirmed) by analyzing speed and acceleration to build models

We also plan to use SparkPython to build the word cloud to find the keywords of this outbreaks by using COVID-19 Open Research Dataset.

**Datasets**

We plan to use three datasets:

1. Tableau COVID-19 Data Hub (<https://www.tableau.com/about/blog/2020/3/coronavirus-data-hub-faq>)

The data comes from the dataset maintained and updated by the Johns Hopkins University Center for Systems Science and Engineering. Tableau cleans, reshapes, and makes this data ready for your analysis. This dataset updated by 9:00am EST every day. This data contains Dates and Confirmed Cases and total deaths for each state by providing for all countries over the word. This data also contains the location (Lat and Long).

1. COVID-19 in USA (<https://www.kaggle.com/sudalairajkumar/covid19-in-usa#us_covid19_daily.csv>)

This dataset has details information from 50 US states and the District of Columbia at daily level. This dataset contains date, total case of states, number of positive and negative cases, pending cases, number of recovered and death, and number of death increasing by daily.

1. COVID-19 Open Research Dataset Challenge (CORD-19)( <https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge>)

In response to the COVID-19 pandemic, the White House and a coalition of leading research groups have prepared the COVID-19 Open Research Dataset (CORD-19). CORD-19 is a resource of over 51,000 scholarly articles, including over 40,000 with full text, about COVID-19, SARS-CoV-2, and related coronaviruses. This freely available dataset is provided to the global research community to apply recent advances in natural language processing and other AI techniques to generate new insights in support of the ongoing fight against this infectious disease.