PREDICTING COVID MORTALITY

Machine Learning-CS314b

08.04.2020

Saumya Mishra(2017230)|Kuhu Pyasi(2017327)|Rushikesh Ghule(2017218)|

# Abstract

Corona Virus(COVID-19) is an infectious disease caused by a newly discovered coronavirus,currently more than 1.4 million people are affected by this disease worldwide.We are trying to predict the state of a person if he gets infected by the virus,i.e if he would survive the virus or not based various data points like age,gender,past travel history,cold symptoms etc.

# Dataset

For the task in hand we have taken the data set for our model from the data collected by the Korean Statistical Information Services.Where they have used following informations-

#### Data of COVID-19 infection cases in South Korea

#### Epidemiological data of COVID-19 patients in South Korea

#### Route data of COVID-19 patients in South Korea

#### Time series data of COVID-19 status in South Korea

#### Time series data of COVID-19 status in terms of the age in South Korea

#### Time series data of COVID-19 status in terms of the gender in South Korea

#### Location and statistical data of the regions in South Korea

#### 

# Proposed Approach

The problem can be broken into the following segments:

1.Understand the dataset available,by performing exploratory data analysis.

2.Gain Insights from the EDA,estimate which factors could affect the net outcome more, perform feature engineering prior feeding the dataset to the machine learning models.

3.Divide the dataset into training,validation and test set and train machine learning models on the training set created

4.Perform cross validation on the validation set and refine the model for optimization

5. Compare performances of various models on the test set and select the best fit model.

# 

# TOOLS & LIBRARIES

1. Python 2-3
2. Following libraries would be used for the above task

|  |  |
| --- | --- |
| LIBRARIES | DESCRIPTION |
| Numpy | used to perform numeric computations on an n-dimensional array |
| Pandas | used to load tabular data in predefined data structure called dataframes and perform computations on that |
| ScikitLearn | used to train all ML models |

|  |  |
| --- | --- |
| matplotlib | used to do plotting in exploratory |