Design and analysis of the epidemic novel corona virus

Nirbhay Mishra¹, Amit Majumder², Dr. Pranati Rakshit³

¹Nirbhay Mishra, M.Tech CSE, JIS College of Engineering
²Amit Majumder, Assistant Professor, CSE Department, JIS College of Engineering
³Dr. Pranati Rakshit, Assistant Professor, CSE Department, JIS College of Engineering

E-mail: mlython28@gmail.com

Abstract. We all know that corona virus first news is get flashed on 31st December 2019 from the Wuhan city of China and Province Hubei is mostly affected by this deadliest version of corona bomb hence it is called COVID-19 named by scientists because it’s completely novel by structure. It’s a respiratory diseases and very much similar to SARS (Severe Acute Respiratory Syndrome) but its molecule structure is not completely similar to SARS hence till now there is no medicine for this daemon CORONA. Currently we can clearly observe that almost all countries of world are suffering from this. At initial stage it is predicted that its animal kind of flu and hence there is no need to worry but as soon as time goes by its started transmitting from animal to men and then men to men and hence transmitted from city of Wuhan and reaches almost all corners of the world, hence it is very dangerous for our civilization and its increasing death rate and suspect cases reveals that it’s a big epidemic that we ever faced in last 100 years. So by observing the severity of this epidemic we need to inform right and exact guideline to every other people. Epidemic is a disease that spread very quickly and covers entire geographical area in no time that what the world are facing off. It can cause significant economic, social, and political disruption. Before this the mankind has experienced several epidemics like SARS, Dengue, Plague, and Spanish Flue which had already ruined the life of world completely in those days when Medical science was not at their pick. But, today when world civilization and modern medical science is at their pick a new RNA structured VIRUS has completely destroyed the world in many ways. So it’s very necessary to know the impact of it on originating venue so that we can analyze its danger and rate of spread ability. So to do this we have applied Machine Learning algorithm and concepts for predicting death and other cases of Covid-19.

1. Introduction

Any epidemic is dangerous and COVID-19 is one of them. We all are very much familiar with such viruses as a SARS, basically corona virus is from the family of SARS but it has some advancement over it featuring some add on respiratory problems in patient. We must be familiar with MERS that’s refers to Middle East respiratory syndrome. Corona is making replica of its own and regularly changes their form, this corona bomb is very aggressive and virulence and can spread a lot faster than a normal flue. At time of SARS before 17-18 years ago 10% have been died and 85% have been badly suffered.
In MERS 1/3 patient are dead, this is a type of corona virus but this new version is more deadly than ever. This is different in sense that the vaccine of last corona is not working on it. The Researcher & scientist from around the world are working on it to invent new vaccine formula and as they found it will be available in retail markets.

Almost all countries have confirmed diagnose cases of CORONA. But the good and preliminary steps been taken by the current government of India (i.e. GOVT. OF INDIA) that the initial checking’s have been pursuing at every Airport of India like passengers specially from China , Japan, US, Hong Kong will be screened and scanned if there is found something suspicious then he/she will isolated for treatment. Till date we have some corona news from India but good thing is that every suspicious of corona is recovered.

Corona acts like a normal viral fever with symptoms like headache, back pain, nausea with cough and cold but it is 10 time aggressive and contagious. If anyone have been suffering from this then it can transmit in 500 people up to distance of 6FT by sneezing or touch by their droplets. If there is something in air like this, we have to take Precaution because we don’t know who is suffering from which disease. So in order to protect our self we have to use medicated N-95 Mask which is a 3 layered and can protect us from such infections, apart from this we have to wash our hands on regular intervals, we have to use tissue paper for sneezing and other things. We have away from those who are suffering from cold and cough. The main problem a corona virus patient is suffering is that at very rapid way they will have problem of breathing means Oxygen level gets decreased and also they will have pneumonia. Which on later stage resulting Multi-organ disorder and patient will just wait at ventilator? From here on recovery is not possible.

At initial stage there is a news from China, that it is a animal to human transmission bacteria so no need to worry and keep yourself away from animals then you will be OK but negative impact have been faced by all countries of WORLD and death rate has been increased dramatically and on later stage it is confirmed that human to human transmission is going on. Hence till date world not get out of this death toll rate.

By the above findings we will try to analyze the death and recovered cases in China and India by applying different Machine learning methodologies and data science statistics with the help of a simple and small dataset of Covid-19.

1.1. Objective
In this paper we analyze and visualize COVID-19 data. We apply machine learning approach for doing prediction on number of deaths or number of recovered cases etc. related to CORONA infection. Our main objectives are:
   a) We try to sketch current scenario of CORONA Infection in china/India.
   b) We have used a small dataset that what we received of India and china.
   c) We try to sketch the real symptoms of it.
   d) We also try to let the people know that it’s spreading and need to be cautious and follow the guideline suggested by Ministry of Health and welfare (India) and WHO.

1.2. Dataset Description
The dataset that we used here has seven (7) different attributes like date, Province, Country, Last updated, confirmed, and Death. We have chosen China and India to visualize the scenario and also perform data analysis and forecasting.

| Sno | Date         | Province/State | Country | Last Update     | Confirmed | Deaths | Recovered |
|-----|--------------|----------------|---------|-----------------|-----------|--------|-----------|
| 1   | 01/22/2020   | Anhui          | China   | 01/22/2020 12:00:00 | 1         | 0      | 0         |
| 2   | 01/22/2020   | Beijing        | China   | 01/22/2020 12:00:00 | 14        | 0      | 0         |
Table 1: Covid dataset

|  | Date       | City  | Country | Time   | Confirmed | Deaths | Recovered |
|---|------------|-------|---------|--------|-----------|--------|-----------|
| 3 | 01/22/2020 | Chongqing | China   | 12:00:00 | 6         | 0      | 0         |
| 4 | 01/22/2020 | Fujian  | China   | 12:00:00 | 1         | 0      | 0         |
| 5 | 01/22/2020 | Gansu   | China   | 12:00:00 | 0         | 0      | 0         |
| 6 | 01/22/2020 | Guangdong | China   | 12:00:00 | 26        | 0      | 0         |
| 7 | 01/22/2020 | Guangxi  | China   | 12:00:00 | 2         | 0      | 0         |
| 8 | 01/22/2020 | Guizhou  | China   | 12:00:00 | 1         | 0      | 0         |

We choose this dataset and its effect as a sample and try to convey a message to world and all Indian people that if it will not be stopped then it will be completely a devastating scenario for all mankind.

2. Literature Survey

We have gone through various literature and publication from the official website of WHO and Ministry of human resource and welfare India and also studied the Journal published paper written by Simon James Fong, Gloria Li, Nilanjan Dey, Rubén Gonzalez Crespo and Enrique Herrera-Viedma on. In this publication they used multiple linear regression, Polynomial Neural Network with Corrective Feedback (PNN+cf) to find different parameter as output which may happen in near future as what we are seeing right now. WHO and ministry of health continuously updating every single news/danger & also awareness about this epidemic, and they regularly trying to communicate world people and let them know about the severity of this daemon corona. Indian Prime Minister Shri Narendra Modi announced medical emergency from 24/03/2020 which may continue for 21 days, this is because corona can spread very easily like touch, sneezing droplets. So Govt. tries to stop this chain by this complete lockdown. But Govt. has to take more effective decision than this People still go outside in a group without any mask or so. We don’t need to panic rather we have to follow guideline of WHO & Ministry of welfare & human resource India.

3. Proposed Methodology and Implementation

In this work we have analyzed and visualized CORONA dataset using tools for analysis and visualization. We also have performed prediction using linear regression technique based on machine learning.

3.1. Machine Learning Technology

We have applied some Machine learning techniques for data analysis and visualization from the data that we have. We have shown the mostly affected province and also the recovered and death cases from it.

3.2. Linear Regression

Linear regression is a very simple method but has proven to be very useful for large numbers of situations. Linear Regression is a machine learning algorithm based on supervised learning. Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y
The dataset that we have used here is collected at a very beginning of this epidemic. Since then we are working on this where we applied different machine learning algorithm and tried to depict the current scenario of India and China in different parameters like confirmed, recovered and death cases at different interval of time and for clear understanding. We have used the concept of visualization along with linear regression. We used here the concept of visualization for showing the current scenario of death and recovered cases along with time and also predicted the death in future using linear regression concept using Covid-19 dataset.

4. Tools used for Prediction and Visualization of Covid Data
We used the concept of folium for showing the global spread of COVID-19 in map. We have used Matplotlib and Plotly libraries for visualization purpose.

4.1. Matplotlib
Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+.

4.2. Plotly in Python
The plotly Python library is an interactive, open-source plotting library that supports over 40 unique chart types covering a wide range of statistical, financial, geographic, scientific, and 3-dimensional use-cases. Plotly's Python graphing library makes interactive, publication-quality graphs which need line plots, scatter plots, area charts, bar charts, error bars, box plots, histograms, heat maps, subplots, multiple-axes, polar charts, and bubble charts. In this paper we visualize the followings.

a) Covid-19 cases in India and China.

b) Total confirmed cases in India and China.

c) Prediction and forecasting of corona virus cases.

d) Graphical representation of the worldwide COVID-19 cases.

4.3. Prophet in Python
Prophet is a forecasting procedure implemented in R and Python. It is fast and provides completely automated forecasts that can be tuned by hand by data scientists and analysts.

4.3. Scikit-learn
Scikit-learn (also known as sklearn) is a machine learning library developed in Python programming language. It can be used to apply various classifications, regression and clustering algorithms including support vector machines, random forests, k-means etc to develop machine learning based system.

5. Dataset and Experimental result
By above all findings from different dataset that we have collected from WHO official website and Ministry of health and welfare India for my data analysis to show the world that the current situation is not good but if we the people of world can’t obey the Instruction and guidelines of WHO & messages of our countries health department then it could be the devastating for entire mankind community. Our dataset contains Covid-19 data of confirmed, recovered and death cases from 22/01/2020 to 19/09/2020. We have predicted as well as forecasted the Covid-19 cases in future. For this purpose we have added 20 new successive dates starting from 20/09/2020. These new dates for which we have forecasted are from 20/09/2020 to 09/10/2020.
The following is the comparative study of China and India in terms of confirmed, death and recovered cases. We have also calculated the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) of prediction for both the countries on confirmed, death and recovered cases of Covid-19 dataset.

5.1. Experiment on Confirmed Cases data for India and China
We have performed prediction and forecasting on confirmed cases data for India and China. The experimental results are generated using Profet and have been shown in Figure-1 and Figure-2.

![Figure 1: Prediction of Confirmed Cases in India](image)

Prediction Error:
RMSE for Prediction => 42884.56
MAE for Prediction => 20426.06
5.2. Experiment on Recovered Cases data for India and China
We have performed prediction and forecasting on confirmed cases data for India and China. The experimental results are generated using Profet and have been shown in Figure-3 and Figure-4.

Prediction Error:
RMSE for Prediction = 1636.28
MAE for Prediction = 808.63
Prediction Error:
RMSE for Prediction = 30443.36
MAE for Prediction = 14513.38

Figure 3: Recovered Cases in India

Figure 4: Recovered Cases in China

Prediction Error:
RMSE for Prediction = 326.28
MAE for Prediction = 220.48
5.3. Experiment on Death Cases data for India and China
We have performed prediction and forecasting on death cases data for India and China. The experimental results are generated using Profet and have been shown in Figure-5 and Figure-6.

![Figure 5: Death Cases in India](image)

Prediction Error:
RMSE for Prediction = 313.35
MAE for Prediction = 177.30
Prediction Error:
RMSE for Prediction = 195.36
MAE for Prediction = 119.04

5.4. Experiment using Polynomial Regression
We have applied polynomial regression with degree 10 on Covid-19 dataset for confirmed, recovered and death cases in India. To apply polynomial regression we consider each of these cases as a function of date variable. Here, date is considered as independent variable and each case variable (from confirmed, recovered and death cases) is considered as dependent variable. Each date value is represented by one integer number which is equal to the time difference of the current date from the starting date in terms of number of days. For example, the date “27/01/2020” is represented by value 5, as the starting date is “22/01/2020”. The Scikit-learn (sklearn) machine learning library has been used for this purpose. The experimental results are shown in Figure-7, Figure-8 and Figure-9.
Figure 7: Death Cases prediction and forecasting using Polynomial Regression

Prediction Error:
RMSE for Prediction => 309.51
MAE for Prediction => 202.19
Figure 8: Confirmed Cases prediction and forecasting using Polynomial Regression

Prediction Error:
RMSE for Prediction = 14238.66
MAE for Prediction = 9241.92
Figure 9: Recovered Cases prediction and forecasting using Polynomial Regression

Prediction Error:
RMSE for Prediction = 11225.84
MAE for Prediction = 8550.19

| Method          | RMSE (Recovered) | RMSE (Confirmed) | RMSE (Death) |
|-----------------|------------------|------------------|--------------|
| Polynomial Regression | 11225.84         | 14238.66         | 309.51       |
| Profet          | 30443.36         | 42884.56         | 313.35       |

Figure 10: RMSE values of prediction on Covid-19 data of India

We have compared the prediction results generated by the Polynomial Regression method and Profet. From Table-10 it is clear that Polynomial Regression method has less error in comparison to the results generated by Profet.

6. Conclusion and Future scope
It is seen by above all analysis and personal findings that covid-19 is really very dangerous it can be transmitted by any means and start influencing one people to other so social distancing, proper hand wash, using hand sanitizer is needed, From research & study We have observed one thing that suspect cases and confirmed cases is increasing exponentially and though especially in India the testing kit is not in ample amount that Doctors can check everyone. We have predicted and forecasted death recovered and confirmed cases on Covid-19 data for India and China. In future we want to tune the parameters used in Profet to check whether performance of prediction and forecasting increases or not. There is a kind of phobia in people that if they are suffering from cough and cold only, they would like to go for test, somehow they are responsible for increasing the crowd where investigation has been
started. So it’s very pathetic for those who are suffering from COVID-19 and they need testing. So we have designed a model where only 100% suspect will go for Covid-19 test on basis of defined group of symptoms. It’s not only about covid-19 when the people have lesser immune system any virus can attack human so man has to intake fruits and the food which have high source of different good proteins. As per the modern medical science no allopathic medicine can kill any viruses inside body. Further research and investigation regarding any diseases and novel viruses is a matter of future research also if we have proper information of patient then it will be possible to build a model where we can find easily the covid-19 patient.

References
[1] Fong SJ, Li G, Dey N, Crespo RG, Herrera-Viedma E.; “Composite Monte Carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction”. Appl Soft Comput. 2020 Aug;93:106282. doi: 10.1016/j.asoc.2020.106282. Epub 2020 Apr 9. PMID: 32362799; PMCID: PMC7195106.
[2] Zhonghua Liu Xing Bing Xue Za Zhi., “The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China” 2020 Feb 10;41(2):145-151. Chinese. doi: 10.3760/cma.j.issn.0254-6450.2020.02.003. PMID: 32064853.
[3] SARS-associated coronavirus by KV Holmes - New England Journal of Medicine, 2003 - Mass Medical Soc
[4] The molecular biology of coronaviruses by MMC Lai, D Cavanagh - Advances in virus research, 1997 – Elsevier.
[5] Identification of a new human coronavirus by L van der Hoek, K Pyrc, MF Jelbink, Nature medicine, 2004.
[6] The genome sequence of the SARS-associated coronavirus,MA Marra, SJM Jones, CR Astell, RA Holt
[7] Coronavirus avian infectious bronchitis virus, D Cavanagh - Veterinary research, 2007 – vetres.
[8] Abroug F, Slim A, Ouanes-Besbes L, et al. Family cluster of Middle East respiratory syndrome coronavirus infections, Tunisia, 2013. Emerg Infect Dis. 2014;20(9):1527-1530. doi:10.3201/eid2009.140378
[9] The molecular biology of coronaviruses PS Masters - Advances in virus research, 2006 – Elsevier
[10] Bats are natural reservoirs of SARS-like coronaviruses by W Li, Z Shi, M Yu, W Ren, C Smith, JH Epstein, 2005 - science.sciencemag.org
[11] Coronavirus genome structure and replication by DA Brian, RS Baric - Coronavirus replication and reverse genetics, 2005 – Springer
[12] Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus ,SR Weiss, S Navas-Martin - Microbiol. Mol. Biol. Rev., 2005 - Am Soc Microbiol
[13] Woo PC, Lau SK, Chu CM, et al. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J Virol. 2005;79(2):884-895. doi:10.1128/JVI.79.2.884-895.2005
[14] Clinical characteristics of coronavirus disease 2019 in China ,W Guan, Z Ni, Y Hu, W Liang, C Ou, J He, New England journal, April 2020 - Mass Medical Soc.
[15] The coronavirus replicaseJ Ziebuhr - Coronavirus replication and reverse genetics, 2005 - Springer