Brainwave: A Multidisciplinary Journal (ISSN: 2582-659X), Vol. 2, Special Issue No. 1, May 2021, pp. 43-49,

Approaches to Mitigate the Impacts of Highly Contagious Disease COVID-19 Using Machine Learning Algorithm

Rahul Banerjee[1,*],Dr.Sandip Roy[2],Subhadip Nandi[3]
[1]*Student of Brainware University*
[2]*Associate Professor,BrainwareUniversity*
[3]*Assistant Professor,Brainware University*
[*]*rahul.banerjee2008@gmail.com*

**Corresponding Author*

Abstract

Coronavirus pandemic is a worldwide life threading disease after H1N1 swine flu 2009-2010. Already 4,805,210 people got affected by this disease. According to source Wuhan (China) 's wet market is the source of this disease and now virus is spreading person to person and reached to all over the world. China reported an overall 2.3% mortality rate among COVID-19 patients. However, a significantly higher mortality rate (14.8%) was reported for senior patients (80 years or older). In Italy, where more than 23% of residents are 65 or older, the overall mortality rate has been about 5%, while the statistics showed a rate of around 20% for senior patients.My main target is to determine why this disease is spreading & how we can control it. To prevent this communicable disease we need proper plan to reduce its effect. Based on available data statistically want to create some pattern to make decision what are the actions can be taken to prevent. Shutdown, lockdown, travel ban is reducing spreading all over the world. Now need proper regional specific plan to flatten the curve. My purpose of analysis is identifying that approaches to mitigate this impact community wise. As demographic pattern & density is one of the key factors to reduce effect, so I have pointed out few factors of indicators in my conclusion section. My estimated result will help to take proactive measure to improve our public health system. Required medicine can be restocked to a different part of the country to maintain proper supply chain. Setting up new medical facility, arrangement of ventilator, PPE, N95 mask, gloves can be arranged beforehand to deal with Corona virus crisis.

Keywords: Logistic Regression, Predicted value, confusion matrix

1. Introduction

According to existing research, fever is the most common symptom, all most 99%. Based on my data I have depicted that for India close to 85 % patient is suffering from fever and then dry cough. So, to minimize community impact if any person is suffering from fever should not wait for any other symptom rather should contact to local medical facility for further testing as early as possible. There is a relationship exist with age group of the people [1-2]. Already established reports are showing mostly aged (more than 60) people are getting effected by this disease. But based on my acquired data for India age is not rightly proportional to affected people. My data shows mostly all group people are getting affected by this deadly disease and recovery ratio is almost distributed among the all age group. I don't see any direct relationship with the age. There is one know facts roaming around that there is a relationship exist among connectivity to outer world is contributing this spreading. In my analysis it is been pointed out that Maharashtra, Tamil Nadu, Delhi where most of the affected people have foreign travel history. Moreover, population density has direct relationship for spreading this disease widely [3-5]. As per centers of disease control & prevention corona virus attack can be reduced with proper preventive control. So rapid testing and identification of sick people,isolation of people with COVID-19, social isolation, such as closing schools and businesses and canceling large gatherings are the solution to prevent this disease for now. Throughout the world total affected is 5,481,142 and death count is 364,0701 [6].

2. Methodology

Logistic Regression – Since the value of the output variable in the case of logistic regression is either 0 or 1 and probability of output variable lies between 0 & 1 we have to find a function which must lie between 0 and 1 sigmoid function or logistic function is such a function and it is used to fit in logistic regression. Logistic regression is a statistical method for predicting the probability of the occurrence of a binary class using logit function. Special cases of linear regression where the target / output variable is categorical in nature. Uses log of odds as dependent variable. Y=m1x1+m2x2+m3x3+…. +C. Y is dependent variable & x1, x2, x3 ... are independent variables [7].

Sometimes it called logistic model or logit model, analyzes their relationship between multiple independent variables and a categorical dependent variable, and estimates the probability of occurrence of an event by fitting data to a logistic curve. There are two models of logistic regression, binary logistic regression and multinomial logistic regression. Binary logistic regression is typically used when the dependent variable is dichotomous, and the independent variables are either continuous or categorical. When the dependent variable is not dichotomous and is comprised of more than two categories, a multinomial logistic regression can be employed.

The logistic curve Logistic regression is a method for fitting a regression curve, y =f(x), when y consists of binary coded (0, 1- -failure, success) data. When the response is a binary (dichotomous) variable and x is numerical, logistic regression fits a logistic curve to the relationship between x and y. Logistic curve is an S-shaped or sigmoid curve, often used to model population growth. A logistic curve starts with slow, linear growth, followed by exponential growth, which then slows again to a stable rate. A simple logistic function is defined by the formula y= ex 1 + ex 1 1 + e-x

3. Data Analysis

I. Installed Anaconda framework with Jupiter notebook for Python to analyze the data. Panda data frame (read.csv) has been used to read the data from csv file and accordingly used python go.figure(),px.bar() functions has been used extensively [8].

II. Based on available data following are the symptoms for Corona virus. Here we can see fever, dry cough, muscle pain, short of breath etc. are common for all. X axis indicate percentage of the symptom & Y indicate symptom name [9].

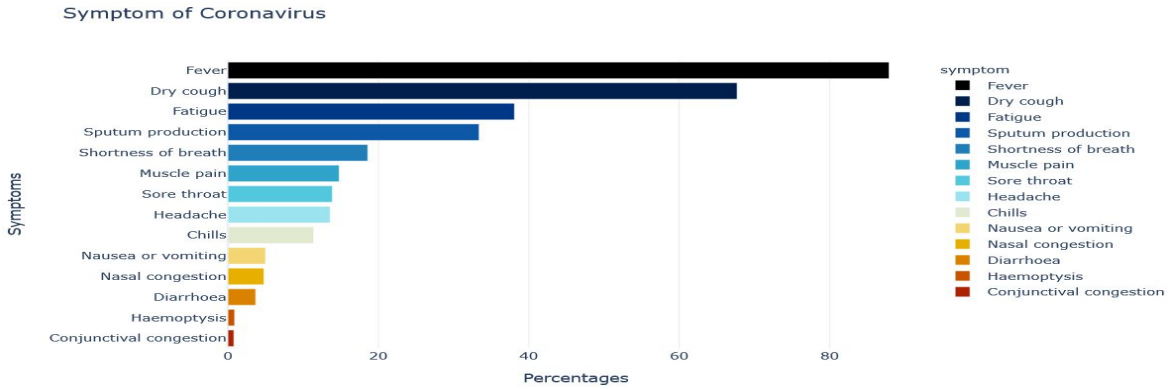

Figure 1:Symptomatic Analysis

III. To visualize state wise impact, create a graph plot using python px.bar function to show the data. Below are my output details. Here Y axis indicate number of active cases & Y axis indicate respective state name [10].

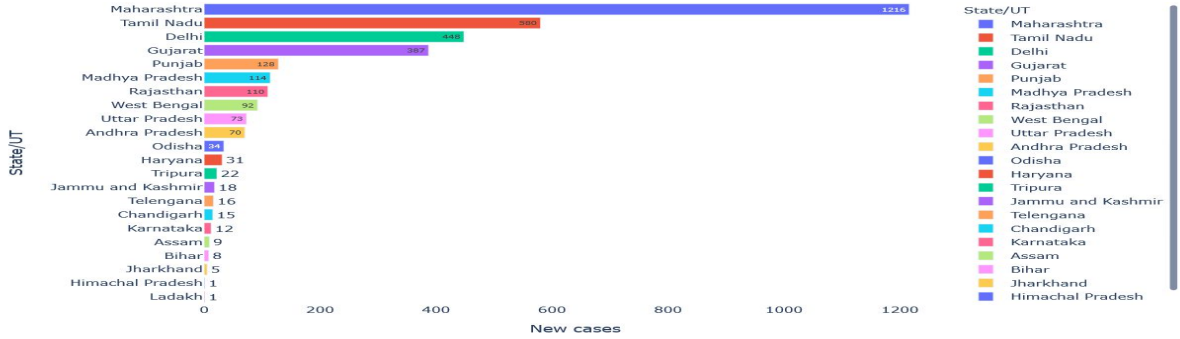

Figure 2:State wise spreading

IV. Gander wise affected details. Blue color indicates Female & Red indicate Male. Here X axis indicates age group of people & Y indicate death count reported. From this graph male are greatly affected than women [11].

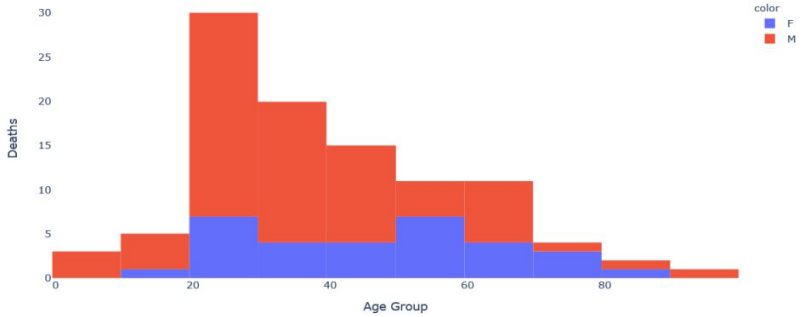

Figure 3: Gander wise affected distribution

V. The below graph is showing increased order confirmed cases over the time. X axis indicate number of count and Y indicate month [12].

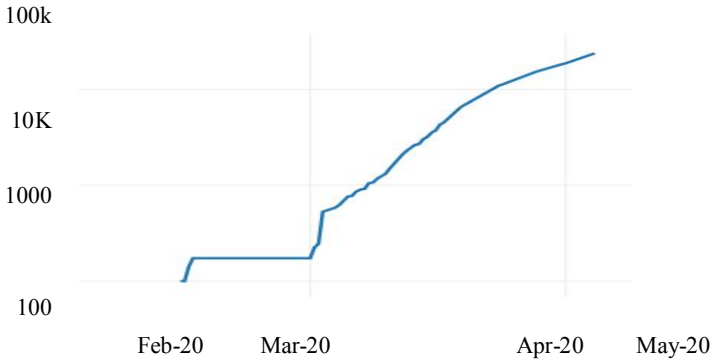

Figure 4: Confirmed case (Logarithm scale)

VI. There is age wise visualization created to estimate which group of people are vulnerable to this corona virus. X axis indicate age where Y indicate no of cases. And there is peak in between 29 to 45 who are mostly outside for their profession/work [13].

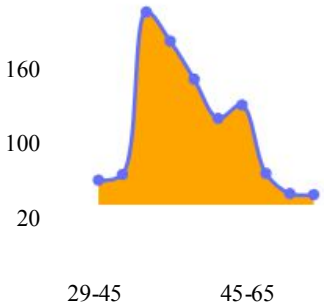

Figure 5: Age wise distribution

VII. Trend of Mortality rate has been determined from my raw data & display based on per 100 confirmed cases. As per my analysis death is 9.5 % for West Bengal which higher than average. And least in 2.27% for Jharkhand [14].

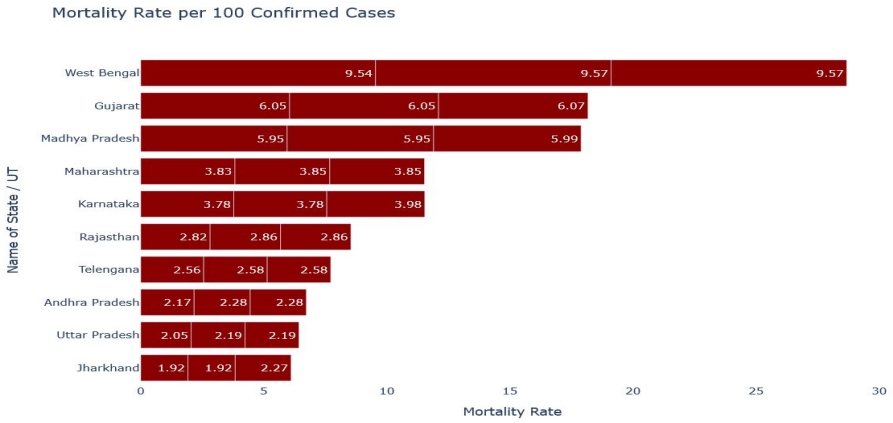

Figure 6: Mortality Chart

4. Result Analysis

To analysis this statistic angle I used most popular logistic regression method with Python as a programming language. Various Python library such as Sklearn, Numpy, Pandas, Json, BeautifulSoup,Matplotlib, Seaborn, Folium, Face Book Prophet, Math. With the help of libraries able to determine symptoms and mostly impacted states in India

Using Logistic regression, I have calculated predicted count which will be by May 2020 is 102668. And it is showing that high density states are mostly affected by this disease. Sikkim, Manipur are less affected compare to West Bengal, Maharashtra, and Tamil Nadu [15]. So, conclusion is if population is high then corona can spread rapidly. There is no relationship with age group any age people can be affected by this virus. The myth of immune system of women is high no longer a true statement. Mostly men are outside

without proper mask & necessary precaution which is spreading this disease. Only shelter in place can reduce spreading this disease.

A) Confirmed cases prediction details are given blow which I determined from my logistic model developed in python. The below trend is showing growth of the affected people starting from 0 days (assuming Feb -2020) and over the time it is keep increasing as per my prediction. X is number of cases & Y is day count.

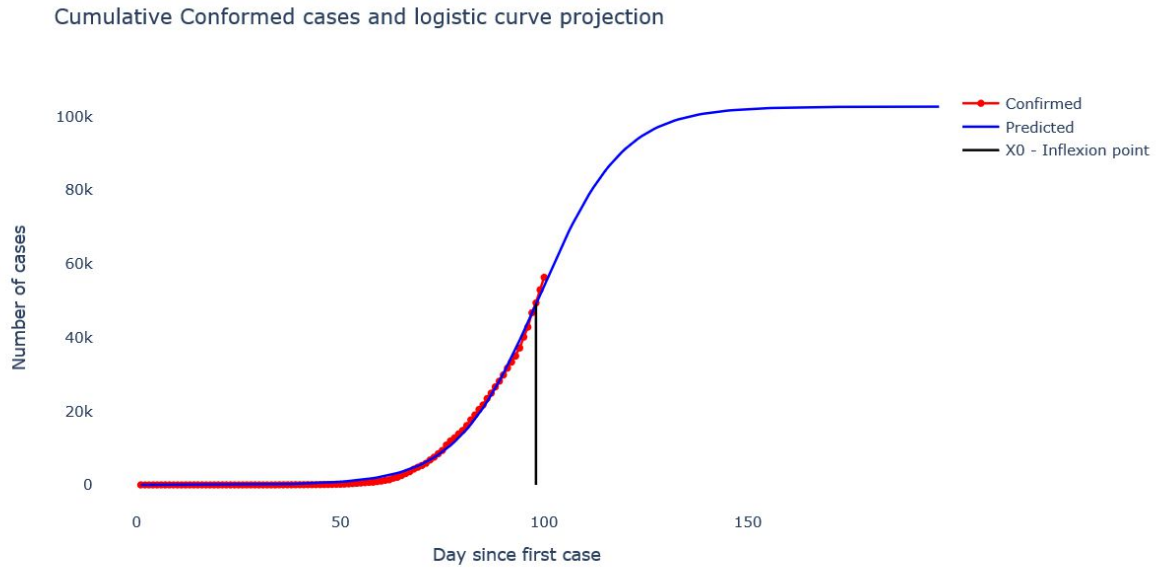

Figure 7: Actual Vs Predicated Analysis

Predicted L (the maximum number of confirmed cases): 102668
Predicted k (growth rate): 0.0984542358239968
Predicted x0 (the day of the inflexion): 98

Aconfusion matrix is a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known [16,17]. Few terms to remember in the context of confusion matrix, refer the table below:
true positives (TP): These are cases in which we predicted yes and are actually yes
true negatives (TN): We predicted no, and no in actual.
false positives (FP): We predicted yes, but actual is no. (Type I error)
false negatives (FN): We predicted no, yes in actual. (Type II error)
Measures of Accuracy
Sensitivity and specificity are statistical measures of the performance of a binary classification test:
Sensitivity/Recall = TP/(TP + FN). When it's actually yes
Specificity = TN/(TN + FP) .When it's actually no
Precision = TP/predicted yes. When it predicts yes

	Predicted No	Predicted Yes
Actual No	0.01	0.01
Actual Yes	0.02	0.98

B) Cumulative Confirmed cases and exponential curve projection. Y axis is representing number of cases whereas X is presenting number of days

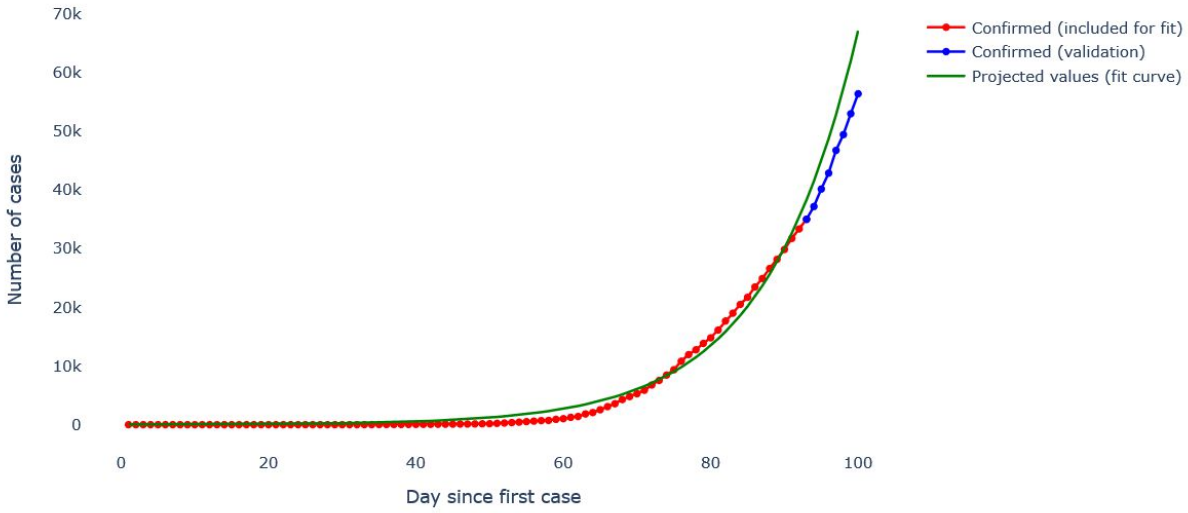

Figure 8: Confirmed cases growth rate

C) Confirmed virus predicted count using Face Book Prophet. Showing the comparison between predicted & confirmed count over the time. X axis is showing number of cases & Y axis is showing month from where we can estimate monthly progression.

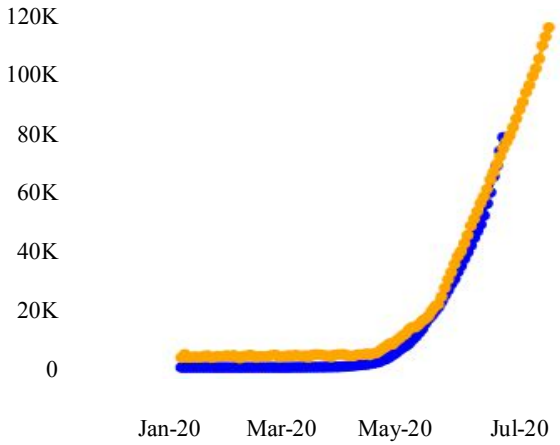

Figure 9: Confirmed cases increasing order over the month

5. Conclusion

Machine learning is helpful in the treatment of COVID-19 infected patients and for their proper health monitoring system, tracks the crisis using data analysis & ensures medical supply.Using the logistic regression model and factor analysis, we were able to determine the significant contributory factors that result in the increasing coronavirus. These factors were subsequently examined in order to determine what measures can be implemented to ensure that the disease can be identified at an early stage and to determine the best approach to undertake in order to reduce the effect of virus. This predictive algorithm will help our medical facility maximize the number of survivors. AI-MLC can help in developing proper treatment regimens, prevention strategies, drug and vaccine development.

References

1. Overview: Data Collection and Analysis Methods in Impact Corona virus Disease (COVID-19) - By Ahmad Yame, Lawrence Technological University
2. SAGE Research Journals
3. Corona Crisis and Inequality: Why Management Research Needs a Societal Turn By HariBapuji, Charmi Patel, GokhanErtug.
4. Social Distancing and Incarceration: Policy and Management Strategies to Reduce COVID-19 Transmission and Promote Health Equity through Decarceration by Brandy F. Henry, PhD, LICSW.
5. Dropout: a simple way to prevent neural networks from overfitting, by Hinton, G.E., Krizhevsky, A., Srivastava, N., Sutskever, I., &Salakhutdinov, R. (2014). Journal of Machine Learning Research, 15, 1929-1958.
6. Deep Residual Learning for Image Recognition, by He, K., Ren, S., Sun, J., & Zhang, X. (2016). CoRR, abs/1512.03385.
7. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, by Sergey Ioffe, Christian Szegedy (2015) ICML.
8. Large-Scale Video Classification with Convolutional Neural Networks, by Fei-Fei, L., Karpathy, A., Leung, T., Shetty, S., Sukthankar, R.,&Toderici, G. (2014). IEEE Conference on Computer Vision and Pattern Recognition
9. Learning deep features for scene recognition using places database, by Lapedriza, À.,Oliva, A., Torralba, A., Xiao, J., & Zhou, B. (2014). NIPS.
10. How transferable are features in deep neural networks, by Bengio, Y., Clune, J., Lipson, H., &Yosinski, J. (2014) CoRR, abs/1411.1792.
11. Do we need hundreds of classifiers to solve real world classification problems, by Amorim, D.G., Barro, S., Cernadas, E., & Delgado, M.F. (2014). Journal of Machine Learning Research.
12. (RF) versions implemented in R and accessed via caret) and the SVM with Gaussian kernel implemented in C using LibSVM.
13. Scalable Nearest Neighbor Algorithms for High Dimensional Data, by Lowe, D.G., & Muja, M. (2014). IEEE Trans. Pattern Anal. Mach. Intell.,
14. Trends in extreme learning machines: a review, by Huang, G., Huang, G., Song, S., & You, K. (2015). Neural Networks,
15. Multi-scale Orderless Pooling of Deep Convolutional Activation Features, by Gong, Y., Guo, R., Lazebnik, S., & Wang, L. (2014). ECCV.
16. Simultaneous Detection and Segmentation, by Arbeláez, P.A., Girshick, R.B., Hariharan, B., & Malik, J. (2014) ECCV .
17. One Millisecond Face Alignment with an Ensemble of Regression Trees, by Kazemi, Vahid, and Josephine Sullivan, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2014