



## To Prediction and Forecast of Corona Virus-19 using Machine Learning

V Anitha

MSC Zoology, Kakatiya University Warangal, Telangana State, India.  
anoo.vadlakonda@gmail.com

### Abstract:

The aim of this Covid-19 Outbreak Prediction paper is to make a model which will forecast the number of confirmed cases covid-19 virus in the upcoming days. Covid-19 is an infectious disease that is affecting a huge number of people all around the world.

This virus was first identified in Wuhan, China, and later spread throughout the world causing a pandemic that forced most countries to go into lockdown.

Various machine learning models and time series forecasting models.

The predictive model will be created using machine learning and using the dataset obtained from Kaggle. Machine learning automates the formation of analytical models. It is a branch of artificial intelligence focused on the principle that data can be learned from processes, It can find patterns and take decisions.

Time series forecasting will be used which is a type of predictive model. Time series forecasting is the use of a model centered on earlier observed values to evaluate future values.

**Keywords:** Covid-19, Classification, CT, Machine Learning, WHO

### I INTRODUCTION

COVID-19 is caused by a virus called SARS-CoV-2. It is part of the coronavirus family, which include common viruses that cause a variety of diseases from head or chest colds to more severe (but more rare) diseases like severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS).

Like many other respiratory viruses, coronaviruses spread quickly through droplets that you project out of your mouth or nose when you breathe, cough, sneeze, or speak.

The word corona means crown and refers to the appearance that coronaviruses get from the spike proteins sticking out of them. These spike proteins are important to the biology of this virus. The spike

protein is the part of the virus that attaches to a human cell to infect it, allowing it to replicate inside of the cell and spread to other cells. Some antibodies can protect you from SARS-CoV-2 by targeting these spike proteins. Because of the importance of this specific part of the virus, scientists who sequence the virus for research constantly monitor mutations causing changes to the spike protein through a process called genomic surveillance.

As genetic changes to the virus happen over time, the SARS-CoV-2 virus begins to form genetic lineages. Just as a family has a family tree, the SARS-CoV-2 virus can be similarly mapped out. Sometimes branches of that tree have different attributes that change how fast the virus spreads, or the severity of illness it



causes, or the effectiveness of treatments against it. Scientists call the viruses with these changes “variants”. They are still SARS-CoV-2, but may act differently.

## 2. LITERATURE REVIEW

There is an outbreak of Corona in early December. This is caused due to severe acute respiratory syndrome coronavirus 2, which is basically the family of SARS virus. Many governments all over the world are issuing their own preventive measures to control the spread of coronavirus. So, we have conducted a literature review regarding this virus, based on the information that is publicly available.

### Background of Literature Review:

China alerted WHO on 31st December 2019 that many people are reported to be suffering from Pneumonia, in Wuhan City. They reported that it started on Dec 8th, 2019, and there were an increasing number of patients who are working or living around the Huanan Seafood Wholesale Market.

When we started working on this paper at the start of February, the Coronavirus was majorly prevalent in China. Initially, at the time of our paper proposal, the mortality rate in China among all the confirmed cases is around 1.2% as of February 2020. And the mortality rate in all other countries, other than china was around only 0.2%. Among all the patients, who were admitted to the hospitals, the mortality rate, was around 11%. COVID-19 is increasing with great speed, and now there is a relatively very high mortality rate

### A Way to Further Research :

So, we have performed this literature review, to analyze the spread of coronavirus. After analyzing how increasingly it's spreading all over the world, we thought of performing our own prediction regarding this virus, so as to make people aware of its

spread, and with this, they can take their own preventive measures, so that they do not fall prey to this dangerous virus.

## III PRESENT SYSTEM

Various work on this problem related to covid-19 is being done. Officials all over the world are using several outbreak prediction models for covid-19 to make informed decisions and implement relevant control measures. Simple statistical models have received greater attention from authorities among the standard models for covid-19 global pandemic prediction. One of the works suggests using SEIR models. SEIR means susceptible-exposed-infected-recovered model.

This model aims to forecast factors like the spread of a disease, the total number of infected, and the span of an outbreak, and estimate different epidemiological parameters like the number of reproductive. Such models can illustrate how the outcome of the disease can be affected by various public health measures.

## PROPOSED SYSTEM

In this paper, we will first collect and evaluate the dataset. We will transform the raw data into an accessible format and visualize it using data preprocessing. Various machine learning algorithms such as Linear regression, polynomial regression, SVM, holt's linear model, Holt's winter model, AR model, ARIMA model, and SARIMA model are used. The tools used in this paper are mainly sklearn for model selection, and NumPy library which is used to work with the arrays and pandas that use a key data structure called a data frame that allows us to store and manipulate tabular data in observation rows and variable columns, matplotlib is a library of plotting that is used to plot graphs. After implementing the model, the model with the

least mean square error will be considered the best-fit model.

#### IV. METHODOLOGY

##### Approach

So, basically, we have followed the below approach to kick-start our Corona Virus Prediction paper:

1. Firstly, we have started with research on choosing the datasets. On performing research on various datasets, we have finalized with John Hoppkins data set, as it gives us the live data on coronavirus.
2. Secondly, we have collected the data and performed our preprocessing operation, so as to make our data ready for future predictions.
3. Next, coming to choosing the machine learning algorithm. We have chosen appropriate machine learning (we will discuss below regarding this).
4. Finally, we have performed our predictions to analyze the active cases, deaths, and recoveries for the next 30 days, based on the data available from the datasets and the chosen machine learning algorithm.

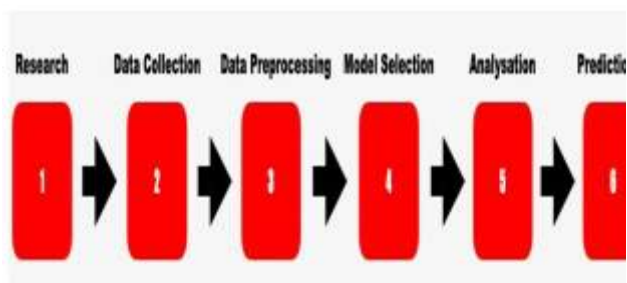


Figure: Approach of prediction of corona virus

#### V CONCLUSION

Finally, to conclude, we have performed prediction using SVR and Polynomial Regression Algorithm.

SVR predictions are mainly for predicting the world case scenario, which includes confirmed, death, and recovered cases.

Polynomial Regression is used for the prediction of US Cases.

Based on the results, we believe that our predictions were almost accurate, with some little differences from the actual values.

This paper can be further scalable, to include the predictions for various individual

#### References

1. Situation report-109. Coronavirus disease 2019 (COVID-19). WHO (2020). Available online at: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports> (accessed May 09, 2020).
2. Wee SL Jr, McNeil DG Jr, Hernández JC. W.H.O. Declares Global Emergencyas Wuhan Coronavirus Spreads. The New York Times (2020). Availableonline at: <https://www.nytimes.com/2020/01/30/health/coronavirus-world-health-organization.html> (accessed February 03, 2020).
3. COVID-19, ICMR. COVID-19. Indian Council of Medical Research.Government of India. ICMR (2020). Available online at: <https://main.icmr.nic.in/content/covid-19> (accessed May 09, 2020).
4. COVID-19 update. COVID-19 INDIA. Ministry of Health and FamilyWelfare. MOHFW (2020). Available online at: <https://www.mohfw.gov.in/>(accessed May 09, 2020).
5. Novel coronavirus-MOHFW. Home. Ministry of Health and Family Welfare.GOI (2020). Available online at: <http://www.mohfw.gov.in/> (accessed May 08,2020).
6. Bureau O. PM Modi calls for 'Janata curfew' on March 22 from 7 AM-9 PM.@businessline (2020). Available online at: <https://www.thehindubusinessline.com/news/pm-modi-calls-for-janta-curfew-on-march-22-from-7-am-9-pm/article31110155.ece> (accessed April 05, 2020).
7. Sangeeta N. Coronavirus Hotspots in India: Full List of 130 COVID-19 HotspotDistricts, All Metro Cities Marked Red Zones. Jagranjosh.com (2020). Availableonline at: <https://www.jagranjosh.com/current-affairs/coronavirus-hotspot-areas-in-india-what-are->



# International Journal For Advanced Research In Science & Technology

A peer reviewed international journal

[www.ijarst.in](http://www.ijarst.in)

ISSN: 2457-0362

**IJARST**

hotspots-know-all-covid-hotspots-1586411869-1 (accessed May 03, 2020).

8. Smith RD. Responding to global infectious disease outbreaks: lessons from SARS on the role of risk perception, communication and management. *Soc Sci Med.* (2006) 63:3113–23. doi: 10.1016/j.socscimed.2006.08.0049.

9. Mackay IM, Arden KE. MERS coronavirus: diagnostics, epidemiology and transmission. *Virology J.* (2015) 12:222. doi: 10.1186/s12985-015-0439-510