

# Predicting Spread of Covid19

Nidhish Kumar Reddy  
ICM2016006

Sarath Chandra  
IRM2016006

Nanda Kishore  
IWM2016002

May 4, 2020

## Abstract

This Paper Contains the Prediction of Covid19 Spread using ML Models

## 1 Problem Statement

We need to create a model that predicts how the virus could spread across different India and World. The goal is to Forecasting the world wide COVID-19 cases using Polynomial Regression and Regression using ANN

## 2 Data-Visualization

### 2.1 India Data-set

This Data-set contains Cured, Deaths, Confirmed in time series manner for each State/Union Territory in India. Here Feature is Date and We need to predict Active cases, Deaths, Confirmed Cases. Here Data Set is Updated up-to 28th April starting from 30th January

Sno	Date	Time	State/Union/Territory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Deaths	Confirmed
1	30/01/20	6:00 PM	Kerala	1	0	0	0	1
2	31/01/20	6:00 PM	Kerala	1	0	0	0	1
3	01/02/20	6:00 PM	Kerala	2	0	0	0	2
4	02/02/20	6:00 PM	Kerala	3	0	0	0	3
5	03/02/20	6:00 PM	Kerala	3	0	0	0	3

Figure 1: India-Dataset

### 2.2 World Data-set

Here we got 3 Data-Sets that is for Confirmed, Deaths, Recoveries in time series manner for each Country. Here Data-sets is updated up-to 30th April starting from 22nd January

Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	...	4/21/20
Western Sahara	24.215980	-12.893800	0	0	0	0	0	0	...	0
Sao Tome and Principe	0.386360	6.613081	0	0	0	0	0	0	...	4
Yemen	15.552727	48.516398	0	0	0	0	0	0	...	1
Comoros	-11.645980	43.333300	0	0	0	0	0	0	...	0
Tajikistan	38.861034	71.276993	0	0	0	0	0	0	...	0

Figure 2: World-DataSet

## 3 Regression

Regression analysis is a form of predictive modelling technique which investigates the relationship between a Dependent and Independent variable. Using the relationship between variables to find the best fit line or the regression equation that can be used to make predictions

### 3.1 Polynomial Linear Regression

Polynomial Regression is a form of linear regression in which the relationship between the independent variable  $x$  and dependent variable  $y$  is modeled as an  $n$ th degree polynomial. Polynomial regression fits a nonlinear relationship between the value of  $x$  and the corresponding conditional mean of  $y$ , denoted  $E(y|x)$ . These are basically used to define or describe non-linear phenomenon such as:

- Progression of disease epidemics
- Growth rate of tissues.

The general equation of a polynomial regression is:

$$Y = a_0 + a_1X + a_2X^2 + a_3X^3 + \dots + a_nX^n$$

where  $Y$  is Dependent Variable and  $X$  is Independent Variable

### 3.2 Regression Using ANN

Artificial Neural Networks (ANN) are comprised of simple elements, called neurons, each of which can make simple mathematical decisions. Together, the neurons can analyze complex problems, emulate almost any function including very complex ones, and provide accurate answers. A shallow neural network has three layers of neurons: an input layer, a hidden layer, and an output layer. A Deep Neural Network has more than one hidden layers, which increases the complexity of

the model and can significantly improve prediction power. Neural networks are reducible to regression models—a neural network can “pretend” to be any type of regression model

**Can Neural Networks be used for regression models?** Some might say it can be an overkill but the right answer would be it can be an interesting and valuable exercise to use neural networks for classic regression. Normal Regression methods work well only when if the regression equation is a good fit for the given data but a Neural Network is more flexible and it very well adapts to the shape and curves of our data. It can dynamically train the neurons to predict the correct values according to the given data unlike fixed learning in Linear Regression.

## 4 Polynomial Regression Outputs

### 4.1 India-Data-Set

#### 4.1.1 Test-Data-India vs Pred-Data-India

Splitting 70 percent for Train-set, 30 percent for Test-set. We can calculate the Active cases of each day by

$$Active = Confirmed - Recovered - Deaths$$

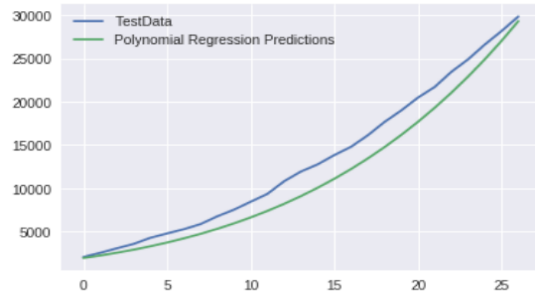


Figure 3: India-Confirmed Cases

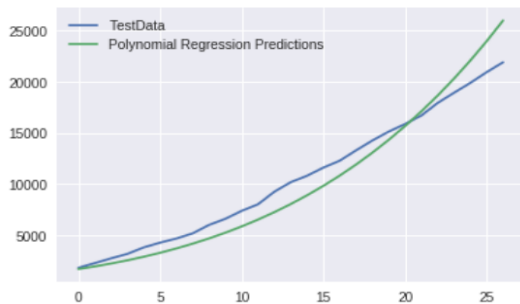


Figure 4: India-Active Cases

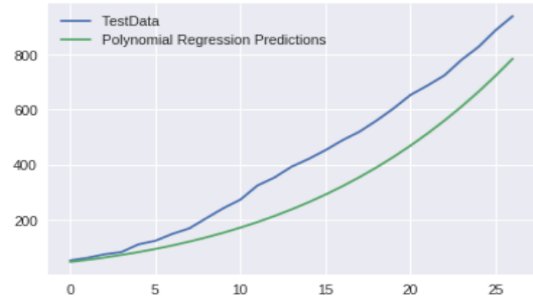


Figure 5: India-Deaths

#### 4.1.2 Next 10 days Predictions in India

Since in the Data-set last date it is updated is 28 April, Generating 10 Days ahead forecast of Confirmed cases, Active cases, Deaths of COVID-19 using Polynomial Regression

Date	ConfirmedCases
29/04/2020	[31723.060693544045]
30/04/2020	[34300.73001864547]
01/05/2020	[37042.283355292704]
02/05/2020	[39955.34348915581]
03/05/2020	[43047.763496358384]
04/05/2020	[46327.63013163711]
05/05/2020	[49803.26721650209]
06/05/2020	[53483.23902739654]
07/05/2020	[57376.35368385649]
08/05/2020	[61491.66653667088]
09/05/2020	[65838.48355604145]

Figure 6: Next 10 days Confirmed cases prediction

Date	Active
29/04/2020	[28076.83923501776]
30/04/2020	[30354.69344371873]
01/05/2020	[32777.1906031861]
02/05/2020	[35351.05196346124]
03/05/2020	[38083.20167439853]
04/05/2020	[40980.769769083956]
05/05/2020	[44051.095147253305]
06/05/2020	[47301.72855871025]
07/05/2020	[50740.435586744585]
08/05/2020	[54375.1996315507]
09/05/2020	[58214.224893645485]

Figure 7: Next 10 days Active Cases Prediction

Date	Deaths
29/04/2020	[848.5373265946528]
30/04/2020	[918.4043989791849]
01/05/2020	[992.7451041805366]
02/05/2020	[1071.7682021970422]
03/05/2020	[1155.6887653522394]
04/05/2020	[1244.7282711140024]
05/05/2020	[1339.1146949136628]
06/05/2020	[1439.0826029651466]
07/05/2020	[1544.8732450840953]
08/05/2020	[1656.734647506999]
09/05/2020	[1774.9217057103233]

Figure 8: Next 10 days Deaths Prediction

#### 4.1.3 Predicting the Peak of the Active Cases in India

Since in the given data set, the active cases of the country are Increasing there is no local minima at all. So i will just predict active cases for entire dates starting from 30th january to 9th march. Then i will just fit a Guassian curve which passes to the cases predicted by the model. So from Guassian Curve i will get a Peak date for Active cases. Assuming the constraints in Ideal Conditions. The peak date i got is

**Peak-Date:: 19-06-2020**

**End-Date:: 08-09-2020**

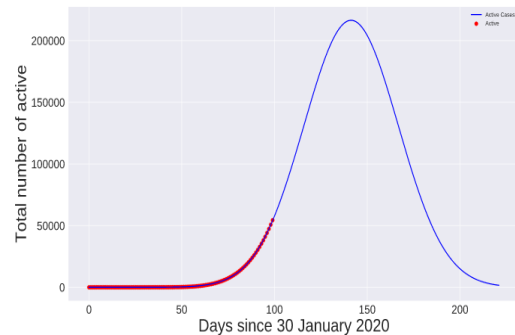


Figure 9: Peak Date Prediction using Active Cases

## 4.2 World-Data-Set

### 4.2.1 Test-Data-India vs Pred-Data-World

Split 70 percent for Train-set, 30 percent for Test-set. We can calculate the Active cases of each day by

$$Active = Confirmed - Recovered - Deaths$$

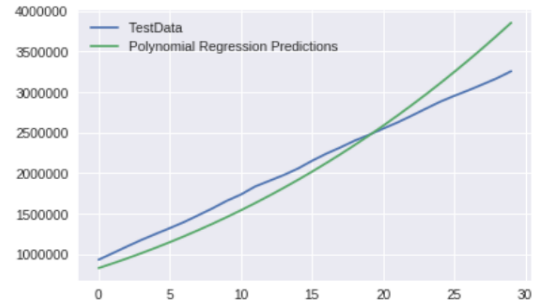


Figure 10: World-Confirmed Cases

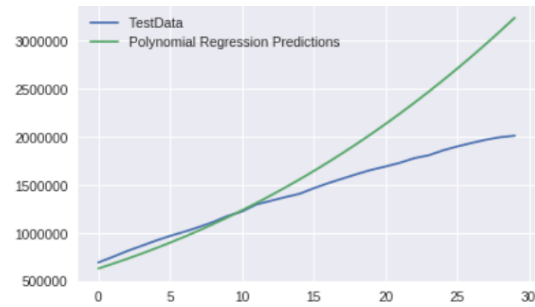


Figure 11: World-Active Cases



Figure 12: World-Deaths

### 4.2.2 Next 10 days Predictions in World

Since in the Data-set last date it is updated is 30 April, Generating 10 Days ahead forecast of Confirmed cases, Active cases, Deaths of COVID-19 using Polynomial Regression

Date	ConfirmedCases
01/05/2020	[4016502.4174010735]
02/05/2020	[4183584.650124598]
03/05/2020	[4355244.842583609]
04/05/2020	[4531544.132786437]
05/05/2020	[4712543.658741411]
06/05/2020	[4898304.5584568605]
07/05/2020	[5088887.969941117]
08/05/2020	[5284355.031202506]
09/05/2020	[5484766.880249361]
10/05/2020	[5690184.655090007]

Figure 13: Next 10 days Confirmed cases prediction

Date	Active
01/05/2020	3.378333e+06
02/05/2020	3.524473e+06
03/05/2020	3.674718e+06
04/05/2020	3.829125e+06
05/05/2020	3.987749e+06
06/05/2020	4.150646e+06
07/05/2020	4.317870e+06
08/05/2020	4.489480e+06
09/05/2020	4.665528e+06
10/05/2020	4.846073e+06

Figure 14: Next 10 days Active Cases Prediction

Date	Deaths
01/05/2020	[208074.39729441862]
02/05/2020	[216833.84519441484]
03/05/2020	[225833.161298609]
04/05/2020	[235075.53127176058]
05/05/2020	[244564.14077862882]
06/05/2020	[254302.1754839733]
07/05/2020	[264292.82105255313]
08/05/2020	[274539.26314912795]
09/05/2020	[285044.68743845704]
10/05/2020	[295812.2795852997]

Figure 15: Next 10 days Deaths Prediction

#### 4.2.3 Predicting the Peak of the Active Cases in World

Since in the given data set, the active cases of the world are Increasing there is no local minima at all. So i will just predict active cases for entire dates starting from 22th january to 10th march. Then i will just fit a Guassian curve which passes to the cases predicted by the model. So from Guassian Curve i will get

a Peak date for Active cases. Assuming the constraints in Ideal Conditions. The peak date i got is

**Peak-Date:: 12-06-2020**

**End-Date:: 03-01-2021**

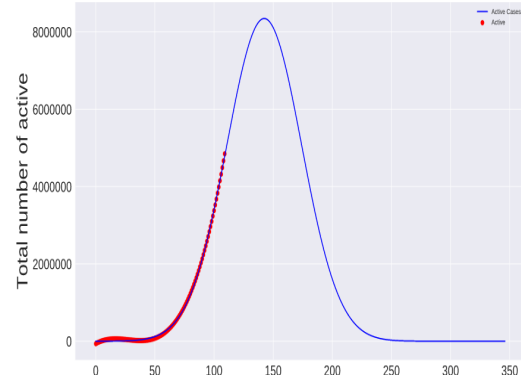


Figure 16: Peak Date Prediction using Active Cases

## 5 Neural Network method Outputs

### 5.1 Whole India:

#### 5.1.1 Prediction of Confirmed Cases

Train and Test split done same as above. Training Data is taken as confirmed cases between 30/1/20 to 1/4/20 and Plotting Between the 'No. of Days from 2/4/20' and predicted cases till '28/4/20'

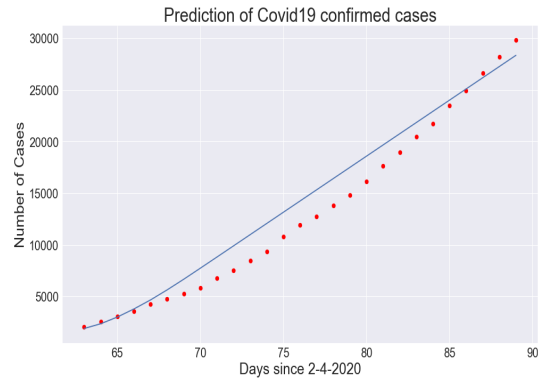


Figure 17: Prediction of Confirmed cases in India

#### 5.1.2 Future Forecasting for 10 Days

Since in the India Data-set last date it is updated is 28th April (89th Day), Generating 10 Days ahead forecast of Confirmed cases COVID-19 using Neural Networks

	Day	Prediction
0	90.0	29424.935547
1	91.0	30509.517578
2	92.0	31594.099609
3	93.0	32678.689453
4	94.0	33763.316406
5	95.0	34847.933594
6	96.0	35932.558594
7	97.0	37017.199219
8	98.0	38101.843750
9	99.0	39186.488281

Figure 18: Forecast of Confirmed cases for next 10 days in India

## 5.2 Focusing on Uttarakhand's trend of Active Cases:

### 5.2.1 Prediction of Active Cases

The model is trained for the trend in current data and tested for last 30% of total dates and the model seems to recognize the trend just as well. This is where Neural Networks perform better compared to regular regression methods as the data doesn't follow a linear trend.

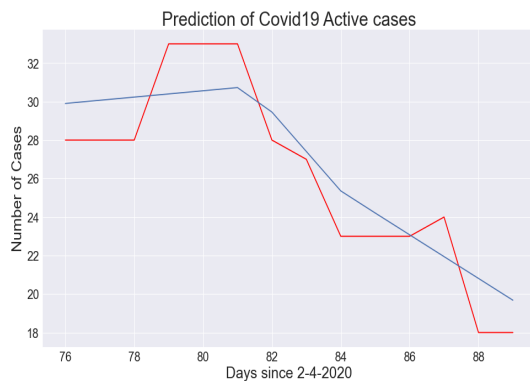


Figure 19: Forecast of Confirmed cases for next 10 days in India

### 5.2.2 Forecasting of Active cases until they become 0

The trend in data is taken into account and the end of Active Cases in Uttarakhand can possibly be predicted According to the model it shows that 20 days after 28th April(89th Day) i.e., 19th May(110th Day) the active cases could be expected to end.

	Day	Prediction
0	90.0	18.543720
1	91.0	17.409239
2	92.0	16.533207
3	93.0	15.660686
4	94.0	14.789351
5	95.0	13.912165
6	96.0	13.025116
7	97.0	12.154822
8	98.0	11.299393
9	99.0	10.444028
10	100.0	9.588305
11	101.0	8.732601
12	102.0	7.857908
13	103.0	6.937689
14	104.0	6.017538
15	105.0	5.097406
16	106.0	4.177374
17	107.0	3.257212
18	108.0	2.337011
19	109.0	1.417364
20	110.0	0.497660

Figure 20: Forecast of Confirmed cases for next 10 days in India

## References

- [1] Wikipedia contributors. "Polynomial regression." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 1 Dec. 2019. Web. 4 May. 2020.
- [2] Wikipedia contributors. "Polynomial regression." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 1 Dec. 2019. Web. 4 May. 2020. Wikipedia contributors. "Artificial neural network." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 24 Apr. 2020. Web. 4 May. 2020
- [3] Kaggle-website-for Indian Dataset  
<https://www.kaggle.com/sudalairajkumar/covid19-in-india>
- [4] Kaggle-website-for World Dataset  
<https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019>