IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Prediction Of Covid-19 Outbreak Using Machine Learning

Swati Powar[1], Nihar Chalke[2], Ketan Gogate[3],Anish Ugale[4]

[1]*Assistant Professor, Department of Information Technology, Finolex Academy of Management And Technology, Ratnagiri , Maharashtra, India*

[2,3,4] *U.G. Student, Department of Information Technology, Finolex Academy of Management And Technology, Ratnagiri ,Maharashtra, India*

Abstract: COVID-19 outbreak affect human lives as a whole and can be a cause of serious Health Problems and death. Artificial intelligence has been proven to be a effective and powerful tool in the fight against COVID-19 pandemic. Machine learning (ML) models are the most remarkable function in disease prediction, such as the Covid-19, in high geared forecasting and used to help decision-makers to understand future spread of COVID-19. The Aim of this paper is to predict the outbreak of COVID-19 using two approaches of Machine Learning viz. Data Visualization and Prediction using Linear Regression and Support Vector Machine. The models are designed to predict Covid-19, depending on the number of confirmed cases, recovered cases and death cases, based on the available datasets. Support Vector Machine (SVM) and Linear Regression (LR) models were used for this study to predict Covid-19 's risk. Predictions would help to get the future count of the COVID-19 cases and thus to establish precautions before the situation goes out of control. All three cases, such as confirmed, recovered and death, models predict deaths over the next 15 days. The experimental result showed that SVM is doing better than LR to predict the Covid-19 pandemic. According to this report, the pandemic has increased by half between the mid of July 2020. Then we will face a number of hospital shortages, and quarantine place.

Index Terms - machine learning, prediction, COVID-19 outbreak, data visualization, linear regression, support vector machine

I.INTRODUCTION

The COVID-19 pandemic, which the world is facing right now has been one of the most deadliest life destruction and has been described by WHO as the global health crisis of our time. Corona viruses are a family of viruses that cause various health diseases such as respiratory disease of gastrointestinal diseases. The COVID-19 disease is caused by the virus called SARS-CoV-2 virus. The present out-break of disease due to COVID-19 was reported in late 2019. The COVID-19 patient first emerged in Wuhan (China) in December 2019.The disease causes respiratory illnesses like influenza, flu, Ebola with common symptoms such as fever, dry cough, difficulty in breathing, and diarrhoea . The people suffering from corona virus start showing the symptoms in 2-14 days. India reported its first case on January 20, 2020 and now has become a global pandemic. This situation should be handled carefully so as to take proper precautions before count goes out of control. Many scientists are taking major efforts to save mankind from this calamity. In this era of technology, Artificial Intelligence and Machine Learning are playing important roles by the proper use of new and advance technology. Many of the data scientists worldwide are involved in preparing the right datasets and creating the strong models in order to fight against this pandemic. AI has been benefiting in various fields like giving early warnings and alerts, forecasting and prediction, data dashboards, social control, treatments and cures, etc. This paper focuses on building the prediction model to predict the future spread of COVID-19 for next 20 days. Linear Regression, polynomial Regression and Support Vector Machine are the two techniques of Machine Learning used for obtaining the results. Python language makes it simpler to obtain the desired Results. It acts as a obligatory tool to discover hidden insights and predict future trends.

1.1 Collecting data for predictions

Data collection is often a very challenging path while developing any machine learning model and it is essential to make hands dirty at this stage. The perfect dataset probably doesn't exist as the tremendous growth of data day by day or rather at every next second. In order to collect data for creating a primary dataset, following steps are to be followed. The dataset is provided and updated periodically by the John Hopkins University and is used for the study factors such as confirmed cases, number of deaths, number of recoveries, locations, active cases, dates and so on. The dataset extracted for this study is from January 22, 2020 for the prediction. The latest dataset is 27 April, 2020.

1.2 Data Visualization

Data Visualization is a field that deals with the graphical And pictorial representation of data in the form of graphs, charts and maps, etc. It is a very efficient way of handling and analysing Large amount of data or example a time-series and producing data driven Results. Data visualization uses statistical graphics, Plots, Information Graphics and some other tools. Data visualization is a technique uses to communicate data or Information by the use of visual objects. As the data and Information are represented using various different colors, it is easier for user to understand and visualize it.

The Libraries such as pandas And matplotlib in order to plot graphs with a python program.

- *Pandas Visualization*

Pandas Library is mainly used for data analysis. We can create basic plots using Pandas.Pandas is very useful when we have to create data analysis plots.It is an open source structure used for data analytics in Python.Pandas can also import files in various formats while .csv is the most popular one

- *Matplotlib*

Matplotlib is a low-level library is Python which is used for creating static, animated and interactive visualizations.It is a data visualization library built on NumPy arrays. It is used in various different applications like Tkinter, wxPython, etc. Pyplot API is a Matplotlib module which is used to plot Areas, Lines and Add labels to the plot.

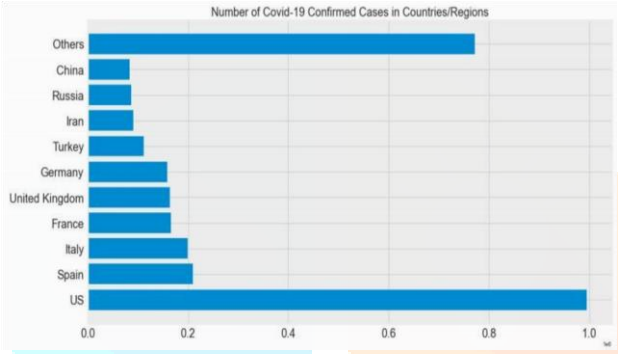

Fig-1: Top 10 countries affected by COVID-19

The above Example shows a bar graph representing the top 10 countries adversely affected by COVID-19. The figure given below (Fig-2) shows the growth of COVID19 Cases over the time across the whole world

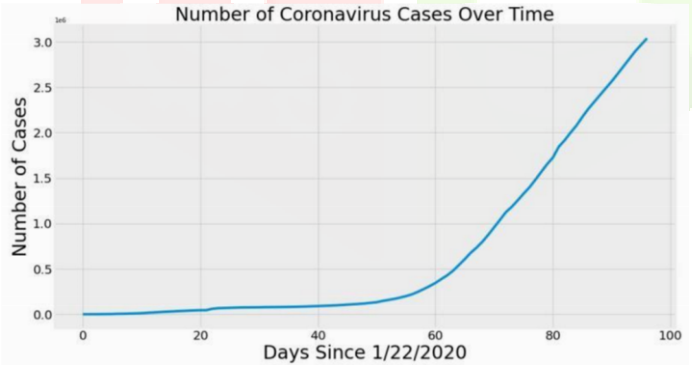

Fig -2: Number of COVID-19 cases over time in the world.

2.PREDICTION

"Prediction" generally refers an the outcome of an algorithm after it is trained with a historical time-series dataset and a model is applied to it. The machine learning predication model allows us to make highly accurate guesses to find outcomes using historical data.

2.1 Regression

Regression consist of mathematical methods that allows us to predict continuous output (y) based on one or more predictor variables (x). There are various different Regression techniques like Linear Regression, Multiple Regression, Polynomial Regression, SVR, etc. These models are based on supervised learning techniques.This paper includes prediction using polynomial regression and support vector regression techniques predicting the outcomes based on the inputs. Regression techniques.

2.1.1 Linear Regression

Linear Regression is a machine learning algorithm which is based on supervised learning. It Basically perform regression tasks. Regression Models target predication values on the basis of independent variables. The given below is a equation for Linear Regression

$y = \theta_0 + \theta_1.x_1 + \theta_2.x_2 + \ldots\ldots + \theta_n.x_n$ Where,

y = predicted value θ_0 = bias term $\theta_{1,\ldots,}\theta_n$ = model parameters $x_1,x_2,\ldots,x_n$ = feature values.

To start working with Linear Regression model, the data should be divided into training and testing sets using train_test_split function of sklearn., then Linear Regression function should be imported from sklearn Library and the data should be fitted in to the model using fit function. Predict function is used for predicting data in Linear Regression. The Mean Squared Error (MSE) and Mean Absolute Error (MAE) are required for calculating the predication error rates and performance of the model.

```
X_train_confirmed, X_test_confirmed, y_train_confirmed, y_test_confirmed =
train_test_split(days_since_1_22, world_cases, test_size=0.25, shuffle=False)

# Using Linear regression model to make predictions

linear_model = LinearRegression(normalize=True, fit_intercept=True)
linear_model.fit(X_train_confirmed, y_train_confirmed)
test_linear_pred = linear_model.predict(X_test_confirmed)
linear_pred = linear_model.predict(future_forecast)
print('MAE:', mean_absolute_error(test_linear_pred, y_test_confirmed))
print('MSE:',mean_squared_error(test_linear_pred, y_test_confirmed))
```

```
MAE: 1459906.8537211397
MSE: 2407543041442.5107
```

Fig-3: Implementation of Linear Regression

The Graph given below shows the comparison between the actual confirm cases and the Linear Regression predication. The graph also predicts the cases for the next 20 days.

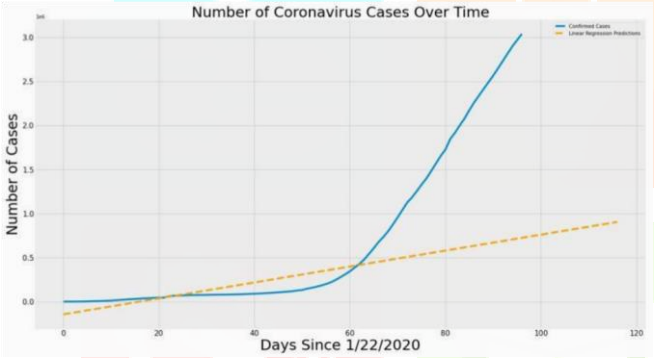

Fig-4: Prediction Graph for Linear Regression

To show the date wise prediction for the next 20 days in tabular form a data frame is created using pandas Library. The table given below shows the predicated no. of cases for the next 20 days.

	Date	Predicted number of cases using Linear Regression
0	04/28/2020	733600.0
1	04/29/2020	742643.0
2	04/30/2020	751685.0
3	05/01/2020	760728.0
4	05/02/2020	769771.0
5	05/03/2020	778813.0
6	05/04/2020	787856.0
7	05/05/2020	796899.0
8	05/06/2020	805941.0
9	05/07/2020	814984.0
10	05/08/2020	824027.0
11	05/09/2020	833069.0
12	05/10/2020	842112.0
13	05/11/2020	851155.0
14	05/12/2020	860197.0
15	05/13/2020	869240.0
16	05/14/2020	878283.0
17	05/15/2020	887325.0
18	05/16/2020	896368.0
19	05/17/2020	905411.0

Fig-5 : Predicted number of cases using Linear Regression

Polynomial Regression

Polynomial Regression is a form of Linear Regression that shows the relationship between a dependent variable (y) and independent variable (x) as nth degree polynomial. Polynomial Regression uses relationship between x and y variables to find the best way to draw a line through the data points. The given below is a equation for Polynomial Regression:

$y= b_0+b_1x_1+ b_2x_{12}+ b_2x_{13}+\ldots\ldots b_nx_{1n}$ where, y = dependent variable $x_{1\ldots}x_{1n}$= independent variables $b_{1\ldots}b_n$= coefficients

b_0 = constantAfter the data is divided into training and testing sets, the data is transformed for Polynomial Regression using Polynomial Features function and fit_transform method. Polynomial Regression model is prepared using LinearRegression function and data is fitted using fit function. The predictions are made using predict function of linear_model.

```
# Transform our data for polynomial regression

poly = PolynomialFeatures(degree=3)
poly_X_train_confirmed = poly.fit_transform(X_train_confirmed)
poly_X_test_confirmed = poly.fit_transform(X_test_confirmed)
poly_future_forecast = poly.fit_transform(future_forecast)
```

```
# Polynomial regression

linear_model = LinearRegression(normalize=True, fit_intercept=False)
linear_model.fit(poly_X_train_confirmed, y_train_confirmed)
test_linear_pred = linear_model.predict(poly_X_test_confirmed)
linear_pred = linear_model.predict(poly_future_forecast)
print('MAE:', mean_absolute_error(test_linear_pred, y_test_confirmed))
print('MSE:',mean_squared_error(test_linear_pred, y_test_confirmed))
```

```
MAE: 188902.0335329103
MSE: 73898204643.1854
```

Fig-6: Implementation of Polynomial Regression

The Graph given below shows the representation of the actual confirmed cases and the cases predicted by Polynomial Regression model.

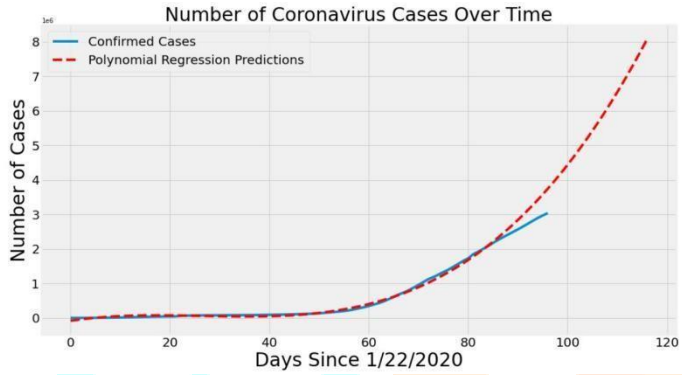

Fig-7: Prediction Graph for Polynomial Regression

The following table shows the date wise cases predicted by Polynomial Regression.

	Date	Predicted number of cases using Polynomial Regression
0	04/28/2020	3909045.0
1	04/29/2020	4080660.0
2	04/30/2020	4257215.0
3	05/01/2020	4438781.0
4	05/02/2020	4625425.0
5	05/03/2020	4817218.0
6	05/04/2020	5014227.0
7	05/05/2020	5216523.0
8	05/06/2020	5424174.0
9	05/07/2020	5637248.0
10	05/08/2020	5855816.0
11	05/09/2020	6079946.0
12	05/10/2020	6309708.0
13	05/11/2020	6545169.0
14	05/12/2020	6786400.0
15	05/13/2020	7033469.0
16	05/14/2020	7286446.0
17	05/15/2020	7545399.0
18	05/16/2020	7810397.0
19	05/17/2020	8081510.0

Fig-8: Predicted number of cases using Polynomial Regression

Support Vector Regression (SVR)

The Support Vector Regression algorithm is based on super wise learning and is used for dealing with classification problems in machine learning. The Objective of Support Vector Regression is to find a hyperplane in an n-dimensional space that classify the data points. The data points are on either sides of the hyperplane and the once closes to it are called Support Vectors.

SVM uses different parameters to build a model, which includes shrinking, kernel, gamma, epsilon, degree and c.

- shrinking : takes Boolean values
- kernel : specifies the kernel type use in an algorithm (Linear/Poly/Sigmoid)
- gamma : kernel coefficient of poly, sigmoid, etc.
- epsilon : specifies a margin of tolerance where no penalty is given to errors

• degree : of polynomial kernel function
• c : regularization parameter

```
# Support Vector Machine
# svm_confirmed = svm_search.best_estimator_

svm_confirmed = SVR(shrinking=True, kernel='poly',gamma=0.01, epsilon=1,degree=5, C=0.1)
svm_confirmed.fit(X_train_confirmed, y_train_confirmed)
svm_pred = svm_confirmed.predict(future_forecast)
```

```
svm_test_pred = svm_confirmed.predict(X_test_confirmed)
plt.plot(y_test_confirmed)
plt.plot(svm_test_pred)
plt.legend(['Confirmed Cases', 'SVM Predictions'])
print('MAE:', mean_absolute_error(svm_test_pred, y_test_confirmed))
print('MSE:',mean_squared_error(svm_test_pred, y_test_confirmed))
```

```
MAE: 286517.13370290556
MSE: 176487747625.62286
```

Fig-9: Implementation of SVR

The following graph shows the actual confirmed cases and the cases predicted by SVR model

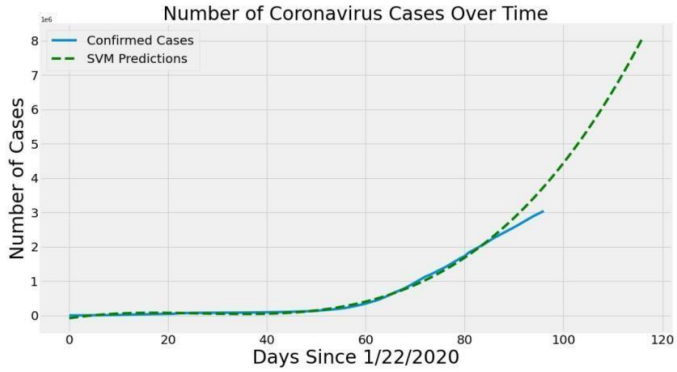

Fig-10: Prediction Graph for SVR model

The following table shows the date wise cases predicted by SVM.

	Date	Prediction number of cases using SVM
0	04/28/2020	4342572.0
1	04/29/2020	4569807.0
2	04/30/2020	4806508.0
3	05/01/2020	5052969.0
4	05/02/2020	5309488.0
5	05/03/2020	5576369.0
6	05/04/2020	5853924.0
7	05/05/2020	6142470.0
8	05/06/2020	6442330.0
9	05/07/2020	6753834.0
10	05/08/2020	7077316.0
11	05/09/2020	7413120.0
12	05/10/2020	7761594.0
13	05/11/2020	8123094.0
14	05/12/2020	8497981.0
15	05/13/2020	8886623.0
16	05/14/2020	9289396.0
17	05/15/2020	9706682.0
18	05/16/2020	10138869.0
19	05/17/2020	10586353.0

Fig-11: Predicted number of cases using SVR model.

*3.*CONCLUSIONS

COVID-19 is affecting human life, world trade, businesses and economy all over the world. As the rate of spread of the virus is high, it is difficult to control it. Under such crisis, Machine Learning and Artificial Intelligence techniques plays a important role to analyze and predict future spread of the virus. The prediction models are prepared using Linear Regression, Polynomial Regression and SVR help to predict the future spread of virus and get the situation under control. Machine Learning is used by Data Scientists all around the world to build a strong model to help the mankind to fight against this Pandemic.

REFERENCES

[1] World Health Oraganization. (2020). Coronavirus disease 2019 (COVID-19): situation report, 88

[2] 5 Tull, 5. Tull, R-Tall, and 5 S. Gill, Predicting the growth and trend of COVI-19 pandemic using machine learning and cloud computing fanernet of Thinge vel. 11, p. 100222, 2020

[3] Upendra Kumar Trivedi and Rizwan Khan "Role of machine learning to predict outbreak of COVID-19 in India," Research Gate, April. 2020.

[4] Mohammad Mehran, Austin George, Umesh Yadav, R. Logeshwari, "Epidemic outbreak prediction using AI", Vol7, Issue 4, April 2020.

[5] Sakshi Deshmukh, Tashfin Ansari, Dr. Almas M.N Siddiqui,Aniket Kotgire, Dr. Gaikwad A. T., "An overview of detection of COVID-19 in medical imaging using machine learning, Vol 7, issue 4, April 2020.

[6] SF Andiblerat. "COVID-19 Outbreak Prediction with Machine Learning 58 Electron, J. 2020

[7] Joseph George and Ranjeesh R Chandran, "Comparison of regression models on covid19 cases", Vol 7, Issue 5, May 2020

[8] Wim Naude,"Artificial Intelligence against COVID-19: An early review", IZA Institute of Labor economics, April 2020

[9] Sujath, R., Chatterjee, J. M., & Hassanien, A. E. (2020). A machine learning forecasting model for COVID-19 pandemic in India . Stochastic Environment Research and Risk Assessment, 1

[10] Rohan Taneja, Vaibhav, "Stock market prediction using regression", Vol 5, Issue 5, May 2018.

[11] Chakraborty, I., & Maity, P. (2020). COVID-19 outbreak: Migration, effects on society, global

[12] Environment and Prevention. Science of the Total Environment, 138882.

[13]https://www.datarobot.com/wiki/predictions- explanations/

[14]https://towardsdatascience.com/an-introduction-tosupport-vector-regression-svr-a3ebc1672c2

[15]https://www.edureka.co/blog/covid-19-outbreak-