IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2320-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 7, July 2022

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.118

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107126 |

Covid-19 and it's Impact Upon the Community

Sudhir Bussa [1] ,Riya Bajaj[2], Tarique Shahidi [3], Anirudh Amla [4]

Assistant Professor, Department of Electronics and Telecommunication, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India[1]

U.G. Student, Department of Electronics and Telecommunication, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India [2]

U.G. Student, Department of Electronics and Telecommunication, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India [3]

U.G. Student, Department of Electronics and Telecommunication, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India [4]

ABSTRACT:The pandemic of coronavirus disease 2019 (COVID-19), was first discovered in Wuhan, China, in December 2019. This has affected people round the world and also the number of infections and mortalities has been growing at an alarming rate. In times like these, a correct study of the disease spread can help in making efficient decisions and predictions. one among the most focus of this project is to use machine learning techniques to analyze and visualize the results of the spreading of the virus country-wise as well as globally during a selected period of time by considering confirmed cases, recovered cases and fatalities. In this project, we performed regression, Support vector machine and XGBoost on the Johns Hopkins University's COVID-19 data to anticipate long effects of COVID-19 pandemic in India and a few other countries. Moreover, we also studied the impact of some parameters like geographic conditions, economic statistics, population statistics, lifespan, etc., in prediction of COVID-19 spread.
.

KEYWORDS- Machine Learning, Data Analysis, Covid-19, Support Vector Machine.

I. TECHNOLOGY

Machine Learning

Machine learning is a field of study or process that teaches a computer to learn the data supplied without being explicitly programmed. this enables computers to take human-like decisions. Currently, this is active in various fields. Examples: Medical, industrial, astronomy, etc.

-LIBRARIES USED

NumPy

NumPy a Python library is used to manipulate arrays. It also provides functions for working within the areas of linear algebra, Fourier transforms, and matrices.

Pandas

Pandas a Python library is used to work with datasets. It's ability to clean, explore data makes it an efficient library. With Pandas, we can analyze big data and draw conclusions based on statistical theory.

Matplotlib

Matplotlib is Python's low-level graph drawing library that acts as a visualization utility.

Pyplot

Most Matplotlib utilities are under the pyplot submodule and are usually imported under the plt alias.

Altair

Altair is a statistical visualization library in Python. it's declarative in nature and uses Vega and Vega-Lite visualization grammars. it's fast becoming the primary choice of individuals searching for a fast and efficient means to visualize

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107126 |

datasets. It is rightly considered as a declarative visualization library since, while visualizing any dataset in Altair, the user only has to specify how the info columns are mapped to the encoding channel

II. METHODOLOGY

DATA MINING

Data cleansing is defined as the process of correcting or deleting incorrect, corrupted, malformed, duplicated, or incomplete data within a dataset. When combining multiple data sources, it's likely that the information is duplicated or mislabeled. If the information is wrong, the results and algorithms are unreliable, whether or not they appear correct. Since the method is different for every dataset, there's no absolute method indicate the precise steps of data mining cleansing process. However, it's important to form a template for the information cleansing process so you'll be able to latch on correctly whenever.

Steps Execution

The data provided by John Hopkins University also required cleaning for better results so we cleaned the data by following the steps mentioned

Removing of useless columns which are Latitude and Longitude (from Covid dataset)

Overall rank, Score, Generosity, Perception of corruption (from World Happiness Report). Followed by changing the indexing of the dataset from Provinces to Country. We also converted the date from string to datetime format. In the end we Replaced missing value NaN

EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis is defined as the initial analysis of provided or extracted data to identify trends, underlying constraints, qualities, patterns, and relationships between different entities within a dataset using descriptive statistics and visualization tools. EDA provides a good idea of which models are better suited to for the data and whether it has to be cleaned and massaged before it goes through advanced modeling techniques or is subjected to machine learning or AI algorithms

Exploratory Data Analysis Performed

We plotted the data for confirmed cases for China, Italy and India-
here n=25

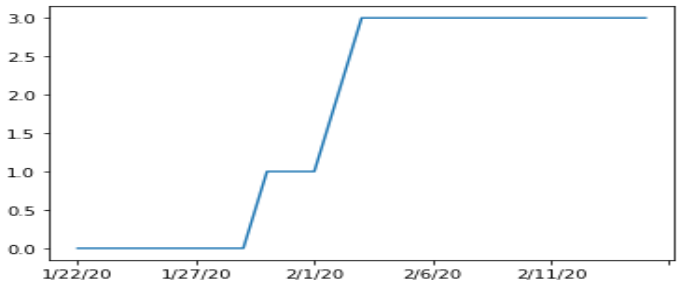

Figure 1- Confirmed Cases for China for 25 days

here no of days is 3(n=3): -

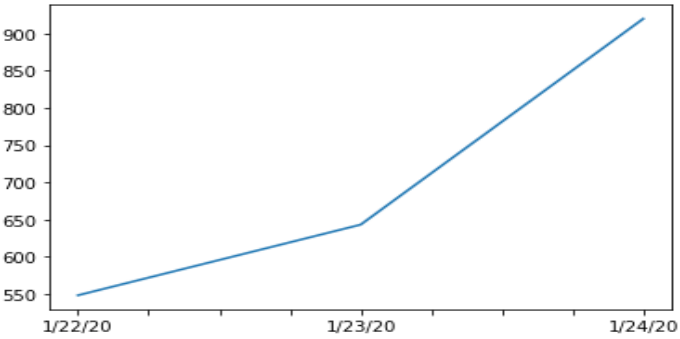

Figure 2- Confirmed Cases for Italy for 3 days

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107126 |

here no of days is 30(n=30):-

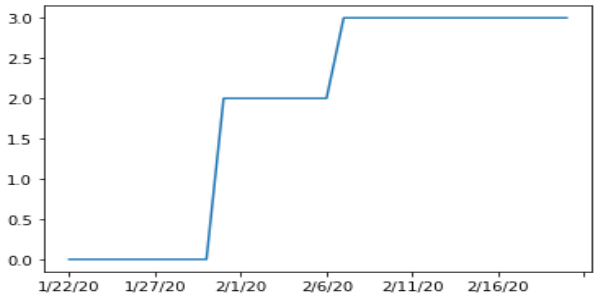

Figure 3- Confirmed Cases for India for 30 days

We also plotted the first derivative for China, Italy and India. This way we can count the number of cases increases each day. This can also be used to calculate the infection rate.

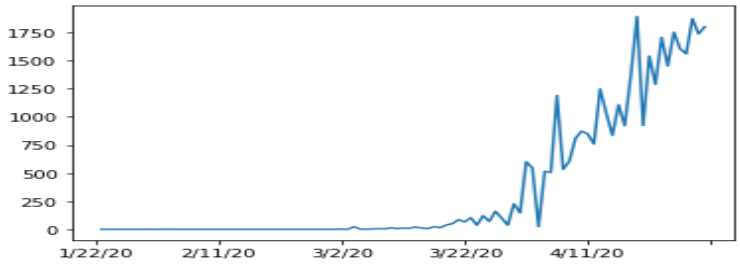

Figure 4- First Derivative for China

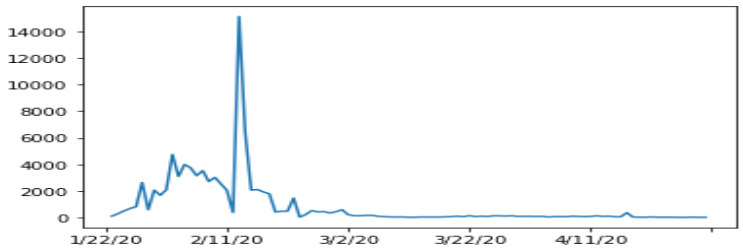

Figure 5- First Derivative for Italy

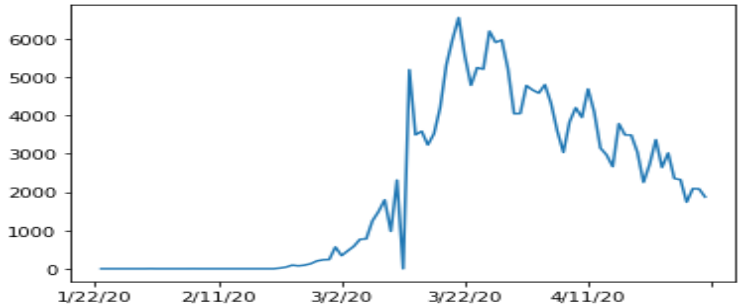

Figure 6- First Derivative for India

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107126 |

Followed by, we calculated a new column that denotes the maximum infection rate for all the countries. Thereafter we created a new data frame with only the maximum infection rate and countries as columns and then used Inner Join to merge the two datasets – the new dataset and the World Happiness Report.

To visualize the correlation between infection rate and factors such as GDP, Health Life Expectancy etc, we used scatter plots, regression lines and heatmaps in our further analysis.

GDP vs Maximum Infection Rate-

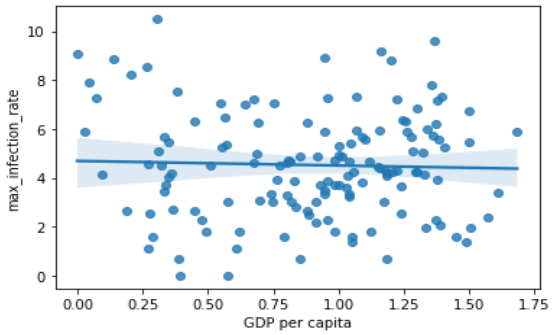

Figure 7 - GDP vs Maximum Infection Rate-

Social Support vs Maximum Infection Rate-

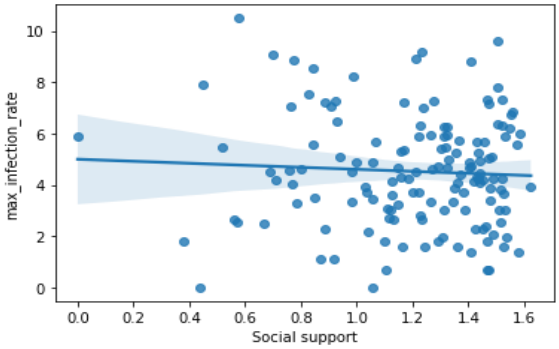

Figure 8 - Social Support vs Maximum Infection Rate-

Healthy Life Expectancy vs Maximum Infection Rate-

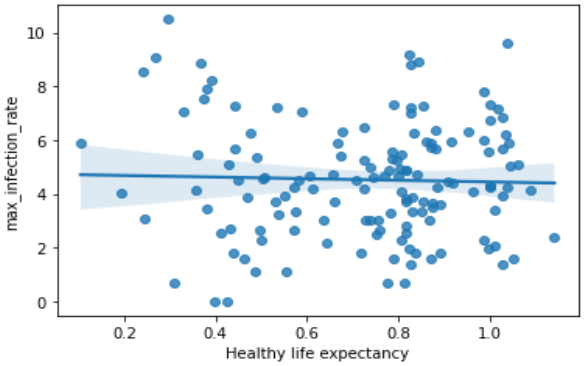

Figure 9- Healthy Life Expectancy vs Maximum Infection Rate

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107126 |

Freedom to make life choices vs Maximum Infection Rate-

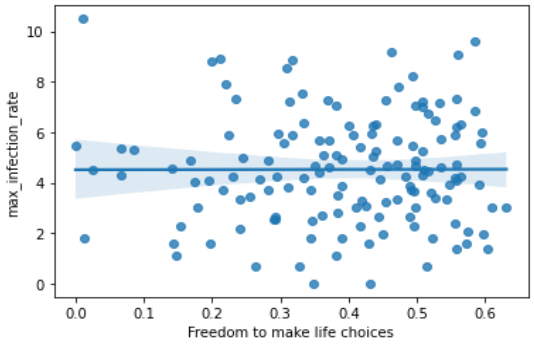

Figure 10- Freedom to make life choices vs Maximum Infection Rate

Plotting a correlation matrix and heatmap

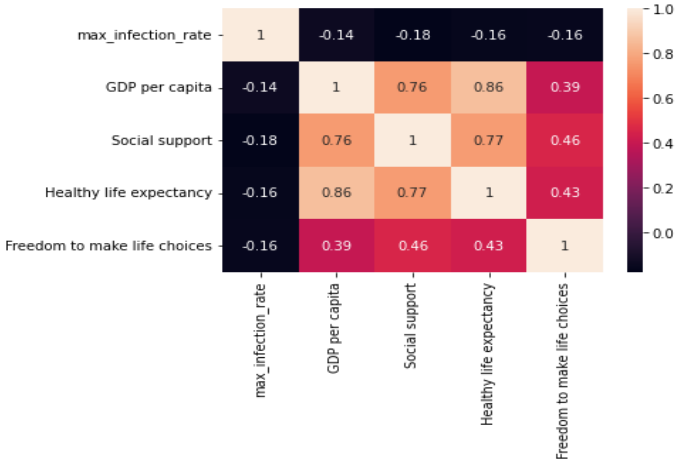

Figure 11- Correlation Matrix

From the above figure, we can interpret that the 'maximum infection rate' has a negative correlation rate with 'Freedom to make life choices', 'Healthy Life Expectancy', 'Social support' and 'GDP per capita' which suggests that as the maximum infection rate increases 'Freedom to make life choices', 'Healthy Life Expectancy', 'Social support' and 'GDP per capita' decreases which we all witnessed during this pandemic outbreak

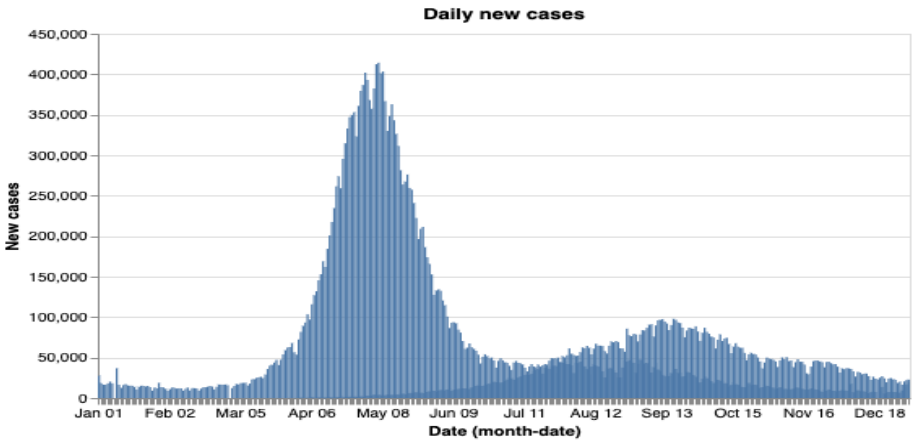

Figure 12-Daily new cases for India

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107126 |

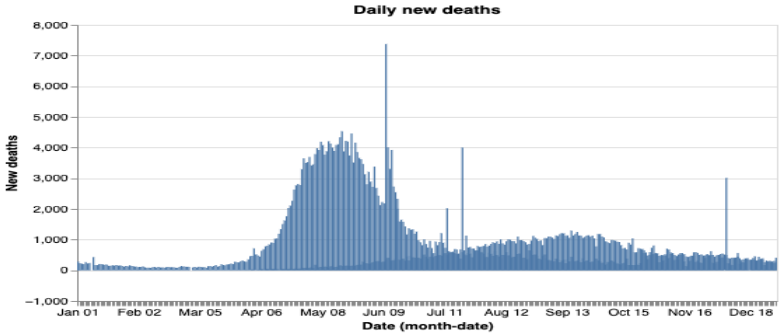

Figure 13-Daily new deaths for India

Machine Learning Algorithms

Machine learning algorithms are mathematical model mapping methods applied to learn or uncover underlying patterns embedded within the data. Machine learning comprises a collection of computational algorithms that can perform pattern recognition, classification, and prediction on data by learning from existing data (training set). Recently recent era we all have experienced the advantages of machine learning techniques from streaming movie services that recommend titles watch on the basis of supported viewing habits to observe fraudulent activity on the basis of spending pattern of the shoppers. It can handle large and sophisticated data to draw interesting patterns or trends in them like anomalies. Machine Learning Algorithms are classified into 4 types

- Supervised Learning
- Unsupervised Learning
- Semi-supervised Learning
- Reinforcement Learning

In our model, we have used Linear Regression, Support Vector Machine and XGBoost Algorithm to compare, see the results as well as anticipate the future effects of Covid19.
We divided the datasets into training and testing datasets-
Linear Regression-
In this process, a relationship is established between independent and dependent variables by fitting them to a line. This line is called as the because the regression line and represented by Y= a *X + b.
In this equation:

- Y – dependent variable
- a – Slope
- X – independent variable
- b – Intercept

The coefficients a & b are derived by minimizing the sum of the squared difference of distance between data points and the regression line.
We performed Linear Regression on the confirmed cases data by dividing the data into two parts-training data and testing data. First we trained the training data to a degree of four and then we tested the testing data on the same model.

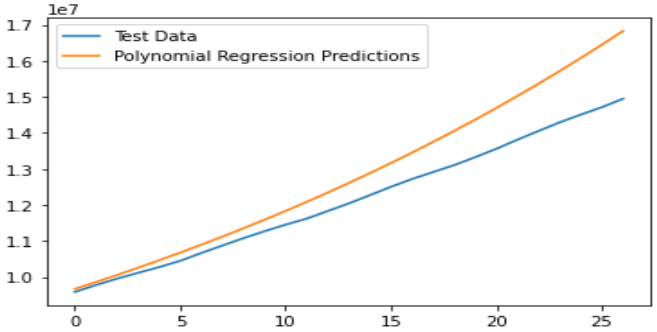

Figure 14- Linear Regression Results

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107126 |

Support Vector Machine

Support Vector Machine is employed for Classification problems in Machine Learning. The purpose of the algorithm is to develop the most effective line that segregates n-dimensional space into classes so that we are able to easily put the new information within the correct category henceforth which is termed as a hyperplane. The algorithm selects the extreme points/vectors that help in creating the hyperplane which are called as support vectors.

We applied Support Vector Machine Algorithm by dividing the dataset into training and testing datasets. First we trained the training dataset and then we passed the testing dataset through that model.

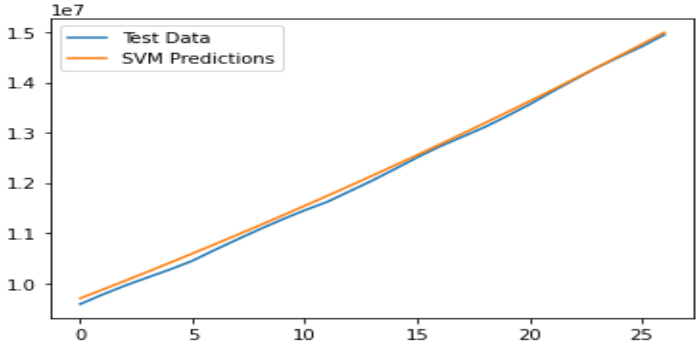

Figure 15- Support Vector Machine Results

[III]-CONCLUSION

In the end, our main aim for making this project was to understand how Covid19 affected our lives on a larger scale. Therefore, we created a correlation matrix and aheatmap visualization which resulted in a negative correlation of maximum infection rate with GDP, SOCIAL SUPPORT, HEALTHY LIFE EXPECTANCY and FREEDOM TO MAKE LIFE CHOICES.

We have applied advance machine learning algorithms- Support vector machine, Linear regression using which we have created a mathematical model for the confirmed cases of covid-19 in our dataset. By dividing the data set for the confirmed cases into 80% training data and 20% testing data. After training and testing our model with Support vector machine and linear regression we came to the conclusion that, Support vector machine proved to be a better fit for our model. In the future, we can say that support vector machine can be used for the purpose of prediction.

REFERENCES

[1] Divyadharani A K, Gayatri A B, Jeyanithy N M, Padmavathi R, Dr. K. Velmurugan,Predictive Analysis for Covid 19 Using Python, International Journal of All Research Education and Scientific Methods ISSN-2455-6211

[2] Batta Mahesh, Machine Learning Algorithms-a Review, International Journal of Science and Research ISSN-2319-7064

[3] Kamlendu Pandey, Ronak Panchal, A Study of Real World Data Visualization of Covid-19 dataset using Python, International Journal of Management and Humanities ISSN-2394-0913

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details