International Journal of Applied Engineering Research ISSN 0973-4562 Volume 16, Number 3 (2021) pp. 181-191


# Analyzing the Trend and Forecasting of Covid-19 Outbreak Using Machine Learning Techniques

**Dr. Suboh Alkhushayni** *suboh.alkhushayni@mnsu.edu*
**Dr. Du'a Alzaleq** *dua.al-zaleq@mnsu.edu*
**Tejaswi Vuyyur** *Tejaswi.vuyyur@mnsu.edu*
**Kayode O. Ayorind** *Kayode.ayorind@mnsu.edu*

**Abstract**

Covid-19 have impacted the whole world negatively since its outbreak and the health care sector has been challenged. There has been lot of pressure mounted on the sector to develop a solution such as a vaccine to treat patients or some suggested measures to slow down the spread following the World Health Organization's (WHO) directives. In this paper, we thoroughly analyzed the trend and predicted the cases of Covid-19 outbreak for the next 90 days across various major countries such as United States, France, Italy, Spain, China, India, and Germany. We used four different Polynomial Regression models, Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and Random Forest. We have taken a close look at the number of active cases, number of deaths, the mortality and recovery rates. The Polynomial Regression model performed better than other models, it came out to be the best fit model for predictions. Forecast was done with the best model. It suggests that positive cases in United States and India continued to increase weekly over the next 90 days. This shows that extra measures needed to be put in place to keep track of increased cases and reduce it.

**Keywords:** Covid-19 Outbreak; Polynomial regression; Support; Vector Machine (SVM); LSTM; Random Forest; World Health Organization (WHO); Prediction Algorithms.

## 1. INTRODUCTION

The year 2020 is a year to be remembered for life due to a deadly outbreak of a virus called SARS-CoV-2 (the virus that causes COVID-19). It was initially assumed that the virus originated from a market in a city called Wuhan in China. Still, according to [1], scientists from the Wuhan Institute of Virology (WIV)' and some other scientific experts around the globe have suggested that the virus may not necessarily originate from The WHO (World Health Organization) has been doing everything in its power to investigate, monitor, assess the risk, and report the status of infected patients since the first outbreak. The first case of the virus was reported outside China, in Thailand, on January 13 2020[2]. The cases started to rise at an unprecedented pace globally, so the WHO declared it a pandemic. Animals infected with this virus can infect dogs, which can spread to humans, especially if there is close contact between humans. An increase in the number of people affected by this virus has continued to change globally every day. On March 11, 2020, the World Health Organization declared Covid-19 a Pandemic, and a new social media campaign called "Be Ready for Covid-19" was launched, encouraging the public to be healthy, intelligent, and kind to each other. In terms of minimizing the spread, many industrialized countries have developed various approaches to combating the virus by encouraging the public to practice social distancing, using a public face mask, reducing person-to-person interactions, closing borders, etc.

The epidemic is an assault on the global health care system, the financial industry, the education sector, the economic sector, and small and large-scale enterprises. Take a closer look at countries that happen to have many infected cases, the number of deaths, and recovery rates, such as the United States, Italy, the United Kingdom, and China. For example, the United States was primarily affected by this epidemic, reported the first coronavirus case on January 21. When testing was extended, there was a rise in the number of cases day-by-day across each state. In the United States, there was a national Emergency Declaration. As of September 19, 2020, the total cases reported by the United States was about 6,955,007, death cases of 203,565 and recovered cases of 4,203,484 [3]. It can take between one week to two weeks for an infected person to be exposed to the symptoms of Covid-19, although that person may be able to infect other persons during this time. However, some infected individuals whose infection is so mild that the person can recover even before the hospitalization due to innate immunity. The primary measure taken to monitor Covid-19 by most governmental agencies is the introduction of lockdown to preserve social distance. To monitor the spread of the disease, this technique is an excellent measure. Even the full lockout could soon be the trigger of a major financial crisis from an economic point of view. Lockdowns in high-density countries may decrease the transmission rate of diseases, while full control may not be feasible [13]. Italy has recorded the highest number of deaths globally; the total number of deaths was obtained through the integration of the population registers and tax register covering 95% of the Italian resident population [4]. According to the lancet's article on "Italy's first wave of the Covid-19 pandemic has ended," the main key indicator of the real impact of Covid-19 is the mortality number reported because there are no mistakes in the certification of the cause of death At the end of January 2020, the United Kingdom experienced the first case of the coronavirus and as of September 19, 2020, they recorded 399,358 cases, 41,759 number of deaths and no record of recovery rates [3]. A stay-at-home order was implemented in March and they were

advised to self-isolate those with symptoms. According to the' South China Morning Post,' which stated that as early as November 17, the first case could have been a 55-year-old patient from Hubei province. It was mentioned that the first confirmed patient was never exposed to the Huanan Seafood Market, this is a market in Wuhan where we thought the virus originated from[5], it was found that after around 9 days the epidemic began with individuals who were on this market. China reported 85,269 cases, 4,634 deaths, and a recovery rate of 80,464 cases as of September 19, 2020[3]. To limit or slow down the spread of the virus, the Chinese authorities have done everything they can. There have been several research papers or papers dealing with the prediction of the virus that use machine learning methods in analyzing the number of deaths rates, hospitalizations, recovery rates, and the number of patients that have tested positive. Still, most papers have never really emphasized how mortality rates can help provide some predictive accuracy. A paper titled "A machine learning forecasting model for Covid-19 pandemic in India" [6], this article focused on presenting a model that could be useful to predict the spread of Covid-19. Various machine learning tools were implemented such as linear regression, multilayer perceptron, and Vector autoregression method. Suggestions were also made on to handle this crucial situation by social distancing. One of the limitations we observed from this paper is that the paper is limited to India alone. The authors use a linear regression that we believe cannot be an accurate measure to predict correctly. Also, the size of data available is huge, resulting in patterns, bias, or an outlier in the dataset. Another paper we reviewed is entitled "Forecasting the novel Covid-19 coronavirus"[7], which focused on the global impact of Covid-19, especially on patients recovering. At this point, the authors pointed out that no prediction is certain, as the future seldom repeats itself in the same way as the past. As predicting the future occurrence cannot be certain, we agreed with this suggestion and that is why we want to estimate what the number of occurrences in the next 90 days might be. The downside of this paper is that the emphasis is only on non-seasonal patterns. By presuming the pattern will continue indefinitely in the future, a realistic approach was suggested. According to a paper [8] titled "predicting the growth and trend of Covid-19 pandemic using machine learning and computing", the focus of this paper is how a machine learning and cloud computing can be deployed effectively to track the Covid-19 disease, predict growth and design strategies to manage the spread. Some of the objectives of this paper are proposing a novel scheme to predict the impact of the COVID-19 pandemic, highlight key future research directions and emerging trends and so on. The biggest limitation of this paper is the biases in the data due to diverse travel histories and contact demographic histories of the people from Wuhan.

We want to concentrate in this paper on analyzing the periodic pattern of Covid-19, that is, seasonal or non-seasonal patterns, using a recent dataset to estimate what the trend for the number of active events, death rates, recovery rates, and mortality rates will be in the next 90 days. We chose to predict using a machine learning approach such as SVM (Support Vector Machine), a Polynomial Regression, LSTM and Random Forest. Using different types of graphs and plots, several comparisons were made between major countries such as China, the United States, Spain, France, Germany, India, the United Kingdom, and their common states. The model's accuracy will be considered using the methods of mean square and mean absolute error.

## 2. DATASET

The data source is from Covid-19 Data Repository by the Center for systems science and Engineering (JHU CSSE) at Johns Hopkins University supported by ESRI living Atlas Team and Johns Hopkins University Applied Physics Lab (JHU APL) and some aggregated data sources contributed to putting together the dataset such as World Health Organization (WHO), European Centre for Disease prevention and Control (ECDC), DXY.cn.Pneumonia 2020, BNO news. The dataset includes the complete list of all sources ever used since January 21.

## 3. METHODOLOGY

The machine learning techniques adopted for predictions are SVM (Support Vector Machine) and Polynomial Regression, Long Short-Term Memory (LSTM) and Random Forest.

SVM is a predictive analysis data-classification algorithm that assigns new data elements to one of its labeled categories. Multiclass SVM is used as classifier on a data set that contains more than one class. Comparing to other classifiers, Support Vector Machines produce robust, accurate predictions that are least affected by noisy datasets and are less prone to overfitting. It is the most suitable algorithm for binary classification. According to paper on Acadgild website, called [9] Understand Power of Polynomials with Polynomial Regression explained why polynomial regression is one of the most important machine learning classifiers. Polynomial regression is a special case of linear regression. It is based on the idea of how features are selected. Polynomial can also be described as a regression algorithm that models the relationship between a dependent (y) and independent variable (x) as nth degree polynomial.

The equation is written as

$$y = b_0 + b_1 x_1 + b_2 x_1^2 + \ b_2 x_1^3 + \cdots + \ b_n x_1^n.$$

Predictions of new cases of Covid-19 is going to be based on these two models or more and comparison is going to be drawn. Long Short-Term Memory is a type of a neutral network which is capable of learning order dependence in sequential way of predicting problem. It has a feedback connection and its mostly used in the field of deep learning. According to Wikipedia, "LSTM can be used to process an entire sequence of data instead of a single process point. LSTM networks are well suited for classifying, processing, and making predictions based on time series data, as there can be lags of unknown duration between important events in a time series[10]. LSTMs have been developed to address the leakage gradient problem that can be encountered when forming traditional (Recurrent Neural Network) RNNs. [11] How does LSTM work? The basic concept of LSTM is cell state, which are different gates. The cell state acts as a highway that carries relative information all the way down the hierarchy. You can think of it as a network 'memory'. The cell's state, in theory, could carry relevant information throughout the sequence processing. Therefore, even information from previous time steps can make the path

to later time steps, reducing short-term memory effects. As the cell state continues its journey, the information is added or removed to the cell state via gates. Gates are various neural networks that determine the information allowed in the cell state. Portals can learn pertinent information to keep or forget during training. [12] Random Forest is one the supervised learning algorithm technique which can be classified or regression. It comprises of trees, the more the trees, the strong the forest will be. Feeding this algorithm with randomly selected data creates decision trees that give a prediction and choose the finest output from each decision tree. The algorithm has some advantages over other methods like, it is most highly accurate which does not hurt by overfitting. It can easily manipulate missing values by either replacing continuous variables with the median or proximity-weighted computation. Most disadvantages are, to create predictions are very slow because of the number of trees creating at a time and it is hard to explain.

## 4. RESEARCH QUESTIONS

Some of the research questions we aimed to answer in this paper are as follows:

- how the Public health infrastructure must be strengthened
- how the spread of the virus has increased over the months
- how the lockdown measures have decreased the spread of the in some countries
- On what condition can government lift the security measures when the spread of virus is low
- what future recommendations can the treatment centre have regarding the number of patients affected with the virus and regarding the number of patients likely to be admitted to intensive care.

## 5. RESULTS AND DISCUSSION:

The number of active cases, number of deaths, number of recoveries, and mortality rate of each nation were extensively examined, but our main focus was on China, the United States, Spain, France, Germany, India, and the United Kingdom and their key states. The data set was collected from the University of John Hopkins. In this dataset, the regular report on the cases of Covid-19 around the world has been preserved. Fig 1. displays a subset of our initial dataset. Details on the names of provinces, regions, their latitude/longitude location coordinates, and Covid-19 data for 282 days from January 22 2020 to October 31 2020 were included in the dataset. The data is cleaned and prepared to fit the model shown in Fig. 2.

Figure 3 shows the sample of the total number of covid-19 cases per country from January 22, 2020.

We can understand that the mortality rate of covid-19 is significantly less compared to rate at which it spread. States in China, the United States, Spain, France, Germany, India, and the United Kingdom have shown rapid increase for covid-19 cases. Fig.4 is the detailed analysis of each state in countries across the globe.

| | Province/State | Country/Region | Lat | Long | 1/22/20 | 1/23/20 | 1/24/20 | 1/25/20 | 1/26/20 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | NaN | Afghanistan | 33.93911 | 67.709953 | 0 | 0 | 0 | 0 | 0 |
| 1 | NaN | Albania | 41.15330 | 20.168300 | 0 | 0 | 0 | 0 | 0 |
| 2 | NaN | Algeria | 28.03390 | 1.659600 | 0 | 0 | 0 | 0 | 0 |
| 3 | NaN | Andorra | 42.50630 | 1.521800 | 0 | 0 | 0 | 0 | 0 |
| 4 | NaN | Angola | -11.20270 | 17.873900 | 0 | 0 | 0 | 0 | 0 |

5 rows × 288 columns

**Fig. 1.** Sample of original dataset

| | 1/22/20 | 1/23/20 | 1/24/20 | 1/25/20 | 1/26/20 | 1/27/20 | 1/28/20 | 1/29/20 | 1/30/20 | 1/31/20 | ... | 10/2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 4( |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 1{ |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 5{ |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 3 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | { |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 263 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 4{ |
| 264 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | |
| 265 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 2 |
| 266 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 1( |
| 267 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | { |

268 rows × 284 columns

**Fig. 2.** Data after required preparation

| | Country Name | Number of Confirmed Cases | Number of Deaths | Number of Recoveries | Number of Active Cases | Mortality Rate |
|---|---|---|---|---|---|---|
| 0 | US | 6771412 | 199294 | 2577446 | 3994672 | 0.029432 |
| 1 | India | 5400619 | 86752 | 4303043 | 1010824 | 0.016063 |
| 2 | Brazil | 4528240 | 136532 | 3936893 | 454815 | 0.030151 |
| 3 | Russia | 1092915 | 19270 | 903545 | 170100 | 0.017632 |
| 4 | Colombia | 758398 | 24039 | 627685 | 106674 | 0.031697 |
| 5 | Peru | 756412 | 31283 | 594513 | 130616 | 0.041357 |
| 6 | Mexico | 694121 | 73258 | 586154 | 34709 | 0.105541 |
| 7 | South Africa | 659656 | 15940 | 589434 | 54282 | 0.024164 |
| 8 | Spain | 640040 | 30495 | 150376 | 459169 | 0.047645 |
| 9 | Argentina | 622934 | 12799 | 478077 | 132058 | 0.020546 |
| 10 | France | 467614 | 31257 | 93586 | 342771 | 0.066844 |

**Fig.3.** Analyzing the number of Covid-19 Cases per Country

| | Province/State Name | Country | Number of Confirmed Cases | Number of Deaths | Number of Recoveries | Mortality Rate |
|---|---|---|---|---|---|---|
| 0 | Maharashtra | India | 1188015 | 32216 | 857933 | 0.027118 |
| 1 | Sao Paulo | Brazil | 931673 | 33927 | 780448 | 0.036415 |
| 2 | California | US | 783313 | 15018 | 0 | 0.019172 |
| 3 | Texas | US | 707940 | 15051 | 0 | 0.021260 |
| 4 | Florida | US | 681233 | 13287 | 0 | 0.019504 |
| 5 | Andhra Pradesh | India | 617776 | 5302 | 530711 | 0.008582 |
| 6 | Tamil Nadu | India | 536477 | 8751 | 481273 | 0.016312 |
| 7 | Karnataka | India | 511346 | 7922 | 404841 | 0.015492 |
| 8 | New York | US | 449038 | 33081 | 0 | 0.073671 |
| 9 | Lima | Peru | 349167 | 14009 | 0 | 0.040121 |
| 10 | Uttar Pradesh | India | 348517 | 4953 | 276690 | 0.014212 |

**Fig. 4.** Total number of Covid-19 cases per state.

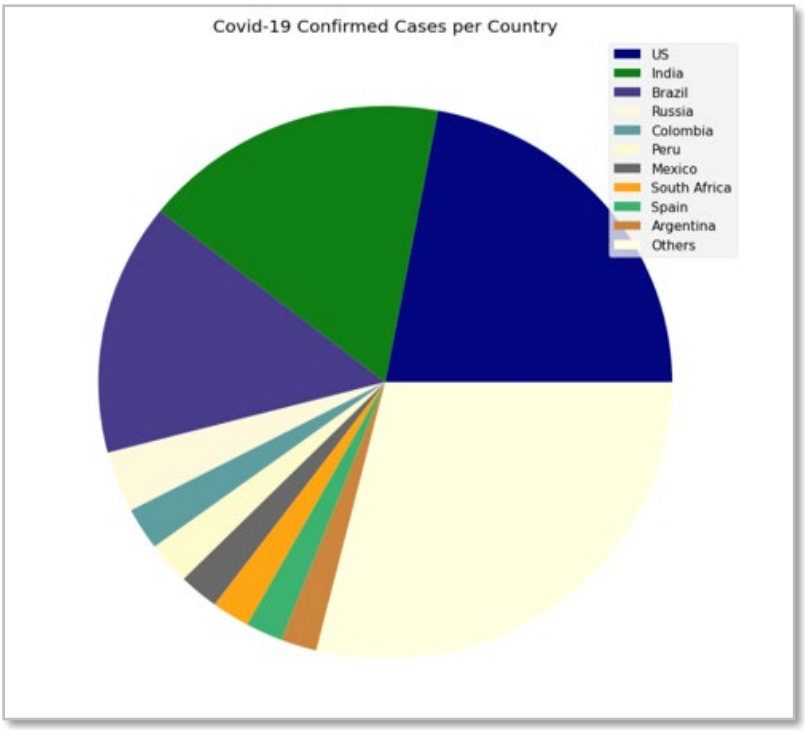


**Fig. 5.** Pie chart visualization of the number of Covid-19 cases per country.

The pie chart in Fig.5, Fig.6a, and 6b give us a simple visualization of the number of cases of Covid-19 in the United States, India, and other major countries around the world per country and per state. A pie chart makes it simple to evaluate and compare different countries with a higher number and fewer cases of Covid-19.

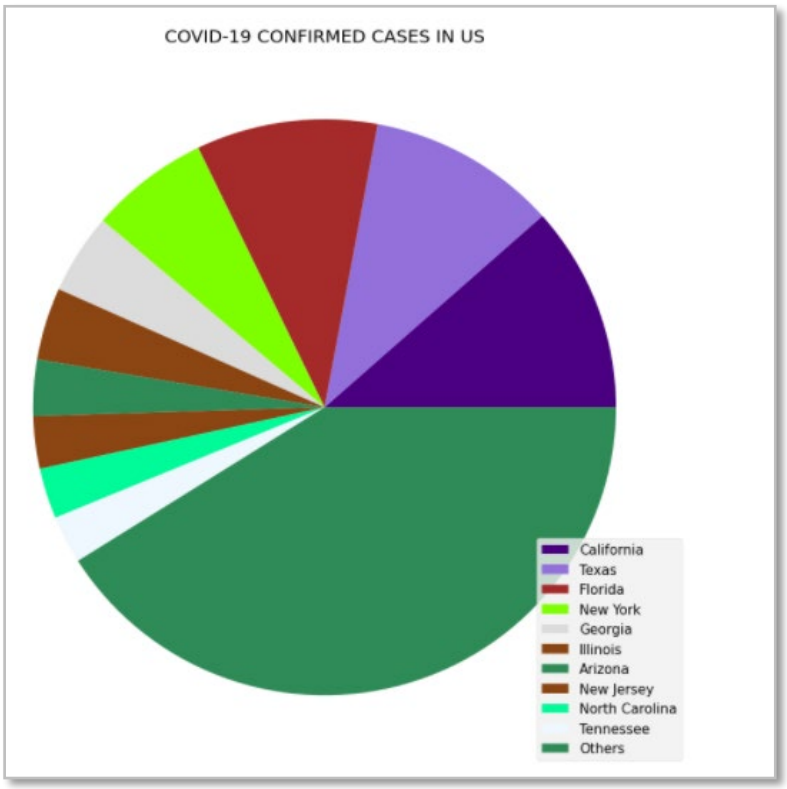


**Fig.6a.** Pie chart visualization of the number of Covid-19 cases per state in United State.

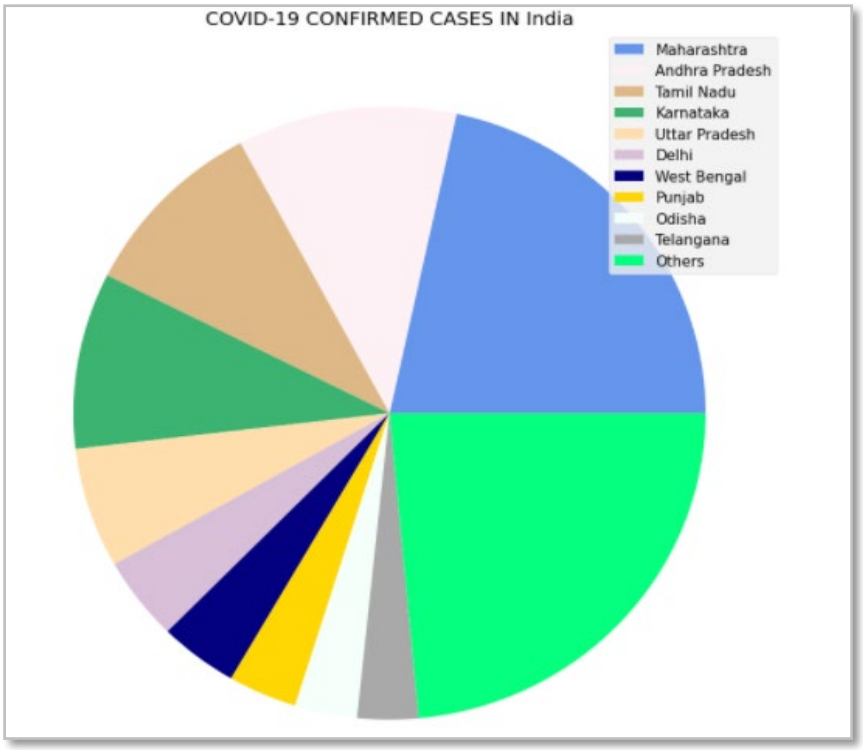


**Fig.6b.** Pie chart visualization of the number of Covid-19 cases per state India.

Our dataset focuses on cases of Covid-19 that were gathered between January 22, 2020, and October 31, 2020. 75% of the data was used for the model preparation and 25% for the testing. We have a total of 282 days and modeling is done with 211 days and the remaining 71 days for testing. We predicted the number of instances using SVM, Random Forest, LSTM and Polynomial Regression.

Fig.7, Fig.8, Fig.9, and Fig. 11 below clearly demonstrate that in estimating the number of Covid-19 cases, the Polynomial Regression did a better job than the SVM, LSTM, and Random Forest. While not all models have done a perfect job, they are quite inaccurate in estimating the number of cases over the next 90 days.

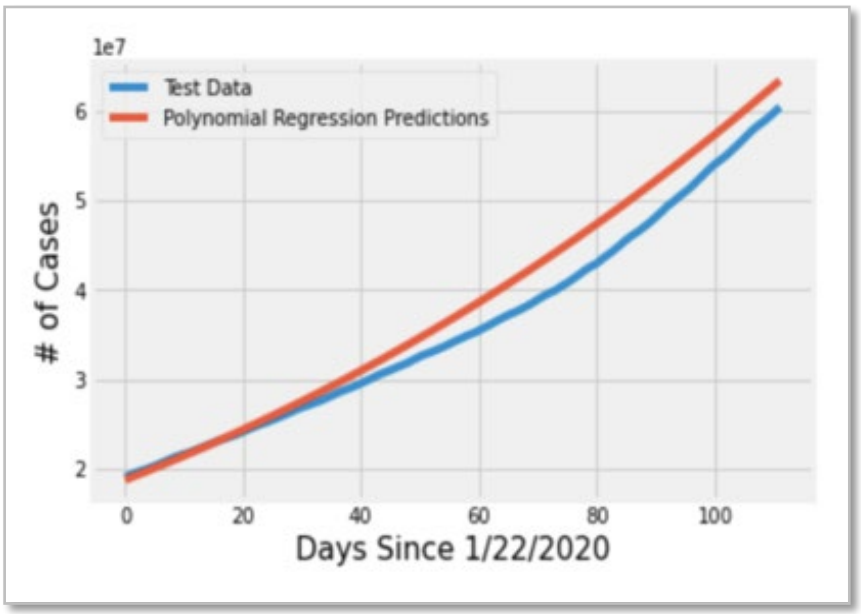


**Fig. 7.** Polynomial Regression Predictions

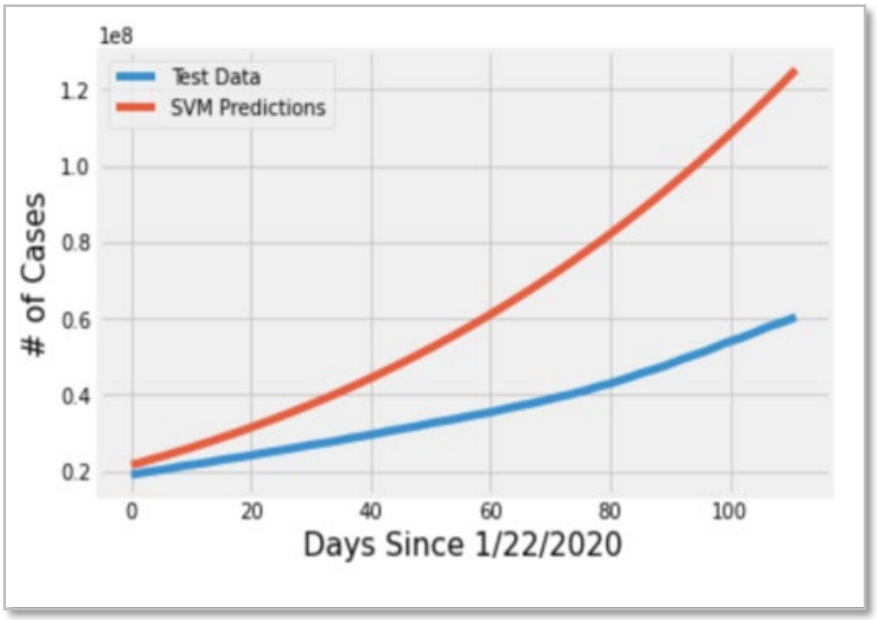


**Fig. 8.** SVM Predictions

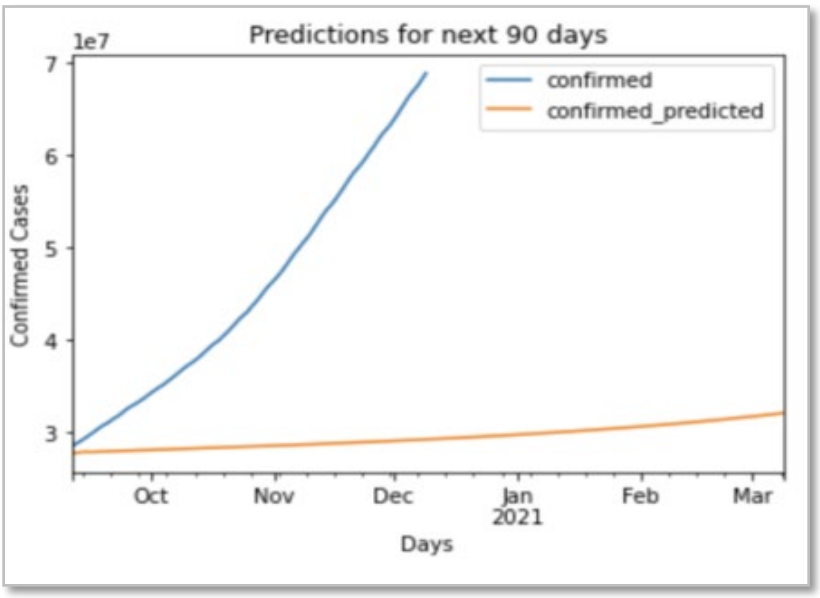


**Fig. 9.** LSTM Predictions

Fig.12. represents the number of cases plotted on the y-axis against the number of days since January 2020 plotted on the x-axis to evaluate the rise in cases of Covid-19 around the globe.

As we have seen from fig. 7 and fig. 8 Polynomial Regression and SVM both showed a positive trend. On the other hand, LSTM and Random Forest are not suitable for forecasting in this case, as shown in Fig. 9 and fig. 10, this is because the data set is not time-series in nature. The Polynomial Regression model is more accurate, with MSE 1441095873039 and MAE 3230574.08 being the lowest. SVM has MSE 332449158283797.8 and MAE 14901449.183, respectively. The model of Polynomial Regression is steadier and more valid. We will be forecasting based on the best model, which is shown in Fig. 12.

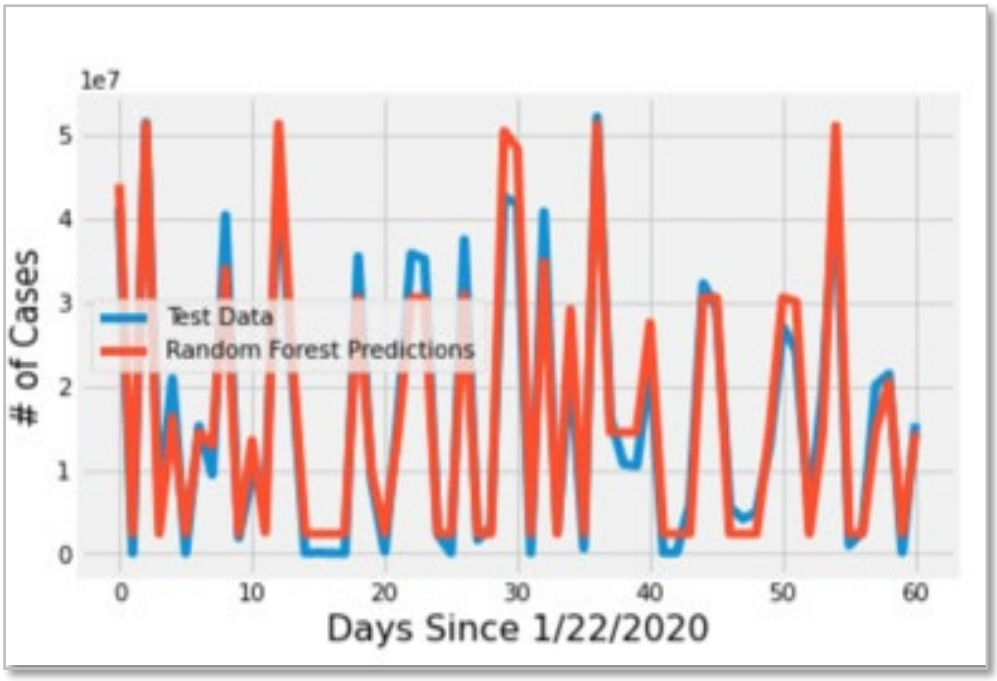


**Fig. 10.** Random Forest Predictions

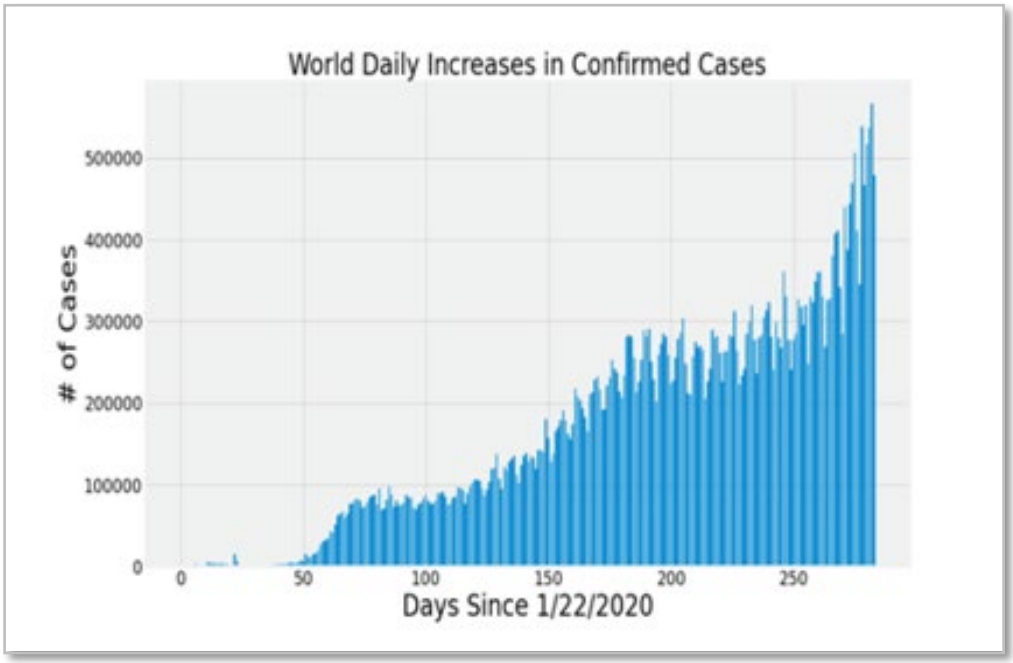


**Fig. 11.** Daily Increase in Covid-19 Cases worldwide

| | Date | Polynomial Predicted # of Confirmed Cases Worldwide |
|---|---|---|
| 0 | 11/01/2020 | 45403013.0 |
| 1 | 11/02/2020 | 45808324.0 |
| 2 | 11/03/2020 | 46215850.0 |
| 3 | 11/04/2020 | 46625597.0 |
| 4 | 11/05/2020 | 47037568.0 |
| ... | ... | ... |
| 85 | 01/25/2021 | 88269987.0 |
| 86 | 01/26/2021 | 88881933.0 |
| 87 | 01/27/2021 | 89496531.0 |
| 88 | 01/28/2021 | 90113787.0 |
| 89 | 01/29/2021 | 90733705.0 |

| | Date | SVM Predicted # of Confirmed Cases Worldwide |
|---|---|---|
| 0 | 11/01/2020 | 101640467.0 |
| 1 | 11/02/2020 | 103073651.0 |
| 2 | 11/03/2020 | 104522002.0 |
| 3 | 11/04/2020 | 105985624.0 |
| 4 | 11/05/2020 | 107464626.0 |
| ... | ... | ... |
| 85 | 01/25/2021 | 288889399.0 |
| 86 | 01/26/2021 | 292029171.0 |
| 87 | 01/27/2021 | 295194503.0 |
| 88 | 01/28/2021 | 298385535.0 |
| 89 | 01/29/2021 | 301602405.0 |

**Fig. 12.** Predictions of number of Covid-19 Cases for the next 90 days

Fig.13, shows that cases of Covid-19 have increased rapidly right from the start in many countries and have been steadily growing, particularly in the United States and India. Still, countries that have followed strict procedures such as lockdown, mask-wearing, social distancing, and isolating have reported fewer increases in cases.

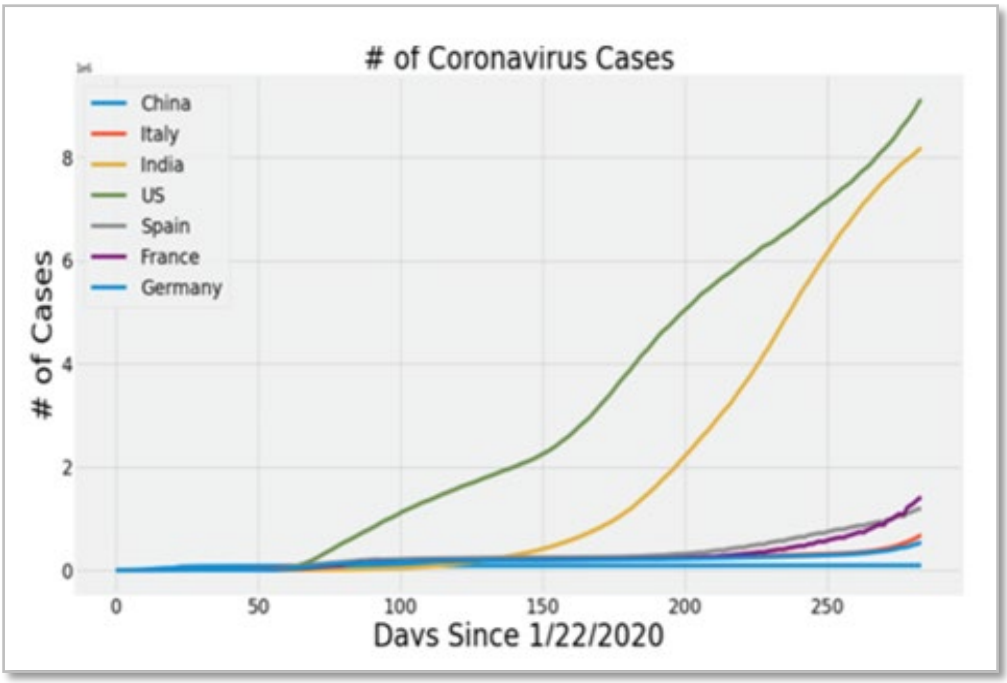


**Fig. 13.** Plot of Covid-19 increase in Cases in major Countries.

## 6. LIMITATIONS

Out of all the four models compared in this paper, Polynomial Regression turns out to be the best, and the second-best is the SVM method. The other two methods are also good in prediction only if the data sample is a time-series. For time-series research, long-short term memory networks (LSTMs) are now widely used. An LSTM is a specific form of a neural network capable of learning sequential dependencies between observations in a series, making them ideal candidates for forecasting time series. There is concern that LSTMs do not necessarily forecast at all [14]. Instead, to allow predictions of the immediately following an observation, they use lagged values. Converting the dataset into a time-series would help predict a good result, especially for the LSTM model we can work on in future papers.

## 7. CONCLUSION

From January 22, 2020, to October 31, 2020, we examined the rise in Covid-19 incidents. According to our study for a number of incidents, a number of deaths and mortality rates per country and state, we found that less Covid-19 cases have been identified in some countries that have introduced measures such as lockdown, face mask use, social distance, isolation, and travel ban restrictions compared to other countries or states that have not implemented those measures. Our forecast model shows that the number of cases of covid-19 for the next 90 days will continue to increase globally, especially in the United States, Indian is rising tremendously, in the next 90 days the United Kingdom, Spain, France and Germany will also experience a spike in cases, but not as much as the United States and Indian. The number of cases shows rapid variations. The rise is difficult to accurately predict as it can rely on numerous variables such as travel, atmosphere, number of vulnerable groups present in a country, etc. Getting some self-awareness, taking appropriate precautions, going for Covid-19 testing, and self-isolation can not only minimize the spread, but also help to enhance many people's everyday survival. Such strict steps must continue to be implemented by the government in each country to keep the spread low pending the time a vaccine is produced. Most treating centers continue to operate on the lack of supplies and are worried about insufficient beds, medical gas, food, and toilet rolls. Treatment centers should try to get more workers to ensure that employees get the required benefits and do anything to keep them safe, as most employees have raised concerns about being afraid of what will happen. Treatment centers do go through overwhelmed capacity sometimes if there are surges and must deal with the capacity problem. Since the future is not certain about the cases of Covid-19, the hospital should work on expanding its capacity level to be able to take in patients.

## REFERENCES

[1] Rafi Letzter. The coronavirus didn't start at that Wuhan' wet market.' 2020. https://www.livescience.com/covid-19-did-not-start-at-wuhan-wet-market.html

[2] Alkhushayni, Suboh M. "mHealth technology: Towards a new persuasive mobile application for caregivers that addresses motivation and usability." (2016).

[3] Alkhushayni, Suboh, and Susan McRoy. "mHealth technology: towards a new mobile application for caregivers of the elderly living with multiple chronic

conditions (ELMCC)." Proceedings of the 6th International Conference on Digital Health Conference. 2016.

[4] Kruse, Ryan, and Suboh Alkhushayni. "Identifying regional COVID-19 presence early with time series analysis." IOP SciNotes 1.2 (2020): 024003.

[5] Alkhushayni, Suboh M., Daniel C. Zellmer, and Ryan J. DeBusk. "Text emotion mining on Twitter." IOP SciNotes 1.3 (2020): 035001.

[6] Alkhushyni, Suboh M., M. Alzaleq Du'a, and Nadine L. Gadjou Kengne. "Blockchain Technology applied to Electronic Health Records." Proceedings of 32nd International Conference on. Vol. 63. 2019.

[7] Al-zaleq, Du. "Optical Fiber Communication with Vortex Modes." (2019).

[8] Andrew Joseph. 2020. Woman with novel pneumonia virus hospitalized in Thailand – the first case outside China. https://www.statnews.com/2020/01/13/woman-with-novel-pneumonia-virus-hospitalized-in-thailand-the-first-case-outside-china/

[9] Coronavirus Update:from COVID-19 Virus Pandemic Worldometer. Retrieved from https://www.worldometer s.info/coronavirus/

[10] Gianfranco Alicandro, Giuseppe Remuzzi, Carlo La Vecchia, 2020. Italy's first wave of the Covid_19 pandemic has ended: no excuses mortality in May. 2020. Vol 393, issue 10253, E27 – E28. https://doi.org/10.1016/S0140-6736(20)31865-1

[11] Josephine Ma. 2020. China is first confirmed Covid-19 case traced back to November 17. South China Morning Post. Retrieved from https://www.scmp.com/news/china/ society/article/3074991/coronavirus-chinas-first-confirmed-covid-19-case-traced-back

[12] Sujath, R., Chatterjee, J.M. & Hassanien, A.E. A machine learning forecasting model for COVID-19 pandemic in India. Stoch Environ Res Risk Assess 34, 959–972 2020

[13] Fotios Petropoulos. Spyros Makridakis. 2020. Forecasting the novel coronavirus COVID-19. https://doi.org/10.1371/journal.pone.0231236

[14] Shreshth Tuli, Shikhar Tuli, Rakesh Tuli, Sukhpal Singh Gill. 2020. Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. 11, 10022

[15] Abhay Kumar. Understand Power of Polynomials with Polynomial Regression. 2018. https://acadgild.com/blog /polynomial-regression-understand-power-of-polynomials

[16] Long short-term memory. 2020 Wikipedia. Retrieved from https://en.wikipedia.org/w/index.php?title= Long_ short-term_memory&oldid=987978443

[17] Michael Phi. 2020. Illustrated Guide to LSTM's and GRU's: A step by step explanation. Medium. Retrieved from https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21

[18] Avinash Navlani. Understanding Random Forest Classifiers in Python. 2018. https://www.datacamp.com/community/tutorials/random-forests-classifier-python

[19] CDC. 2020. Coronavirus Disease 2019 (COVID-19). Centers for Disease Control and Prevention. Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/global -covid-19/index.html

[20] Manotosh Mandal, Soovoojeet Jana, Swapan Kumar Nandi, Anupam Khatua, Sayani Adak, T.K. Kar. 2020. A model-based study on the Dynamic of Covid-19: Prediction and Control. https://doi.org/10.1016/j. chaos.2020.109889

[21] Phoebe Shanahan. 2020. Time Series Forecasting: Limitations of LSTMs. https://morioh.com/p/9202f c245ad9

[22] Datasets source: https://github.com/CSSEGISandData/ COVID