PREDICTION OF COVID-19 USING MACHINE LEARNING MODEL

Abstract

SARS-CoV-2 is officially known as Covid-19, which was first detected in Wuhan, China in December 2019. This virus has produced a global pandemic affecting all over the world. Predicting the rate of Covid-19 spread and modeling of its course have critical impact on both health system and policy makers, which makes hot research topics for machine learning researchers. Covid-19 virus is presently considered as a potential threat to mankind. This study demonstrates machine learning approaches to predict and forecast the upcoming spread of covid-19 in many countries. The ML models have been used from a long time in different application domains for identification and prioritization. The forecasting mechanisms have proved their significance to anticipate in preoperative outcomes to improve the decision making on the future course of actions. In particular, standard forecasting models such as support Vector Machine (SVM), Linear Regression (LR), and Random Forest (RF) have been used in this study to forecast the upcoming factors of Covid-19 patients. This study used four different predictions for each of these models, such as number of confirmed cases, the number of recoveries, the number of active cases, and the number of death cases. The results show that Random Forest algorithm perform best among all the used models, while SVM performs failingly in all the prediction scenarios given the available dataset.

Keywords: The forecasting mechanisms have proved their significance to anticipate in preoperative outcomes to improve the decision making on the future course of actions.

Authors

Dr. K.Santhi
Professor
CSE
S V College of Engineering
Tirupati.

Manduru Sujitha Sree
Student
CSE
S V College of Engineering
Tirupati.

Maddela Sandhya
Student
CSE
S V College of Engineering
Tirupati.

N Dedeepya
Student
CSE
S V College of Engineering
Tirupati.

I. INTRODUCTION

Machine learning is a powerful technique that can predict outcomes by learning patterns and their relationship to data through algorithms without explicit programming. It is a branch of artificial intelligence that uses acquired knowledge to make accurate predictions or decisions. Machine learning algorithms can be used to predict a variety of outcomes such as weather conditions, stock prices, customer behavior, medical diagnoses, and more. The power of machine learning lies in its ability to analyze large amounts of data and identify patterns that may be difficult for humans to discern. Another benefit of machine learning is its ability to recognize complex patterns and relationships in data. Traditional statistical methods may not be able to capture these dependencies, but machine learning algorithms can find them automatically. This can lead to more accurate predictions and information that may not have been discovered using other methods. Overall, machine learning is a valuable predictive tool in many areas. Its ability to learn from data and identify complex patterns makes it a powerful tool for making accurate predictions and improving decision-making.

The impact of the COVID-19 pandemic on people is enormous. In particular, it has raised serious public health concerns as millions of people have been infected with the virus and are suffering from serious illnesses or dying, straining health care systems and frontline workers who work hard to deliver care. The pandemic is also having a significant impact on the economy and society. Governments around the world have implemented lockdowns and other restrictions to limit the spread of the virus, disrupting daily life and causing economic hardship for businesses and individuals. In addition, the pandemic has exposed pre-existing inequalities in society, which are disproportionately affected by disadvantaged communities and people with comorbidities. Mental health has also been affected as many people suffer from high levels of stress, anxiety and depression. Overall, the COVID-19 pandemic has had a profound and significant impact on people's lives. It underlines the need for effective public health measures, social support systems and economic resilience to prepare for and deal with future crises. This pandemic continues to pose a number of challenges to medical systems around the world, including increased demand for hospital beds and acute shortages of medical equipment as many healthcare workers have become infected. The ability to make immediate clinical decisions and use healthcare resources effectively is therefore crucial. The well-established COVID-19 reverse transcriptase polymerase chain reaction (RT-PCR) diagnostic test has long been missing in developing countries. This contributes to increased infection rates and delays essential preventive measures. Effective and timely testing is critical to diagnosing COVID-19 and reducing the burden on healthcare systems. For this purpose, prediction models were developed to estimate the risk of infection by combining different characteristics. These models can help healthcare professionals assess patients, especially when healthcare resources are limited. Features used in these models include computed tomography (CT), clinical signs, and laboratory tests. Some models include all of these features.

Our goal with this study is to help in the current global crisis by developing a COVID-19 prediction system. Predictions are made for three important variables number of New confirmed cases, number of death cases, number of recoveries. In this study, the prediction problem was treated as a regression problem and as such is based on some of the supervised machine learning models Like Support Vector Machine (SVM), LinearRegression (LR) and Random Forest (RF).

1. MATERIALS AND METHODS

- **Details of Dataset:** The aim of this study is to predict the covid-19 cases by using ML models. These mainly focused on the number of confirmed cases, the number of death cases, the number of recovery cases and the number of active cases. The data set used in this study has taken from the GitHub repository named (csse_covid-19_time_series) by Johns Hopkins University. This repository contains daily time series summary tables, which includes covid-19_confirmed_global, covid-19_deaths_global, covid-19_recovered_global and covid-19_dailyreports_global. Data samples from the files are shown in Tables 1, 2, 3 and 4 respectively.

Table 1: Historical Series of Confirmed Cases of Covid-19 Worldwide

	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	...	2/28/23	3/1/23	3/2/23	3/3/23	3/4/23	3/5/23
0	NaN	Afghanistan	33.93911	67.709953	0	0	0	0	0	0	...	209322	209340	209358	209362	209369	209390
1	NaN	Albania	41.15330	20.168300	0	0	0	0	0	0	...	334391	334408	334408	334427	334427	334427
2	NaN	Algeria	28.03390	1.659600	0	0	0	0	0	0	...	271441	271448	271463	271469	271469	271477
3	NaN	Andorra	42.50630	1.521800	0	0	0	0	0	0	...	47866	47875	47875	47875	47875	47875
4	NaN	Angola	-11.20270	17.873900	0	0	0	0	0	0	...	105255	105277	105277	105277	105277	105277

5 rows × 1147 columns

Table 2: Time Series of Deaths From Covid-19 Patients Worldwide

	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	...	2/28/23	3/1/23	3/2/23	3/3/23	3/4/23	3/5/23	3/6/
0	NaN	Afghanistan	33.93911	67.709953	0	0	0	0	0	0	...	7896	7896	7896	7896	7896	7896	78
1	NaN	Albania	41.15330	20.168300	0	0	0	0	0	0	...	3598	3598	3598	3598	3598	3598	35
2	NaN	Algeria	28.03390	1.659600	0	0	0	0	0	0	...	6881	6881	6881	6881	6881	6881	68
3	NaN	Andorra	42.50630	1.521800	0	0	0	0	0	0	...	165	165	165	165	165	165	1
4	NaN	Angola	-11.20270	17.873900	0	0	0	0	0	0	...	1933	1933	1933	1933	1933	1933	19

5 rows × 1147 columns

Table 3: Chronological Series of Recoveries From Covid-19 in the World

	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	...	2/28/23	3/1/23	3/2/23	3/3/23	3/4/23	3/5/23	3/6/
0	NaN	Afghanistan	33.93911	67.709953	0	0	0	0	0	0	...	0	0	0	0	0	0	
1	NaN	Albania	41.15330	20.168300	0	0	0	0	0	0	...	0	0	0	0	0	0	
2	NaN	Algeria	28.03390	1.659600	0	0	0	0	0	0	...	0	0	0	0	0	0	
3	NaN	Andorra	42.50630	1.521800	0	0	0	0	0	0	...	0	0	0	0	0	0	
4	NaN	Angola	-11.20270	17.873900	0	0	0	0	0	0	...	0	0	0	0	0	0	

5 rows × 1147 columns

Table 4: Latest Daily COVID-19 Patient Reports, Global Time Series

FIPS	Admin2	Province_State	Country_Region	Last_Update	Lat	Long_	Confirmed	Deaths	Recovered	Active	Combined_Key	Incidence_Rate
45001.0	Abbeville	South Carolina	US	2020-07-16 04:44:59	34.223334	-82.461707	165	1	0	164	Abbeville, South Carolina, US	672.728014
22001.0	Acadia	Louisiana	US	2020-07-16 04:44:59	30.295065	-92.414197	1510	46	0	1464	Acadia, Louisiana, US	2433.717463
51001.0	Accomack	Virginia	US	2020-07-16 04:44:59	37.767072	-75.632346	1045	14	0	1031	Accomack, Virginia, US	3233.692289
16001.0	Ada	Idaho	US	2020-07-16 04:44:59	43.452658	-116.241552	4836	28	0	4808	Ada, Idaho, US	1004.179930
19001.0	Adair	Iowa	US	2020-07-16 04:44:59	41.330756	-94.471059	17	0	0	17	Adair, Iowa, US	237.695749

II. SUPERVISED LEARNING MODELS

Supervised learning involves using labeled datasets to train algorithms to make predictions on new data based on relationships learned from past datasets. This approach is used in various fields such as computer vision, speech recognition, and medical diagnostics. Several supervised learning algorithms exist, including linear regression, logistic regression, decision trees, random forest, k-nearest neighbors (KNN), support vector machine (SVM), and neural networks.

In this Paper it includes the algorithms of Supervised Learning Machine Learning Models

1. Linear Regression
2. Support Vector Machine
3. Random Forest

1. **Linear Regression:** Linear regression is the simplest statistical regression method used for predictive analysis in machine learning. Linear regression shows the linear relationship between the independent variable and the dependent variable known as Linear Regression.

LR is to obtain a line which is best fits the data. Error is the difference between the point to the regression line. There are two factors (x, y) that are involved in linear regression analysis. The equation below shows how y is related to x known as.

$$\boldsymbol{y = \beta 0 + \beta 1_X + \varepsilon}$$

Here, ε is the error term of linear regression. The error term here uses to account the variability between both x and y, β0 represents y-intercept, β1 represents slope. To get the best fit implies the difference between the actual values and predicted values should be minimum, so this minimization problem can be represented as: minimize.

$$\boldsymbol{minimize \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2}$$

2. **Support Vector Machine:** Support Vector Machine chooses the data points that will help in creating the hyperplane and those are known as support vectors, so it is termed as

Support Vector Machine. SVM is used to find the best fit line in the hyperplane which has maximum number of points. Hyperplanes is used classify the data points. To do this SVM finds the largest margin between the hyperplanes that means highest distances between the two classes.

This technique depends on mathematical functions. The set of functions called kernel transforms the data inputs into the desired form. SVM solves the regression problems using a linear function, so while dealing with problems of non-linear regression, it maps the input vector to n-dimensional space called a feature space. The linear function is as follows.

$$\boldsymbol{f(X) = X^r \beta + b}$$

The objective is to make it as flat as possible thus to find the value of f (x) with (βrβ) as minimal norm values. So, the model, the increase in new features problem fits in minimization function as:

$$\boldsymbol{J(\beta) = \frac{1}{2} \beta^r \beta}$$

3. **Random Forest:** Random Forest is a type of Ensemble Learning method used to improve the accuracy of predictive models by combining multiple decision trees. Instead of relying on a single decision tree, Random Forest uses a collection of trees to generate predictions and makes a final decision based on the majority vote. The more trees in the forest, the higher the accuracy and the lower the risk of overfitting.

Random Forest also provides a measure of feature importance, which can be useful for feature selection and data interpretation. This measure is obtained by calculating the importance of each feature in each tree and then normalizing it between 0 and 1 based on the total number of trees in the forest.

It gives the measure of feature importance, helpful for feature selection and data interpretation. Normalizing value between 0 and 1 by as follows.

$$\boldsymbol{normfi_i = \frac{fi_i}{\sum_{j \in all\ features} fi_i}}$$

Feature's importance value on each tree is calculated and divided by the total number of trees as follows.

$$\boldsymbol{RFfi_i = \frac{\sum_{j \in all\ tress} normafi_{ij}}{T}}$$

- $RFfi_i$ = Importance of Features
- $Normafi_{ij}$ = Normalized Feature Importance
- T = Total of trees

III. PERFORMANCE EVALUATION

In this Paper it demonstrates each trained model to test the performance by some of the parameters like R-Square, Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE).

1. **R-Square (R^2 Score):** R-Square is a statistical measure in a regression model that determines the proportion of variance in the dependent variable followed by the independent variable. R-Squared gives how well the data fit the regression model (the goodness of fit).

   The high R^2 Score shows the how well the trained model is. R^2 Score is a linear model that gives the percentage of variation independent variable. It can be found as:

$$R^2 = \frac{Variance\ explained\ by\ model}{Total\ variance}$$

2. **Mean Absolute Error (MAE):** Mean Absolute Error (MAE) is a statistical measure that evaluates the accuracy of a predictive model. It calculates the average magnitude of errors in a set of predictions, regardless of their direction, to indicate how far the predictions are from the actual values on average. A limitation of MAE is that it treats all errors equally, irrespective of their magnitude. Despite this, MAE remains a popular choice for evaluating model performance.

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|y_i - \hat{y}_j\right|$$

   MAE tells us how big of an error, can expect from the forecast on average.

3. **Mean Squared Error (MSE):** Mean Squared Error (MSE) is a statistical measure used to evaluate the accuracy of a predictive model. It considers the average squared difference between the predicted and actual values in a set of predictions, accounting for the magnitude and direction of the errors.MSE is more sensitive to outliers than Mean Absolute Error (MAE), which makes it less appropriate in some scenarios. Therefore, it is often used in combination with other error metrics to assess model performance. It is measured as follows.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

4. **Root Mean Square Error (Rmse):** Root Mean Squared Error (RMSE) is a widely used statistical measure for evaluating the performance of regression models. It is closely related to Mean Squared Error (MSE). RMSE takes into account both the magnitude and direction of errors and is often used in situations where larger errors have a more significant impact. It is more sensitive to outliers than the Mean Absolute Error (MAE). So, it is best used in conjunction with other error metrics to provide a more comprehensive evaluation of model performance.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

IV. ARCHITECTURE

This Paper deals with Covid-19 Predictions. As we know Corona Virus has ruled our lives since 2019. Thousands of deaths from near and dear families. As there was no cure to the disease and became threat to human life. To contribute the society, this study has conducted forecasting for next 10 upcoming days. How the cases would be for next 10 days will be forecasted and gives the results.

In this paper, the dataset which contains time series summary tables which includes deaths, confirmed, recovered, active cases. Initially,pre-processed the dataset to find the number of confirmed cases, number of death cases, number of recovered cases on particular dates. Initial Datasets shown in Table 5 and the Resultant Datasets shown in Table 6 & 7.

Table 5: Time Series Table of Cases of Covid-19 Data

	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	...	2/28/23	3/1/23	3/2/23	3/3/23	3/4/23	3/5/23
0	NaN	Afghanistan	33.93911	67.709953	0	0	0	0	0	0	...	209322	209340	209358	209362	209369	209390
1	NaN	Albania	41.15330	20.168300	0	0	0	0	0	0	...	334391	334408	334408	334427	334427	334427
2	NaN	Algeria	28.03390	1.659600	0	0	0	0	0	0	...	271441	271448	271463	271469	271469	271477
3	NaN	Andorra	42.50630	1.521800	0	0	0	0	0	0	...	47866	47875	47875	47875	47875	47875
4	NaN	Angola	-11.20270	17.873900	0	0	0	0	0	0	...	105255	105277	105277	105277	105277	105277

5 rows × 1147 columns

Table 6: Country-Wise Covid Cases of Sample Data

	Country Name	Confirmed	Deaths	Recoveries	Active	Mortality Rate
0	US	3513573	137285	1075882	2300406	0.039073
1	Brazil	1973933	75604	1350098	548231	0.038301
2	India	968857	24914	612768	331175	0.025715
3	Russia	745197	11753	522375	211069	0.015772

Table 7: State-Wise Covid Cases of Sample Data

	Province/State Name	Country	Confirmed Cases	Deaths	Recoveries	Mortality Rate
0	New York	US	404006	32263	0	0.079858
1	Sao Paulo	Brazil	393176	18640	246941	0.047409
2	California	US	363222	7300	0	0.020098
3	Florida	US	301810	4521	0	0.014980

Initially Covid-19 Dataset was pre-processed to get the required features extraction like Country-wise Covid Cases, State-wise Covid Cases and world-wide covid cases and thishelps to forecast the covid cases for next 10 days. it splits the data into Training and Testing data.

The Machine Learning Algorithms like Linear Regression, Support Vector Machine and Random Forest used for the training dataset. These models are trained for future forecast of covid cases. Then, it has been evaluated with parameters like R-Square, Mean Square Error, Mean Absolute Error and Root Mean Square Error. This work flow is represented in the Figure 1.

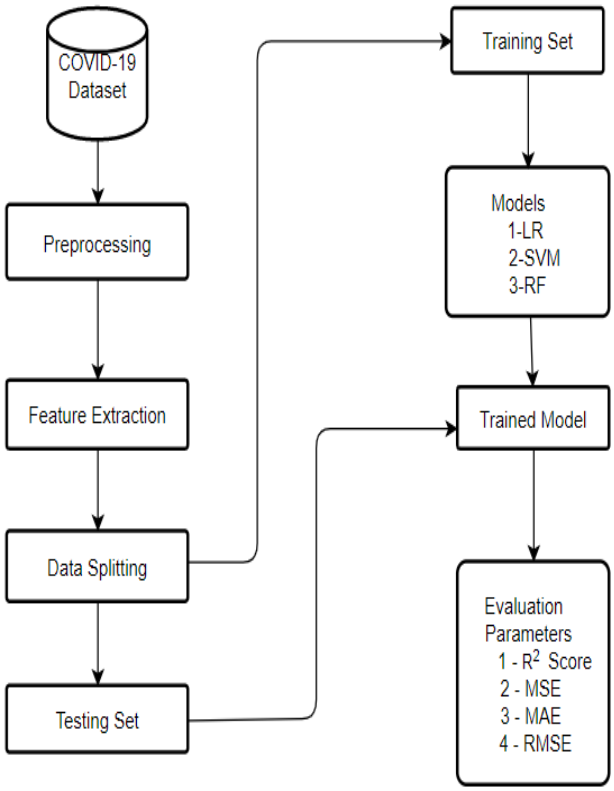


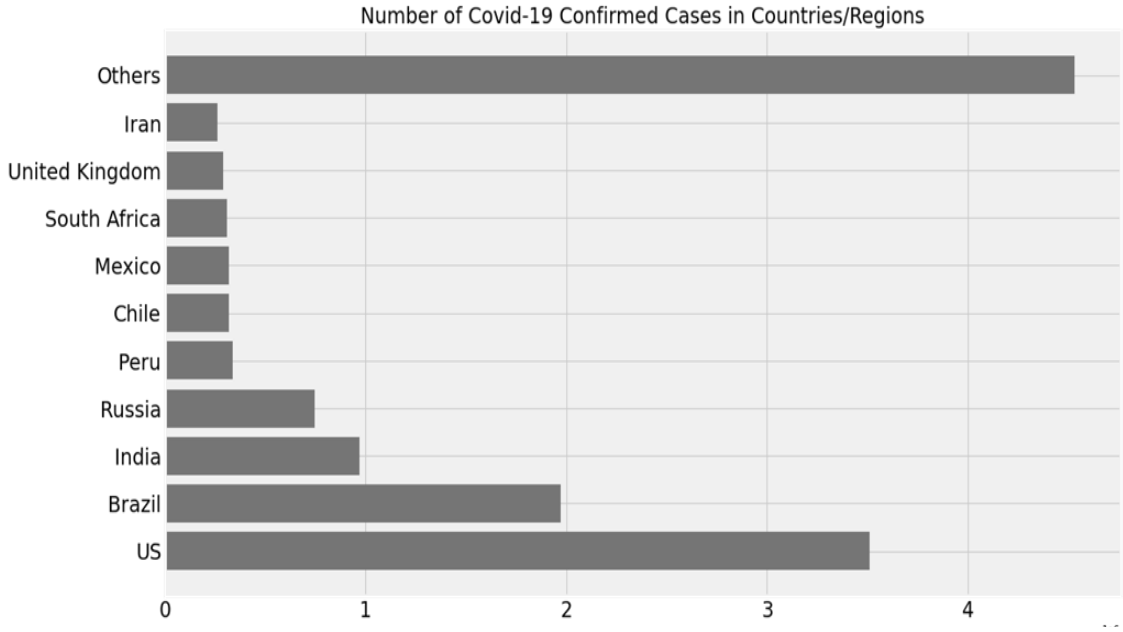
Figure 1: Process Flow of Covid-19

V. VISUALIZATIONS

In this paper it includes some of the visualizations that has compared the features like countries, states while training the models.

In Figure 2, It shows the number of confirmed cases of covid-19 in Countries.

Figure 2: Confirmed Cases in Countrywide



Then compared Confirmed Covid Cases in India and it shown in figure 3

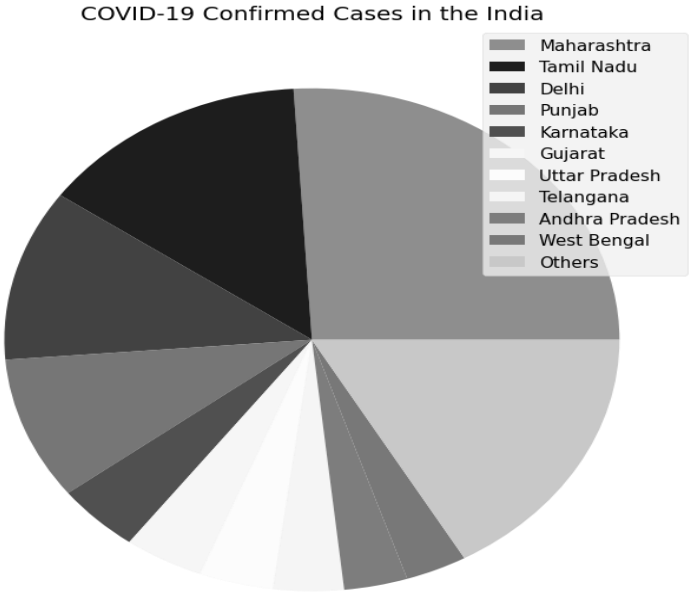


Figure 3: Confirmed Covid Cases in India

This figure 4, compares the ConfirmedCovid Cases within Provinces/States.

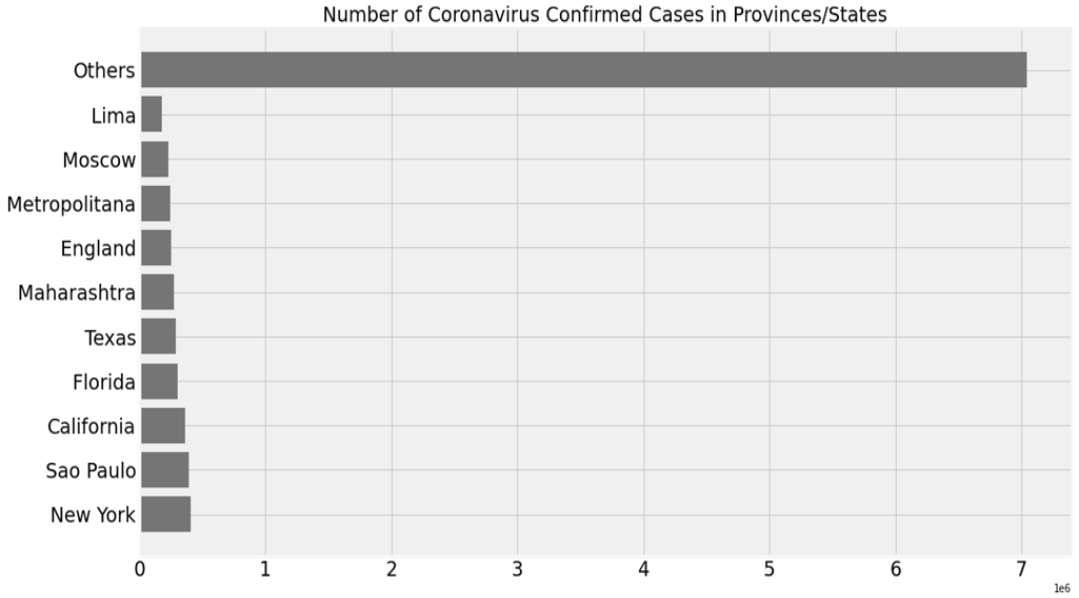


Figure 4: Confirmed Cases in Province/Regions

Here compared worldwide daily increases of covid confirmed cases and it was shown in the figure 5.

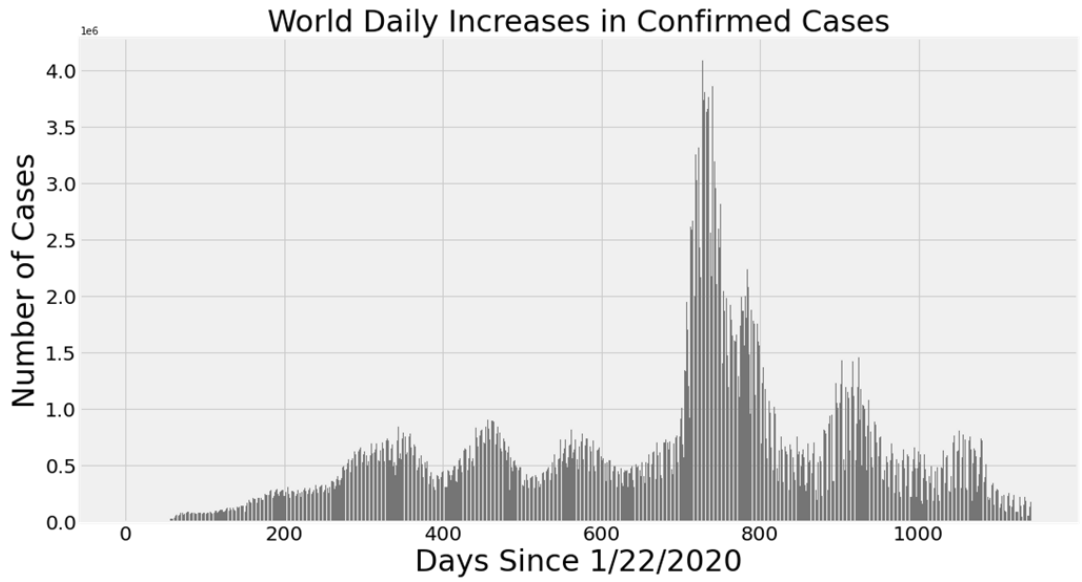


Figure 5: Worldwide Daily Increase of Covid Confirmed Cases

VI. RESULTS AND ANALYSIS

This study develops a model to Future Forecast the Covid-19 cases by using Machine Learning Models.

The Dataset used contains the daily time series of number of confirmed cases, number of death cases and number of recovery cases of Covid-19 from worldwide. As number of death cases was increasing day by day. It panics the people towards the Covid. As this paper demonstrates the number of confirmed cases, number of deaths and number of recovery cases would be for next 10 days will be known and this will be predicted using the machine learning models SVM, LR and RF.

1. **Confirmed cases prediction:** Here In this model, Random Forest performs better to predict the number of confirmed cases of covid-19 compared to the remaining and Support Vector Machine performs poor. It is shown in the Table 8.

Table 8: Model Performance of Future Forecasting for Confirmed Cases. This Below Figure 6,7,8 Shows the Prediction of Covid-19 Cases Performances by using LR, SVM and RF

MODEL	R-SQUARE	MAE	MSE	RMSE
LR	0.98	21408367.124	732705031733152.2	27068524.742
SVM	0.53	207739921.743	6.5144	255234841.954
RF	0.99	643745.328	830421044095.41	911274.40

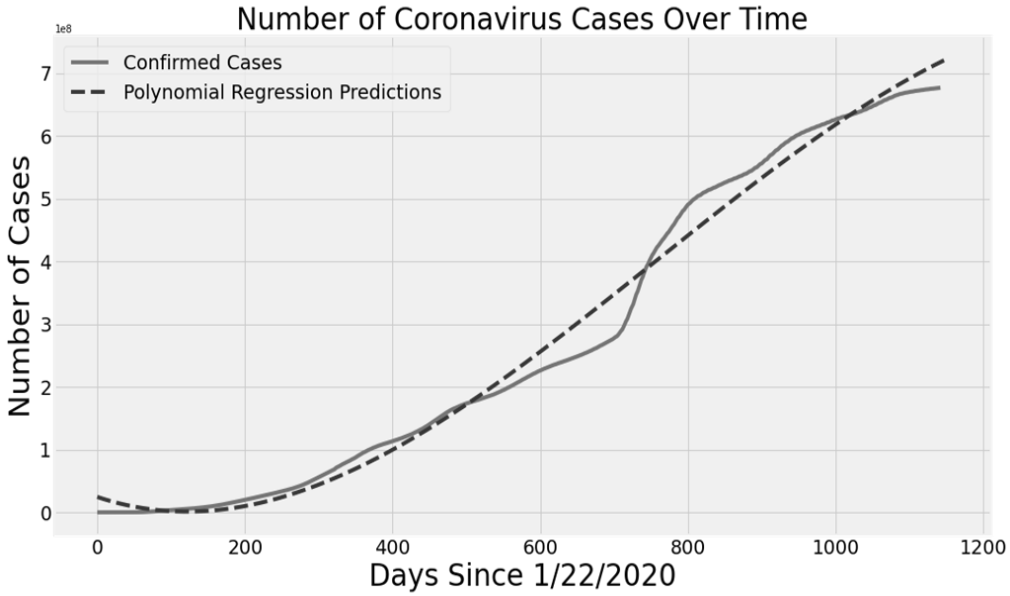


Figure 6: Confirmed Cases prediction by LR

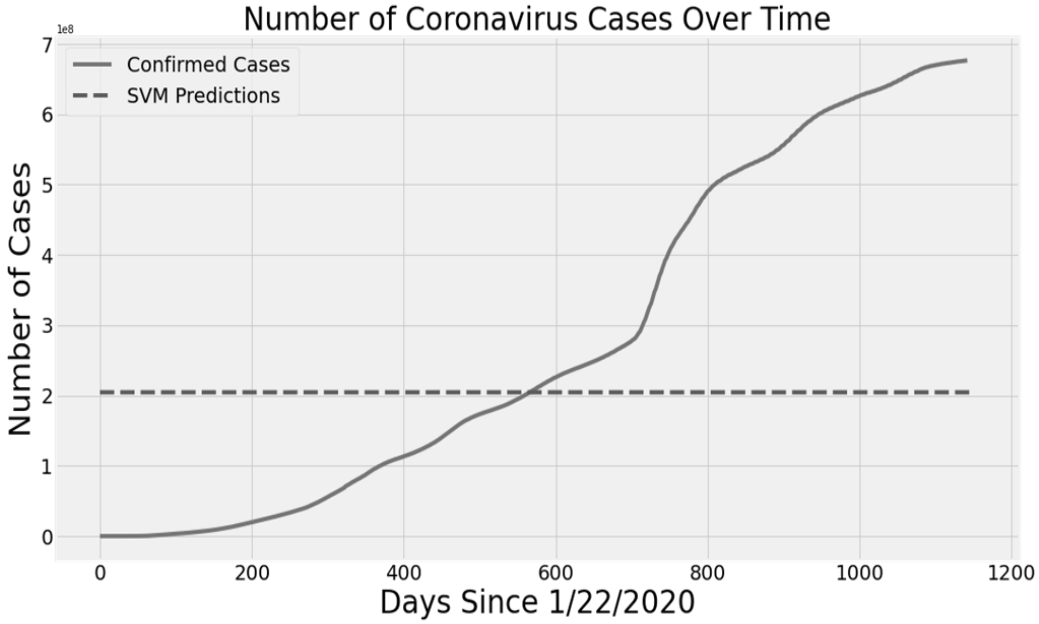


Figure 7: Confirmed Cases prediction by SVM.

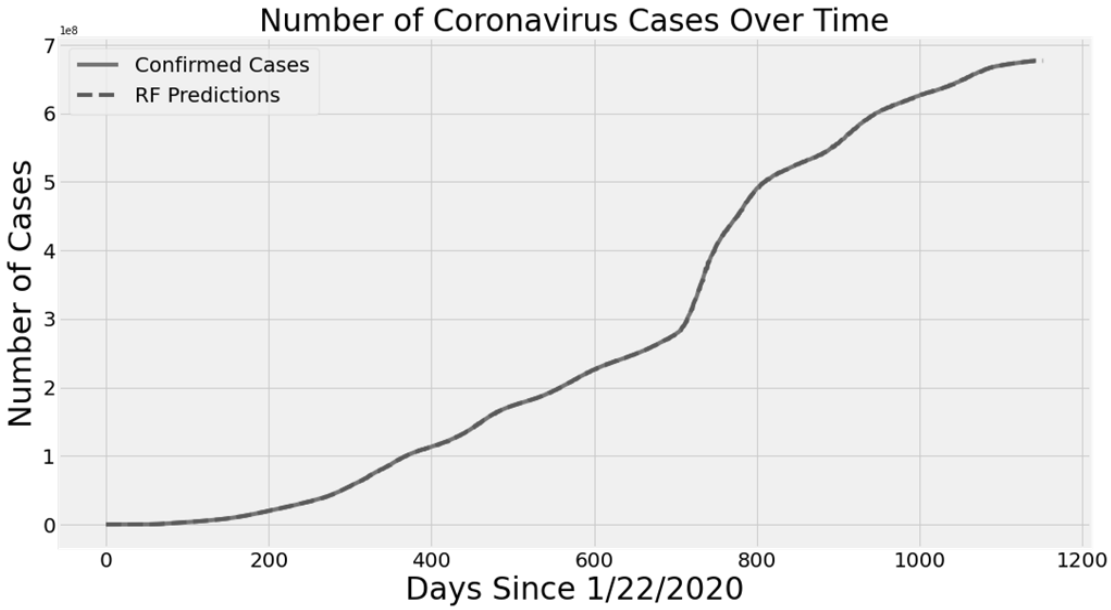


Figure 8: Confirmed Cases prediction by RF.

By Above graphs, shows the Random Forest predictions on Confirmed Cases of Covid-19 was accurate and the confirmed cases predicted by SVM was poor.

2. **Death Cases Prediction:** Here, in this Model Number of Death cases in Covid-19 was predicted by the four models of ML, it will look into the RF, SVM and LR predictions of Death Cases of Covid-19. Its performance evaluation was shown in the Table 9.

Table 9: Model Performance of Future Forecasting of Death Cases

Now, Figure 9,10,11 shows the Future Forecasting of Death Cases of Covid-19 using SVM, RF and LR. These graphs

MODEL	R-SQUARE	MAE	MSE
LR	0.97	280709.41	113241575495.19
SVM	0.55	2112089.30	5652341916550.36
RF	0.99	6113.38	63258807.12

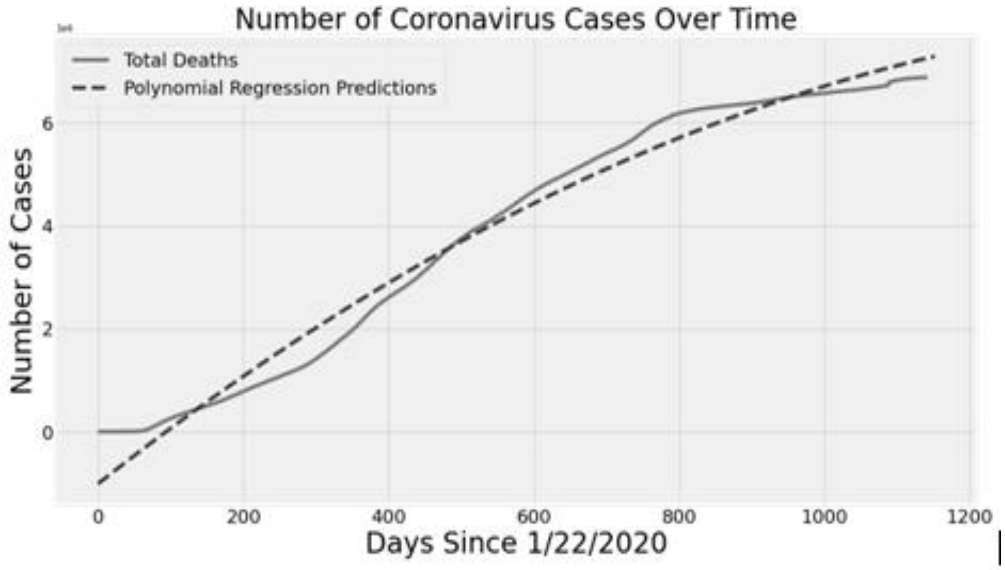


Figure 9: Death Cases Prediction by LR

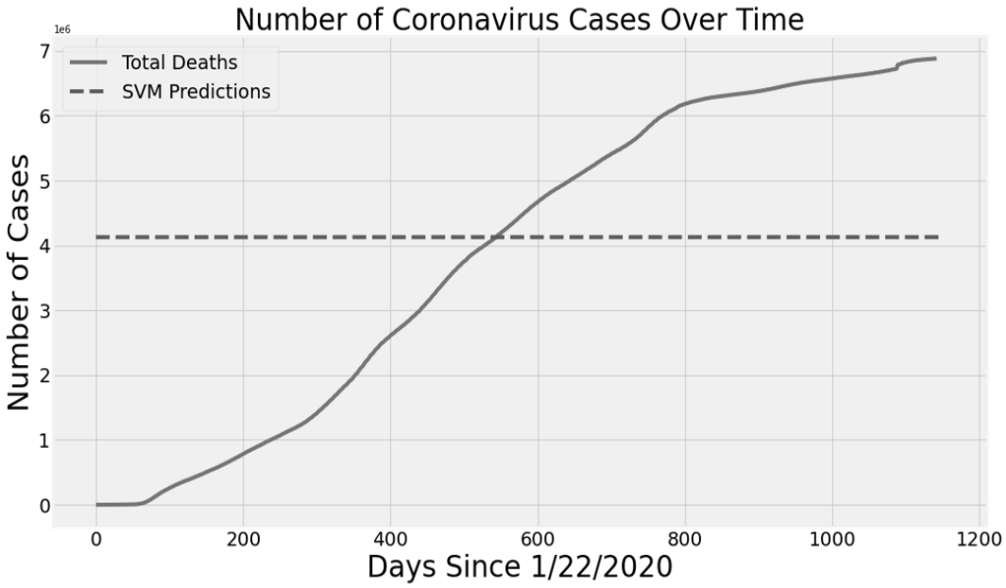


Figure 10: Death Cases Predictions by SVM

MODEL	R-SQUARE	MAE	MSE
LR	0.58	11886616.01	357859487946601.94
SVM	0.53	18227375.77	1169706429124762.2
RF	0.99	109533.47	40193607450.30

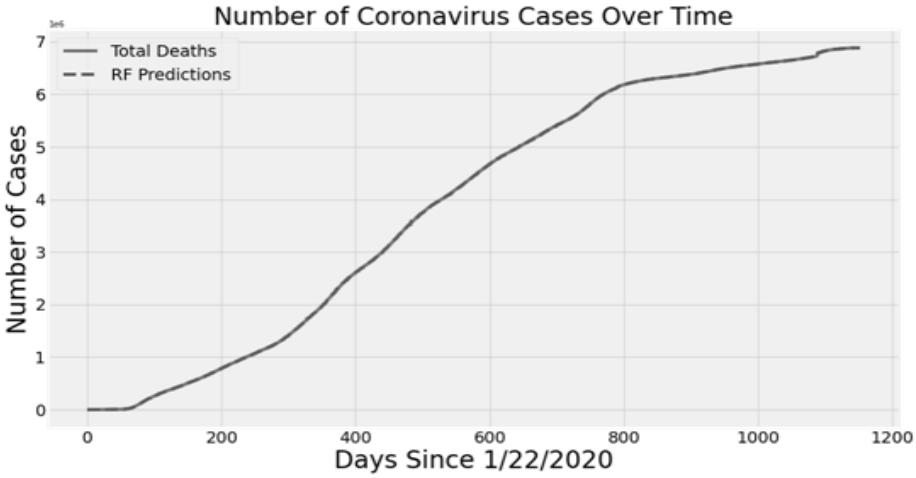


Figure 11: Death Cases Prediction by RF

The above graph predicts the Random Forest algorithmis more accurate when compared with the SVM algorithm to predict the number of Death Cases of Covid-19.

3. **Recovery Cases Prediction:** Here In this model, Random Forest performs better to predict the number of Recovered cases of covid-19 compared to the remaining and Support Vector Machine performs poor. It is shown in the Table 10.

Table 10: Model Performance of Future Forecasting of Recovery Cases

Now, Figure 12,13,14 shows the Future Forecasting of Recovery Cases of Covid-19 using SVM, RF and LR. These graphs

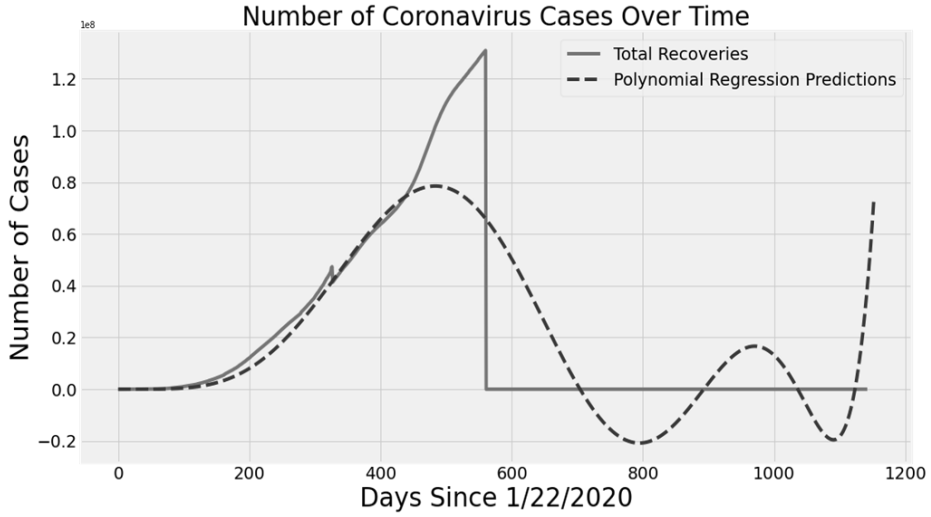


Figure 12: Recovery Cases Predicted by LR

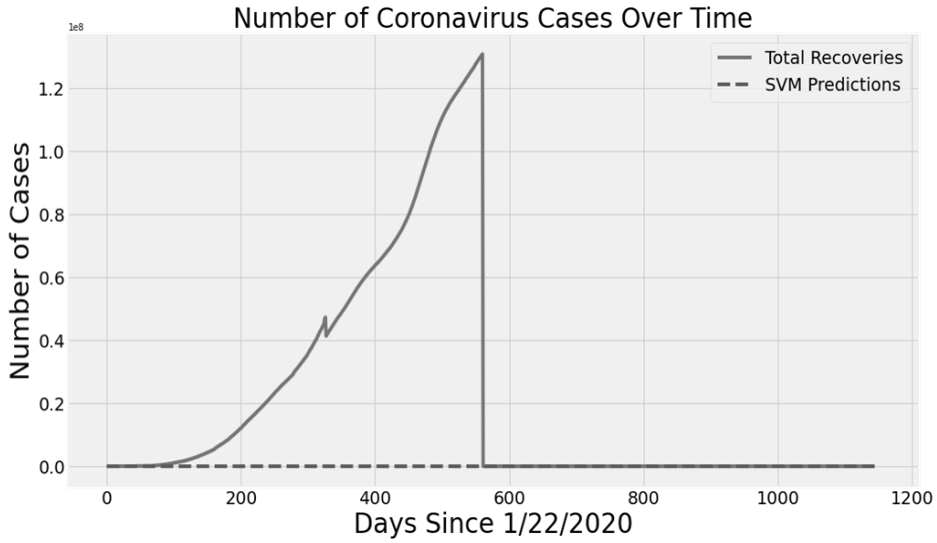


Figure 13: Recovery Cases Predicted by SVM

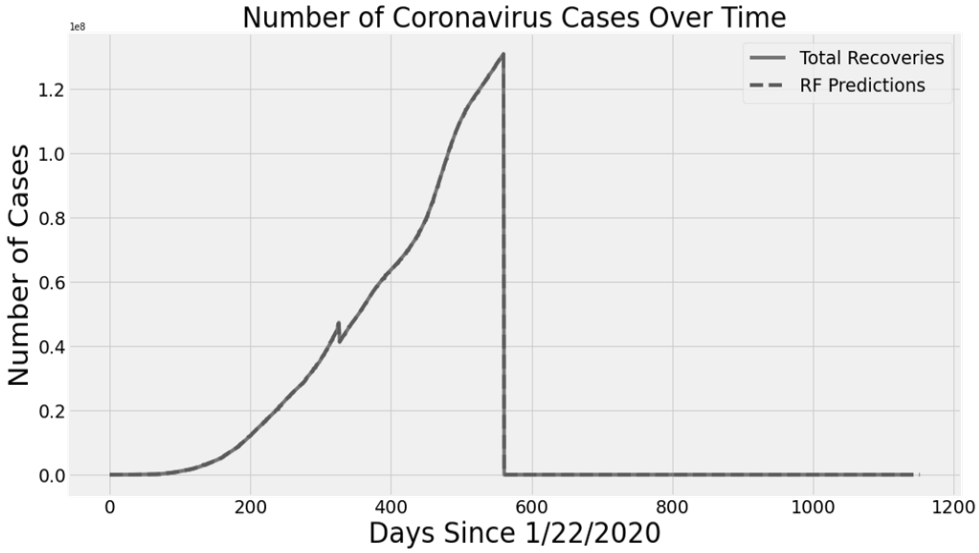


Figure 14: Recovery Cases Predicted by RF

VII. CONCLUSION

The Period of COVID-19 pandemic gave a massive global crisis. Some researchers and higher authorities' agencies throughout the world have declared that the pandemic can affect a massive portion of the world population. In this study, an ML-based prediction model has been built for predicting the risk of COVID-19 outbreak globally. The system analyses dataset containing the day-wise actual past data and makes predictions for upcoming days using machine learning algorithms. The results of the study prove that RF performs best in the current forecasting domain given the nature and size of the dataset. LR also perform well for forecasting to some extent to predict death rate and confirm cases. According to the results of these two models, the death rates will increase in upcoming days, and recoveries rate will be slowed down. SVM produces poor results in all scenarios because of the ups and

downs in the dataset values. Overall,this study concludes the model predictions according to the current scenario, which may be helpful to understand the upcoming situation. The study helps to aware the covid-19 cases for the next 10 days. This study will be enhanced continuously in the future course, next it plans to explore the prediction methodology using the updated dataset and use the most accurate and appropriate ML methods for forecasting. Real-time live forecasting will be one of the primaries focuses in our future work.

REFERENCES

[1] Statistical and Machine Learning forecasting methods: Concerns and ways forward and available at https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0194889
[2] COVID-19 Future Forecasting Using Supervised Machine Learning Models and available at IEEE: https://ieeexplore.ieee.org/document/9099302
[3] Machine learning strategies for time series forecasting,https://www.researchgate.net/publication/236941795_Machine_Learning_Strategies_for_Time_Series_Forecasting.
[4] Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis, Available at https://www.researchgate.net/publication/352572400_Machine_Learning_Algorithms_For_Breast_Cancer_Prediction_And_Diagnosis.
[5] Long-term forecasts of the COVID-19 epidemic: a dangerous idea and available at https://pubmed.ncbi.nlm.nih.gov/32876321/
[6] Johns Hopkins University Data Repository, Accessed: Mar. 27, 2021. [Online]. Available at https://github.com/CSSEGISandData/COVID19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_US.csv.
[7] Machine learning techniques in disease forecasting: a case study on rice blast prediction. Available at https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-7-485
[8] Consideration of manufacturing data to apply machine learning methods for predictive manufacturing. Available at https://www.semanticscholar.org/paper/Consideration-of-manufacturing-data-to-apply-for-Han-Chi/1ee465cbb62c966ef756b8326eb530c0d8fd85bd
[9] Review Paper for Detection of COVID-19 from Medical Images and/ or Symptoms of Patient using Machine Learning Approaches. Available at https://ieeexplore.ieee.org/document/9336799
[10] COVID-19 Infection Detection Using Machine Learning and available at https://ieeexplore.ieee.org/document/9671700.
[11] K. Santhi, Dr. A. Rama Mohan Reddy, " A Novel Approach for Heart Disease Prediction Based on Risk Factors Using Machine Learning" in Lecture Notes in Networks and Systems, , Springer, September 2021.
[12] K. Santhi, Dr. A. Rama Mohan Reddy, "An Integrated Ensemble Learning Model for Multi Class Imbalanced Datasets" , International Journal of Advanced Science and Technology Volume 29, Issue 8, 573-587,2020.
[13] K. Santhi, Dr. A. Rama Mohan Reddy, "An Automated Framework For Coronary Analysis From Coronary Cine Angiograms Using Machine Learning and Image Analysis Techniques" , IT in Industry,Volume 9, Issue 1, 2203-1731,2021.
[14] K. Santhi, Dr. A. Rama Mohan Reddy, "A Systematic Methodology on Class Imbalanced Problems involved in the Classification ofReal-World Datasets" , International Journal of Recent Technology and Engineering Volume 8, Issue 3, 2277-3878,2019.

