International Journal of Scientific Research in Science, Engineering and Technology

Print ISSN - 2395-1990
Online ISSN : 2394-4099

OPEN ACCESS

Available Online at :www.ijsrset.com
doi : https://doi.org/10.32628/IJSRSET25122182

Forecasting Covid-19 Trends Utilizing Linear Regression for Predictive Modelling Of Daily Incidence, Mortality and Recovery Rates

Saloni Bharat Gangar[1], Pratibha C. Kaladeep (Pratibha S. Yalagi)[2]

[1]Student, Department of Computer Science & Engineering, Walchand Institute of Technology, Solapur, Maharashtra, India

[2]Associate Professor, Department of Computer Science & Engineering, Walchand Institute of Technology, Solapur, Maharashtra, India

ARTICLEINFO

Article History:

Accepted : 24 April 2025
Published: 29 April 2025

Publication Issue :
Volume 12, Issue 2
March-April-2025

Page Number :
796-802

ABSTRACT

The dissemination of the Coronavirus across various regions worldwide has led to a substantial number of fatalities and has triggered a decline in the global economy. This situation continues to serve as a significant warning regarding public health and is recognized as one of the major pandemics in the annals of history. The purpose of this project is to offer an in-depth exploration of how different Machine Learning models can be effectively utilized in practical situations. Forecasting techniques based on data analytics have proven their value in predicting perioperative outcomes, which aids in making informed decisions regarding future actions. Regression models in data analysis have been applied in numerous contexts that necessitate the identification and highlighting of adverse factors that pose risks. Several predictive methodologies are primarily utilized to tackle forecasting issues. This research showcases the potential of models to estimate the number of forthcoming COVID-19 patients, which is currently viewed as a significant threat to public health. In particular, linear regression, a standard forecasting model, has been employed in this study to assess the cautionary factors related to COVID-19. The regression analysis models yield three distinct predictions: the anticipated number of new cases, recoveries, and deaths over the next 14 day

Keywords - COVID-19, Machine learning, Models, Prediction, Time series forecasting, linear Regression, Database, algorithm.

INTRODUCTION

The COVID-19 pandemic began in Wuhan, China, on November 17, 2019, and has since spread across the globe, leading to a significant health crisis. As of December 31, 2020, the virus had resulted in nearly 1.8 million deaths and infected over 95 million people in 228 countries. The pandemic has not only caused extensive economic turmoil but has also led to immense suffering worldwide. In response, various countries have enacted a range of precautionary measures. To curb the transmission of the virus, nations have undertaken numerous proactive steps. The disease follows specific growth patterns that are nonlinear and dynamic, with case numbers fluctuating based on seasonal factors, population density, and other influences. Research into COVID-19 predictive analysis has become a vital focus, assisting governments in strategizing to prevent further spread. To effectively combat the pandemic, it is essential to project the disease's trajectory by considering case counts alongside mortality and recovery statistics. Reliable forecasting can empower health and governmental bodies to develop effective health strategies to mitigate the pandemic's expected repercussions. The second wave of COVID-19 was notably more fatal than the first, resulting in an increased death toll.

AIM

The basic aim of the project is to analyze the Corona virus cases and available data to predict the future behavior of Covid-19 in Solapur city Maharashtra. Which helps to take proper decisions and preventive measures related to the virus.

PROBLEM STATEMENT

On March 11, 2020, the World Health Organization declared the coronavirus outbreak a global pandemic. With nearly 5 million deaths reported globally, the COVID-19 pandemic is viewed as the most severe and critical event following World War II. Given these unprecedented conditions, the predictive analysis of COVID-19 has become a vital research focus to aid governments in planning and preventing the further escalation of this disease.

PROPOSED SOLUTION

It is essential to comprehend the rapid development and dynamics of the pandemic. In this project, a Linear Regression method utilizing machine learning has been established for the detection and forecasting of active cases, fatalities, and recoveries from the COVID-19 database. These models for detection and prediction are instrumental in estimating future cases across various regions of the country. The application of Linear Regression through machine learning serves as a valuable tool for enhancing decision-making regarding future forecasts. Machine learning encompasses numerous predictive models for future scenarios, which are based on training and testing datasets, including both supervised and unsupervised algorithms. A variety of prediction methods are employed to address forecasting challenges. This project illustrates that the number of new infections, recoveries, and deaths associated with SARS-CoV-2 are currently regarded as significant threats to human health by machine learning models. The Support Vector Machine model, a type of supervised machine learning tool, has been utilized for predictions, including the linear regression method. Two predictions are made: the number of newly infected cases and the number of recovery cases, which provide insights into the overall situation. Two datasets are analyzed, one prior to the administration of a pre-tested vaccine and the other following it. The results of this dataset are derived from the linear regression machine learning.

OBJECTIVES

The forecasting is done with the help of for the three important variables are as follows.

1) The number of new confirmed cases.

2) The number of recovered cases.
3) The number of death cases

METHODOLOGY AND INVESTIGATION

A. Model Selection and Implementation

The process of implementation begins with the necessary libraries being imported, which include pandas, numpy, matplotlib, sklearn, seaborn, and warnings. Following this, the matplotlib converter is registered, the seaborn style is set, and warnings are filtered out. The next phase involves reading the dataset named "combined_Kazakhstan.csv," which consolidates three time series datasets. Daily confirmed, recovered, and death cases are calculated by determining the daily figures from the Confirmed, Recovered, and Death columns, utilizing the shift method to subtract the previous day's cases. Furthermore, new columns—Confirmed past day, Confirmed 2 days ago, and Confirmed 3 days ago—are established to incorporate the prior values of the Daily confirmed column as features for predicting the current daily confirmed value. This methodology is employed to capture the trends and patterns inherent in the time series data. By including these lagged values as features, the model can effectively learn from historical data patterns, resulting in more precise predictions. The same methodology is applied to the recovered and death cases. For the prediction model concerning Covid-19 cases, a selection of machine learning algorithms has been made.

B. Linear regression method by machine learning

The application of machine learning in linear regression analysis represents a highly effective approach for understanding the relationship between two variables. In this framework, one variable is designated as the dependent variable, and the other is considered the independent variable. A model that incorporates a single independent variable is known as linear regression, whereas a model that includes multiple independent variables is referred to as multiple linear regression. The standard representation of the linear regression equation is articulated in the following manner

$y = P_0 + P_1 X + e$

A significant drawback of the linear regression method is its tendency to utilize the mean values of both input and output variables to define their interrelationship. The mean alone fails to adequately represent the complexities of a single variable, rendering the linear regression model an unreliable method for predicting relationships between variables. As a result, various factors that limit the application of Simple Linear Regression (SLR) have been elucidated through the Multiple Linear Regression (MLR) model. In contrast to SLR, which involves a single independent variable, the MLR technique employs two or more independent variables in conjunction with one dependent variable to evaluate their interconnections. The general expression for multiple linear regression is articulated as follows

$y = P_0 + P_1 X_1 + P2\ X2$……....$Pn\ Xn + e$

Where,

y - The predicted value of the dependent variable

P0 - Value of y when all dependent variables are equal to zero

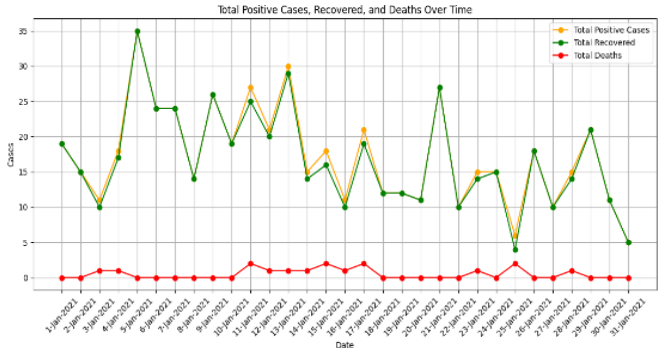
$P_1\ X_1$- The regression coefficient (P1) of the first independent variable (X1)

Pn Xn - The regression coefficient of the last independent variable

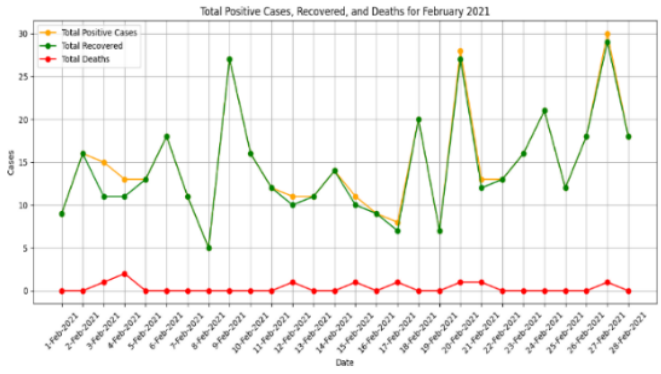
e – Errors in Models.

RESULTS AND DISCUSSION

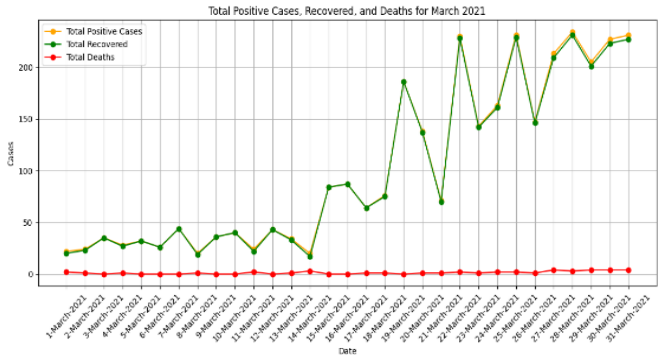
The data collected over the pandemic period from 1-jan-2021 to 31-dec-2021 were analyzed. For the detailed study of following graphs you will get an idea about the active cases, recovered cases and death cases are displayed on graph. Also graphs showing rate of transmission and give idea to take right decision on time to control covid-19 active cases and reduce the death ratio.



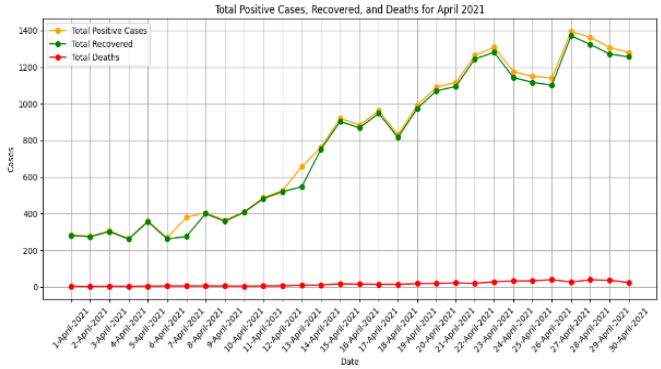
Graph No – 1 (Jan 2021)



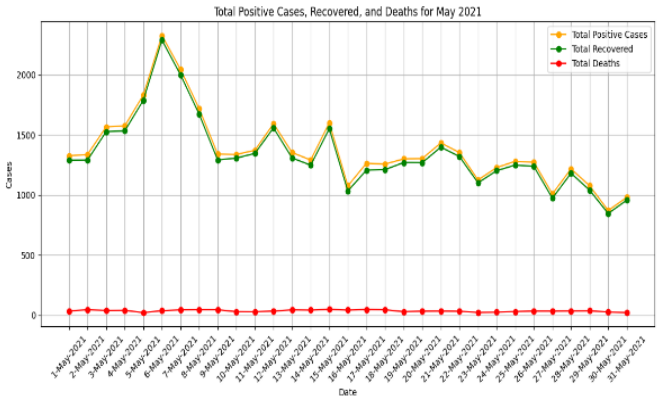
Graph No – 2 (feb 2021)



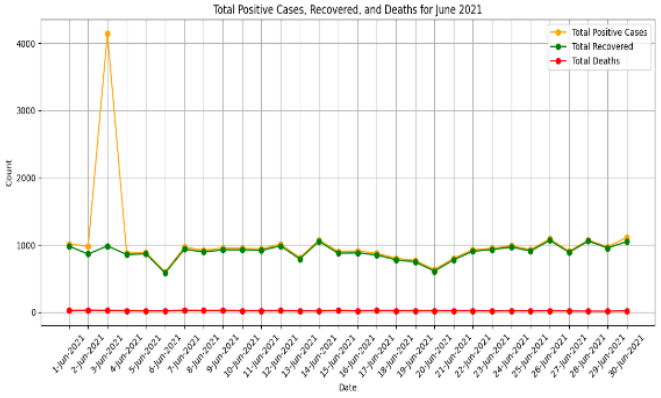
Graph No – 3 (March-2021)



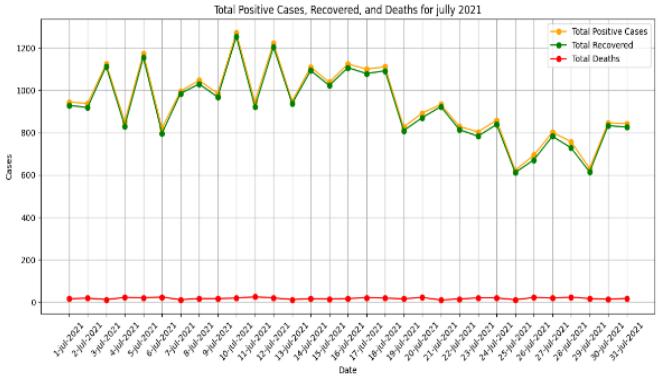
Graph No – 4 (April – 2021)



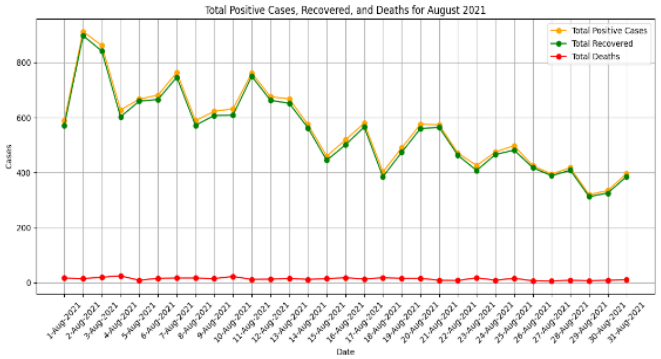
Graph No – 5 (May – 2021)



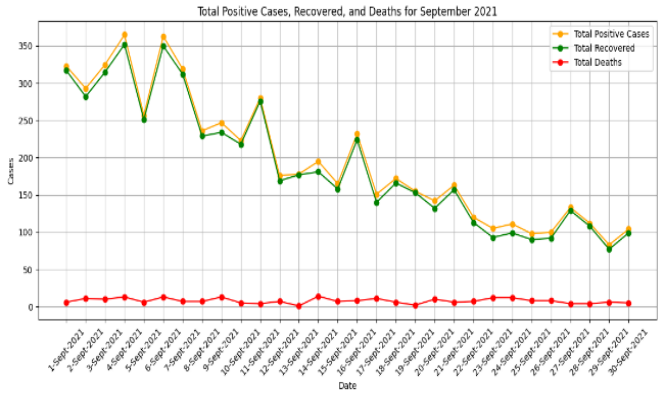
Graph No – 6 (Jun – 2021)



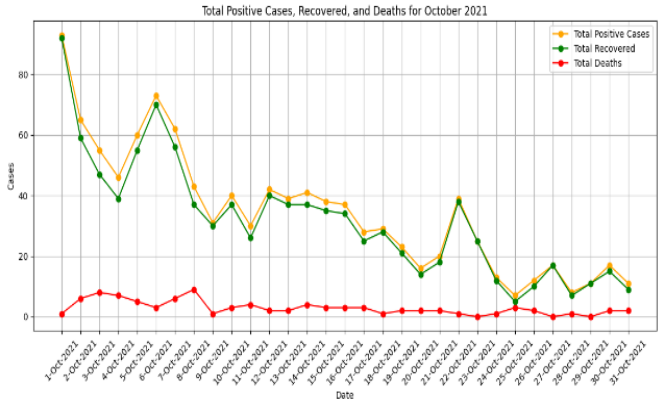
Graph No – 7 (Jully – 2021)



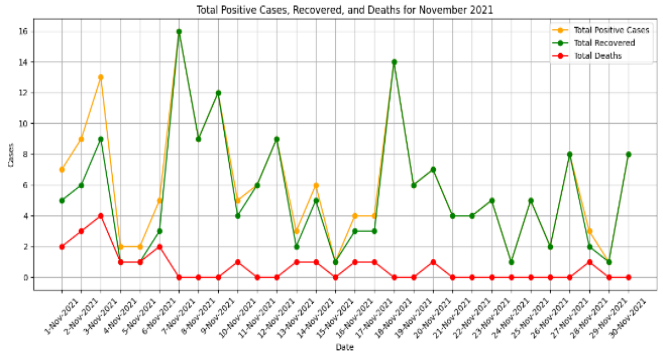
Graph No – 8 (Aug – 2021)



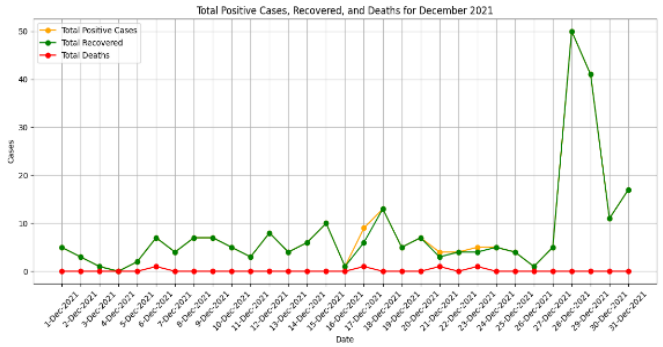
Graph No – 9 (Sept – 2021)



Graph No – 10 (Oct – 2021)



Graph No – 11 (Nov – 2021)



Graph No – 12 (Dec – 2021)

All the above graphs shows on X-Axis yellow colour represent the total active cases , green colur represent total recovered casese out of all active cases and red colur represent total death cases.

CONCLUSION AND FUTURE SCOPE

1) This project work concludes that the knowing and understanding of several factors which can help to asses death cases of the Covid-19 (decrease or increase the) which will further help the Government to take necessary decision and implement useful policies, to implement a proper way to overcome the disaster.
2) The COVID-19 pandemic has posed a significant challenge for nations across the globe, prompting extensive efforts to manage the transmission of the virus and mitigate its impact on public health and the economy.
3) This project involved a detailed analysis of COVID-19 data, which was utilized to create machine-learning algorithms aimed at predicting future cases of the coronavirus.
4) The work undertaken in this project consisted of a Systematic Literature Review to determine the most effective Machine Learning algorithm for the prediction of Covid-19 cases.
5) After analyzing journal papers, I have concluded that regression analysis methods are the most fitting and superior algorithms for the prediction of COVID-19 cases
6) Following an extensive review and analysis of existing literature on prediction methodologies, Linear Regression algorithms were selected for the experiment. These algorithms were subsequently trained and evaluated using a dataset comprising confirmed cases, recoveries, and fatalities.
7) The analysis demonstrates that machine learning algorithms can effectively contribute to the prediction of Covid-19 cases, thereby aiding legislators in making well-founded decisions

concerning public health strategies and the allocation of resources.

8) Future research could investigate various avenues to enhance the findings of this thesis and to further refine the precision of COVID-19 predictions through the application of machine learning algorithms.

REFERENCES

[1]. World Health Organization (WHO), "Coronavirus disease (COVID-19) pandemic," 2020,https://www.euro.who.int/en/health topics/health-emergencies/coronavirus-covid19/novel-coronavirus-2019-ncov.View at: Google Scholar.

[2]. World Health Organization (WHO), "WHO coronavirus disease (COVID-19) dashboard," 2020, https://covid19.who.int/.View at: Google Scholar.

[3]. L. Wang, J. Li, S. Guo et al., "Real-time estimation and prediction of mortality caused by COVID-19 with patient information based algorithm," The Science of the Total Environment, vol. 727, Article ID 138394, 2020.View at: Publisher Site | Google Scholar.

[4]. M. Ozaslan, M. Safdar, I. H. Kilic, and R. A. Khailan, "Practical measures to prevent COVID-19: a mini-review," Journal of Biological Sciences, vol. 20, no. 2, 2020.View at: Publisher Site | Google Scholar.

[5]. M. U. G. Kraemer, C.-H. Yang, B. Gutierrez et al., "The effect of human mobility and control measures on the COVID-19 epidemic in China," Science, vol. 368, no. 6490, pp. 493–497, 2020.View at: Publisher Site | Google Scholar.

[6]. W. Preiser, G. van Zyl, and A. Dramowski, "COVID-19: getting ahead of the epidemic curve by early implementation of social distancing," South African Medical Journal Suid Afrikaanse Tydskrif Vir Geneeskunde, vol. 110, no. 4, p. 12876, 2020.View at: Publisher Site | Google Scholar.

[7]. M. Klompas, "Coronavirus disease 2019 (COVID-19): protecting hospitals from the invisible," Annals of Internal Medicine, vol. 172, no. 9, pp. 619-620, 2020.View at: Publisher Site | Google Scholar.

[8]. J. Pang, M. X. Wang, I. Y. H. Ang et al., "Potential rapid diagnostics, vaccine and therapeutics for 2019 novel coronavirus (2019-nCoV): a systematic review," Journal of Clinical Medicine, vol. 9, no. 3, 2020.View at: Publisher Site | Google Scholar.

[9]. N. Hasan, "A methodological approach for predicting COVID-19 epidemic using EEMDANN hybrid model," Internet of Things, vol. 11, 2020.View at: Publisher Site | Google Scholar.

[10]. Z. Car, S. Baressi Šegota, N. Anđelić, I. Lorencin, and V. Mrzljak, "Modeling the spread of COVID-19 infection using a multilayer perceptron," Computational and Mathematical Methods in Medicine, vol. 2020, Article ID 5714714, 10 pages, 2020.View at: Publisher Site | Google Scholar.

[11]. N. Feroze, "Forecasting the patterns of COVID-19 and causal impacts of lockdown in top five affected countries using Bayesian structural time series models," Chaos, Solitons and Fractals, vol. 140, 2020.View at: Publisher Site | Google Scholar.

[12]. V. Papastefanopoulos, P. Linardatos, and S. Kotsiantis, "COVID-19: a comparison of time series methods to forecast percentage of active cases per population," Applied Sciences (Switzerland), vol. 10, no. 11, 2020.View at: Publisher Site | Google Scholar.

[13]. R. Salgotra, M. Gandomi, and A. H. Gandomi, "Evolutionary modelling of the COVID-19 pandemic in fifteen most affected countries," Chaos, Solitons, and Fractals, vol. 140, Article

ID 110118, 2020.View at: Publisher Site | Google Scholar.

[14]. F. Shahid, A. Zameer, and M. Muneeb, "Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM," Chaos, Solitons, and Fractals, vol. 140, Article ID 110212, 2020.View at: Publisher Site | Google Scholar.

[15]. S. Singh, K. S. Parmar, J. Kumar, and S. J. S. Makkhan, "Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties cases of COVID-19," Chaos, Solitons, and Fractals, vol. 135, Article ID 109866, 2020.View at: Publisher Site | Google Scholar.