iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** III **Month of publication:** March 2022

DOI: https://doi.org/10.22214/ijraset.2022.40990

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

Covid 19 Data Analysis in India Using Linear and Polynomial Regression Algorithms

Ekta Rahangdale[1], Sujata More[2], Shipali Narnaware[3], Gayatri Sahu[4], Sayali Gujar[5]
[2]*Assistant Professor at department of Information Technology, Nagpur Institute of Technology, Nagpur, Rashtrasant Tukadoji Maharaj Nagpur University*
[1, 3, 4, 5]*UG Students at Department of Information Technology, Nagpur Institute of Technology, Nagpur, Rashtrasant Tukadoji Maharaj Nagpur University*

Abstract: ***In past two years there is a pandemic called covid-19, which has shook the world. The world has suffered a lot and suffering till now by the disastrous effect of corona virus globally. It has affect the world in all parameter i.e. economically, mentally and so on. The world don't when will this pandemic end yet can make forecast by utilizing AI calculations to make moves in the event that this occurs in later days ,how might human and government make counteractions from Covid. This project "Analysis on covid-19 in India using Linear and Polynomial algorithms" analyze the covid-19 datasets from 01-03-2021 to 08-05-2021 for India and also for its top 4 states ,having more number of confirmed cases and predicted the results by using machine learning algorithms (linear regression and polynomial regression with degree of 5).The predicted results will be helpful for government to take actions against this pandemic.***

Keywords: Data Analysis, Linear Regression, Polynomial Regression, Preprocessing of data, Flask

I. INTRODUCTION

As we all know about corona outbreak in the world from last two years .We all suffering from this in somehow differ situations. According to the reports ,there are significant differences across the states ,countries in terms of test availability , hospital maintenance , beds and many more reasons .India has been suffering from two waves of covid-19 and there are so many conflicts over beds availability , death rate through delay in treatment .Our project will describes the analysis on Covid-19 related datasets , through which government can analysis the accurate death rates , where the impact of corona was huge .So , in future if these kind of pandemic will occur, then the analysis could help them about the requirements of the treatments . Although, several studies in the context of India have been reported recently by many researchers to understand and analyze the dynamics of COVID-19 spread, but there are very limited studies on state wise analysis of the outbreak. Taking a gander at the variety in populace, populace thickness and geological circumstances, the investigation of India overall may not give genuine status of the scourge, in this manner, each conditions of India which has huge populations as compared to the other part of world, need to analyze separately for the spread of corona virus. Measurable models are significant instruments to investigate the constant information examination of irresistible infection. In this project, we have utilized the linear and polynomial regression model to analyze the pandemic data of India and its different states. It is vital to make reference to that the expectation made in this study is basically as great as the nature of information accessible and deviation from the patterns before very long may change the forecasts also.

II. LITERATURE REVIEW

1) P.Jain, N.Darapaneni R.Khattar, M.Chawla, R.Vaish and A.R.Paduri," Analysis of data and Prediction of COVID-19 Pandemic in India" in 29-Aug 2020.In this paper they have examined the COVID-19 in India and their three most follow up on Indian states (for example Maharashtra, Tamil Nadu and Andhra Pradesh).They have used time series information for India and applied the SIR model and the FbProphet model.
2) Ramjeet, Yadav, Singhdoi," Data examination of COVID-2019 pestilence utilizing AI strategies: a contextual analysis of India", in 26 May. 2020.This paper introduced, the data set of COVID-19 has been examined from March 1, 2020, to April 11, 2020.In this review, they have been used 6th degree, and remarkable polynomial individually for the COVID-2019 dataset
3) Apurbalal, Senapati," An original system for COVID-19 case forecast through piecewise relapse in India.", 1-8. 10 Nov. 2020.In this paper concentrate on the AI (ML) directed straight relapse model has been utilized to address the various kinds of COVID-19 related issues. The straight relapse model has been fitted into the dataset to manage the all out number of positive cases.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

III. PROBLEM STATEMENT

Covid is rising universally and pandemic influence the entire world. There will be no closure appears while crown going to be end .Corona shook the world with respect to various boundaries for example actually, intellectually, monetarily, etc. It is challenging for the public authority likewise to keep up with their country in this difficult stretch.Thus, assuming there is an expectation and nitty-gritty investigation of the pandemic, it will accommodating for the public authority to make a prompt move on the off chance that these sort of pandemic occurs in future.There are a few examinations with regards to India have been accounted for as of late by numerous analysts to comprehend and investigate the elements of COVID-19 spread, however there are exceptionally restricted investigations on state astute examination of the flare-up.On the off chance that,there will be state astute information investigation and expectation additionally ,so it will accommodating for the state government likewise to keep up with the circumstances. Also, there algorithm examination between , which calculation is giving the highest precision for forecast.

IV. METHODOLOGY

A. Collection of Datasets

The collection datasets consists of:

1) Datasets of Indian covid-19 data,which was collected by Indian government official website (https://www.covid19india.org/).
2) Data collection of Indian state-wise data,which is also collected by Indian government official website (https://www.covid19india.org/).

B. Preprocessing of the given datasets

The process of data pre-processing consist of the following steps:

1) *Data reduction*: Reduce data dimensions as given data is huge and machine learning models will predict well there will be right dimensions of data.
2) *Cleaning of the data*: It Enhance data quality and reduce the unnecessary data.
3) *Transform the datasets according to the need*: Ensure data compatibility with algorithms analysis and prediction. As date is not compatible for visualization, so we import date-time to change string data-type into date object.
4) *Partitioning of the datasets*: For prediction we split the datasets into testing and training datasets , so that model can predict according to given datasets.

C. Visualization of data

The proposed project used the different libraries for visualizing different parameters i.e. regression models, predictions, plotting the state and Indian covid-19 data .There are different libraries to beautifully visualize the given data in to graphs, charts. Some of the libraries, which our project has used are:

1) *Pandas***:** This project used panda library for manipulating the given datasets.
2) *numpy :* This project numpy library for dealing with the numerical data values.
3) *Matplotly***:** For visualizing projects datasets in a significant way ,matplotly library has been used.
4) *Plotly:* It was used in our given project for visualizing the state-wise datasets, time-series datasets of India and the for showing the model prediction and regression models visualization.

D. Data Analysis

In given project the different states of India with highest number of confirmed cases (top 4) has been analyzed and doing the analysis different libraries for beautiful and understandable representation.

E. Applying Machine Learning Algorithms

Proposed given project has used two of the machine learning algorithms i.e. linear regression and polynomial regression algorithms for analyzing and predicting the future values.

For proposed project ,Linear regression and polynomial regression is used.The algorithms are as follows:

1) *Linear Regression Algorithm:* Linear regression is one of the simplest and most well known AI calculations. A factual strategy is utilized for prescient analysis.Linear relapse calculation shows a straight connection between a reliant (y) and at least one free (y) factors, henceforth called as direct relapse. Since direct relapse shows the straight relationship, and that implies it observes how the worth of the reliant variable is changing as indicated by the worth of the autonomous variable.

Types of Linear Regression:
Linear regression can be additionally separated into two sorts of the calculation:

a) Simple-Linear-Regression
b) Multiple-Linear-Regression

In proposed project, linear regression model prediction and analysis is used for the Covid-19 dataset for India from 01 March to 08 May (68 days) and the same for its states Maharashtra, Karnataka, Kerala, Uttar Pradesh(having highest number of confirmed cases) and after that we have calculated the r2_score and mean-squared-error values for the Indian dataset and its states.
The calculation of the r2 value is given by:
R2=(1−SSResidual/SSTotal)

2) *Polynomial Regression:* Polynomial Regression is a relapse calculation that models the connection between a dependent(y) and autonomous variable(x) as furthest limit polynomial. The Polynomial Regression condition is given beneath:

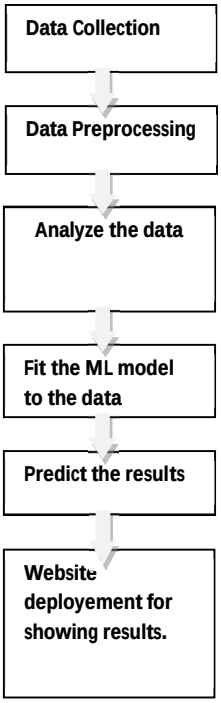
y= b0+b1x1+ b2x12+ b2x13+..... bnx1n

It is a straight model with a change to expand the precision. The dataset utilized in Polynomial relapse for preparing is of non-straight nature. It utilizes a straight relapse model to fit the convoluted and non-direct capacities and datasets.On the off chance that we apply a straight model on a direct dataset, it gives us a decent outcome as we have found in Simple Linear Regression, yet assuming we apply a similar model with next to no alteration on a non-direct dataset, then, at that point, it will create an intense result. Because of the blunder rate will be high, and precision will be decreased.So for such cases, where information focuses are organized in a non-direct style, we really want the Polynomial Regression model.In proposed project, linear regression model prediction and analysis is used for the Covid-19 dataset for India from 01 March to 08 May (68 days) and same for its states Maharashtra, Karnataka, Kerala, Uttar Pradesh(having highest number of confirmed cases) and after that we have calculated the r2_score and mean-squared-error values for the Indian dataset and its states.

F. Web Deployment

Flask gives the developer varieties of choice when developing web applications. Our project has deployed in website using flask, which is showing the comparison between machine learning algorithms and showing the accuracy rate and mse value.
It accepts the input as state name and shows the results as in the form of table format and the accuracy result using ML algorithms

V. FLOWCHART



VI. RESULT

The result of our project analysis on covid-19 in India and its states, we have proposed the linear regression and polynomial regression based machine learning approach for the prediction of actual positive cases and recovery cases of four different states in India, which has highest number of confirmed cases from march 01-2021 to may 08-2021. The main novelty of the proposed scheme is that we have applied linear regression method and polynomial regression. As a result, the proposed model produces an r2_score and mean-square-error predicted result .Hence we have compare between the two machine learning algorithms and found that the polynomial regression model's r2_score is highest than the linear regression model and the value of mean square error is less than the value of linear regression mean square value.

A. Linear Regression Model Prediction:

Country/State	R2_score	MSE
India	-10.90	438864866.47
Maharashtra	-10.06	179074848594.39
Karnataka	-488.91	176528035074.69
Uttar Pradesh	-308.69	159299855079.90
Kerala	-536.95	131623343243242.37

B. Polynomial Regression Model Prediction:

Country/State	R2_score	MSE
India	0.47	21790445975.54
Maharashtra	0.71	5559129556.16
Karnataka	0.79	7960674955.19
Uttar Pradesh	0.53	8375613055.31
Kerala	0.47	21790445975.54

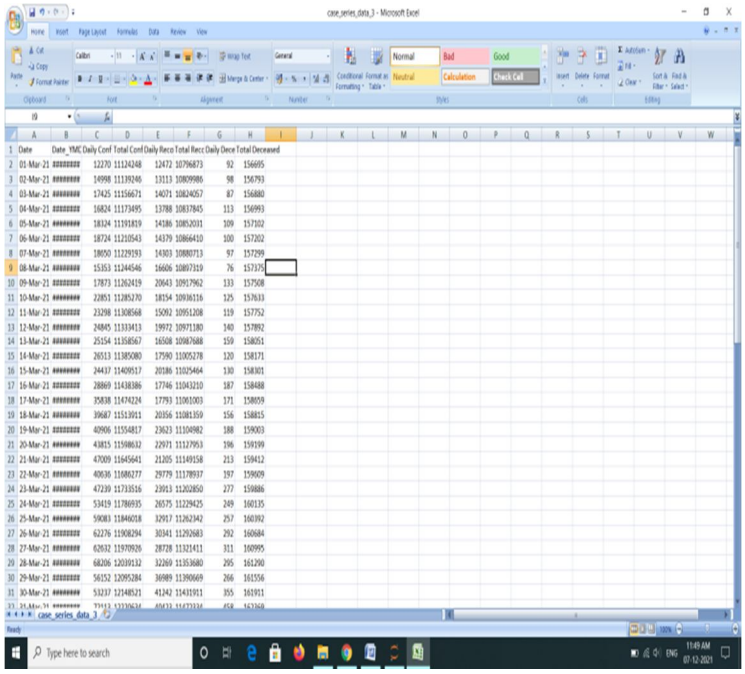


Fig: Dataset of Indian data from 01-03-2021 to 08-05-2021

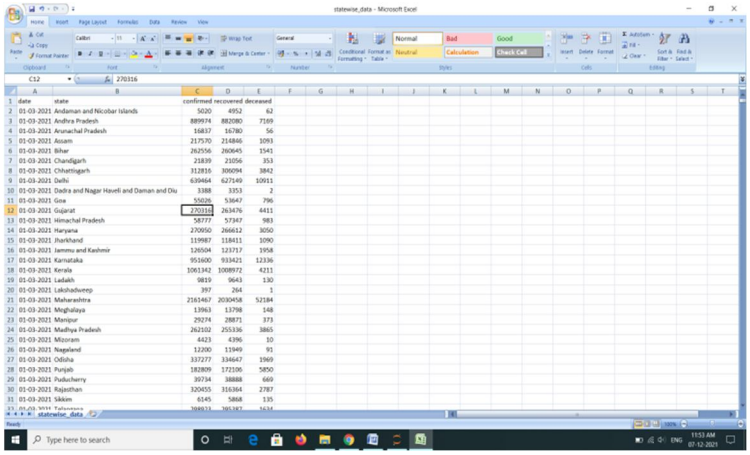


Fig: Dataset of Indian States data from 01-03-2021 to 08-05-2021

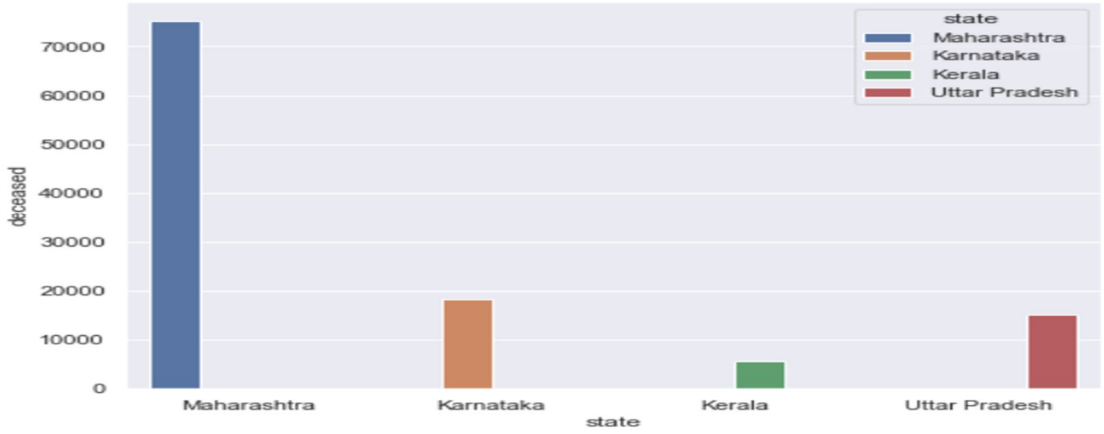


Fig: Top 4 states ,having highest number of confirmed cases from 01-03-2021 to 08-05-2021

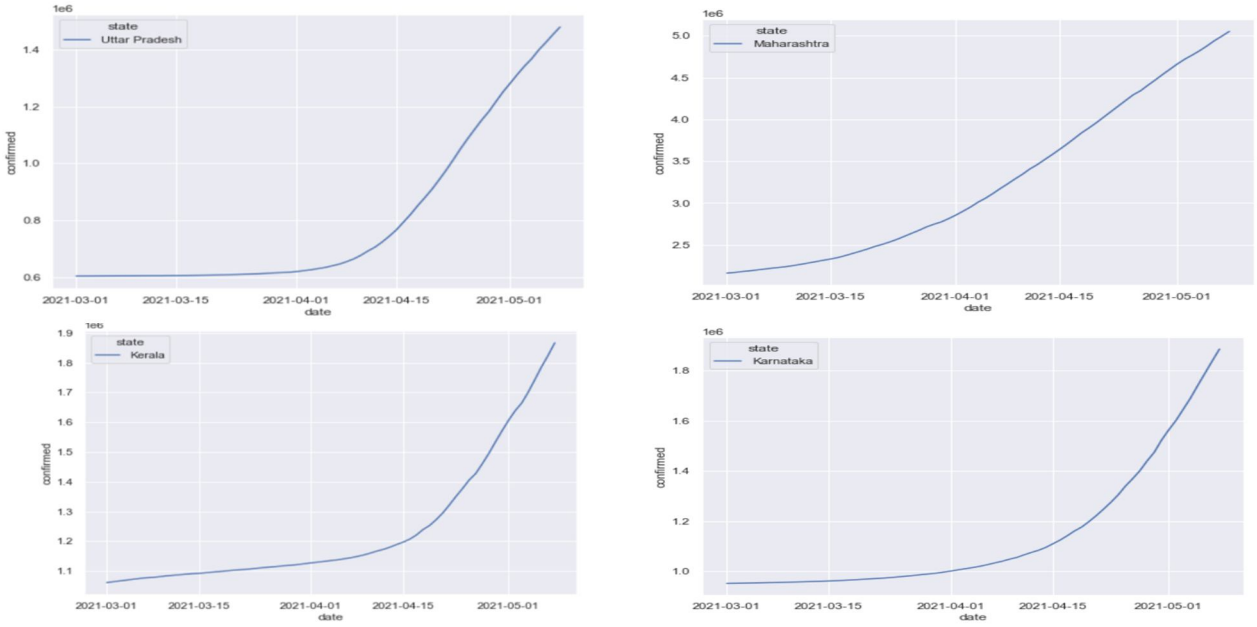
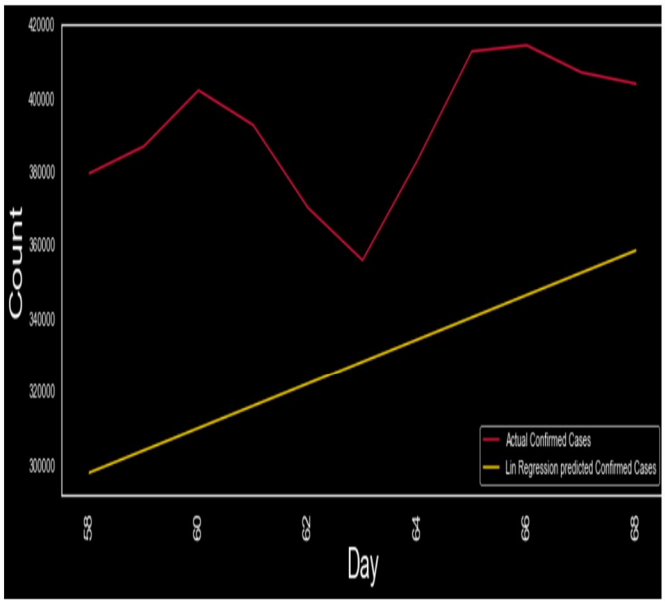
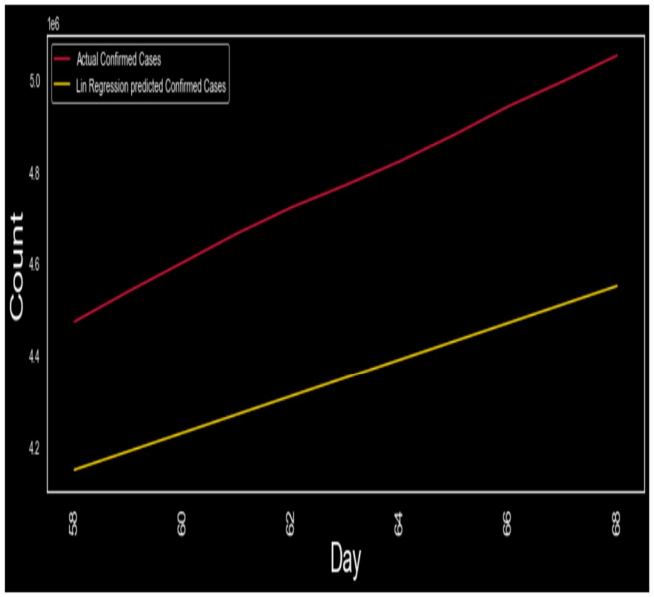


Fig: Confirmed cases in Indian states(Maharashtra, Karnataka, Kerala, Uttar Pradesh) from 01-03-2021 to 08-05-2021

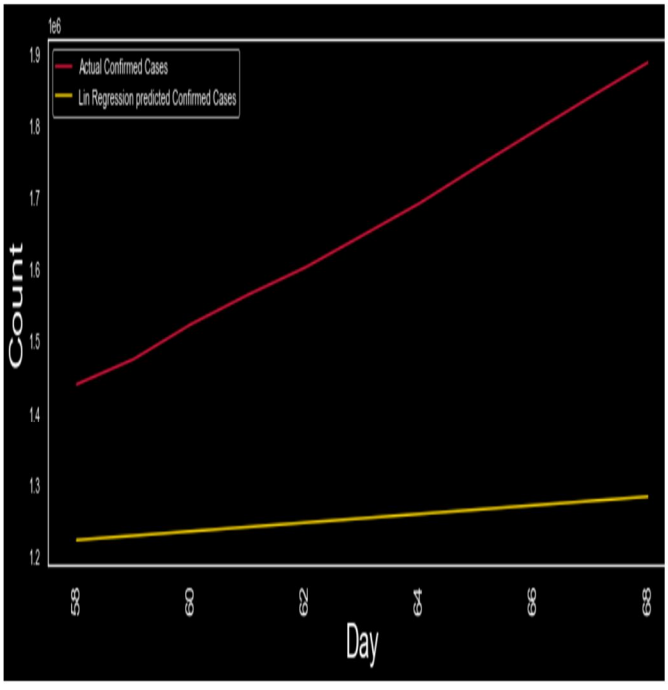
-10.900320006386755
4388649866.47048



-10.067175494747561
179074848594.39645



-486.9170206806863
176528035074.69855



-308.6937668604833
159299855079.90497

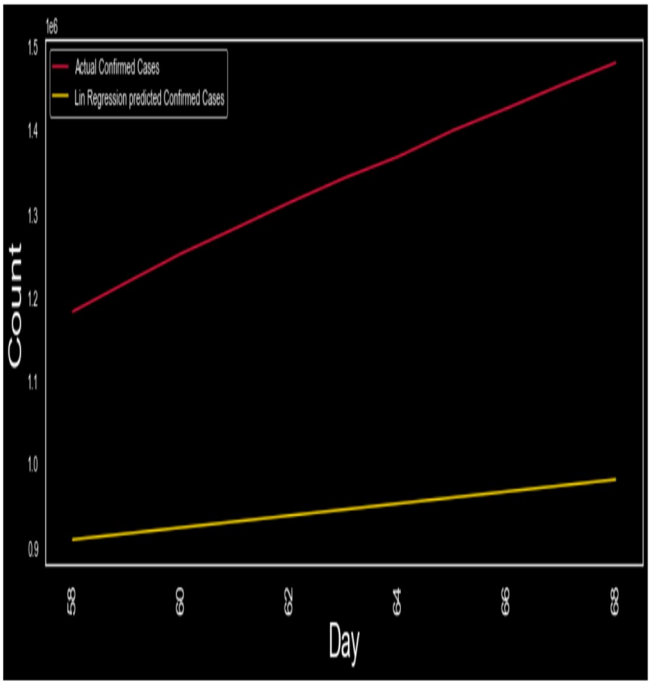


Fig: Regression graph using Linear regression model in Indian states(Maharashtra, Karnataka, Kerala, Uttar Pradesh) from 01-03-2021 to 08-05-2021 having r2_score value and mse value

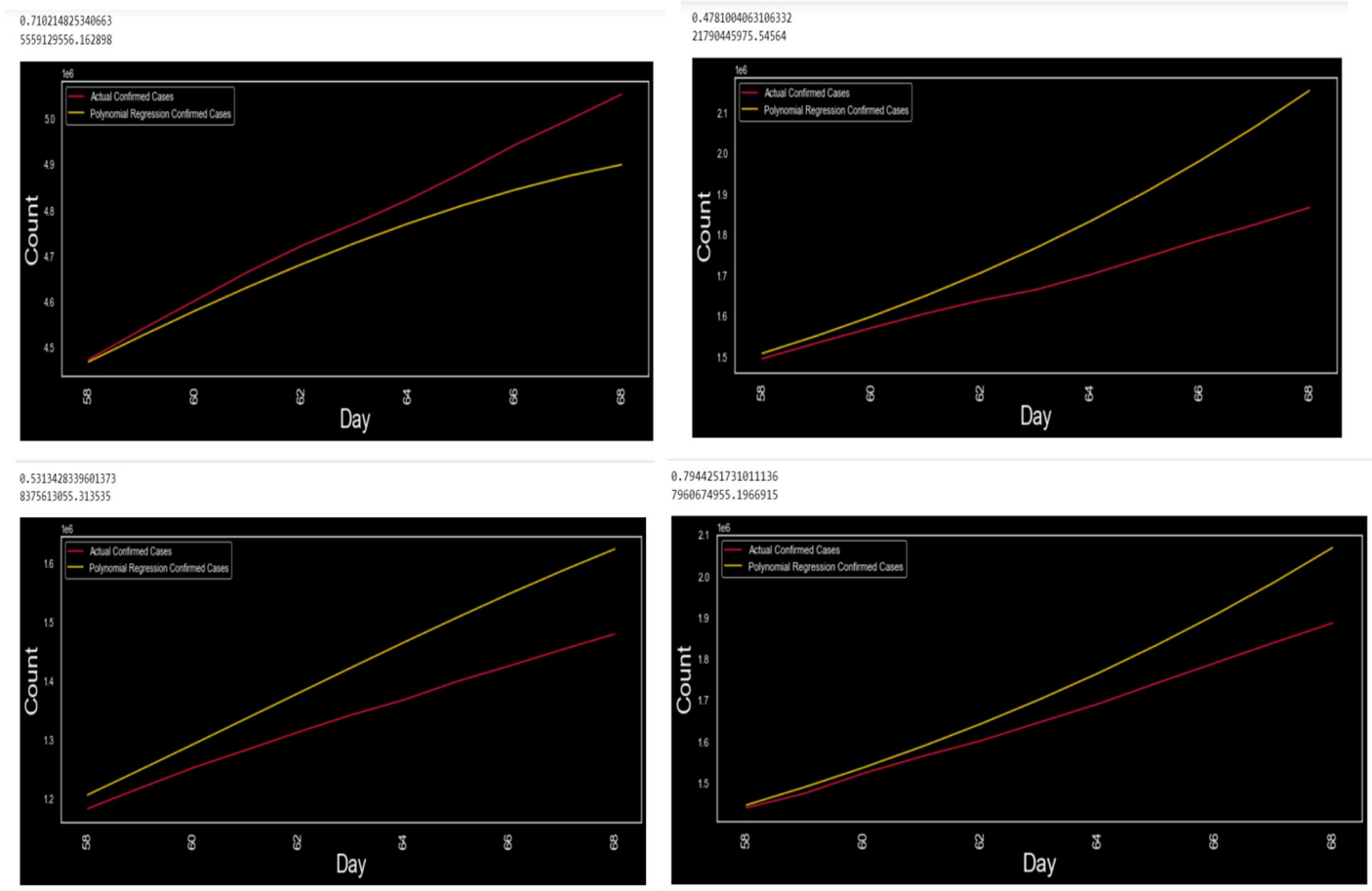
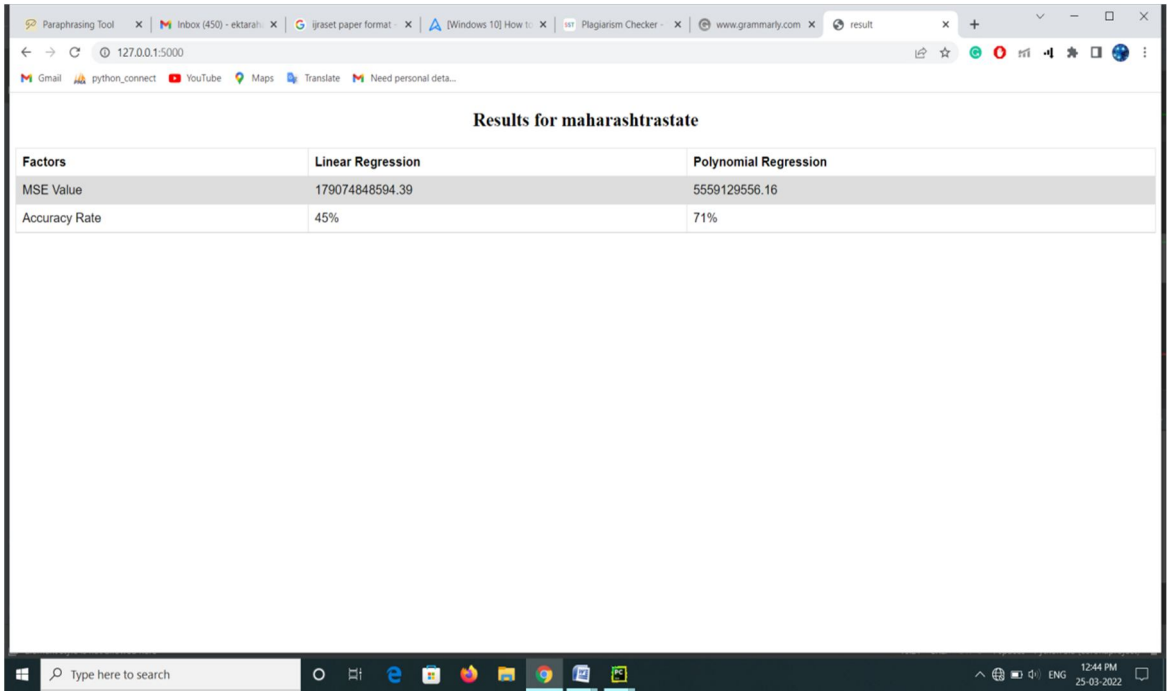


Fig: Regression graph using Polynomial regression model in Indian states(Maharashtra, Karnataka, Kerala, Uttar Pradesh) from 01-03-2021 to 08-05-2021 having r2_score value and mse value



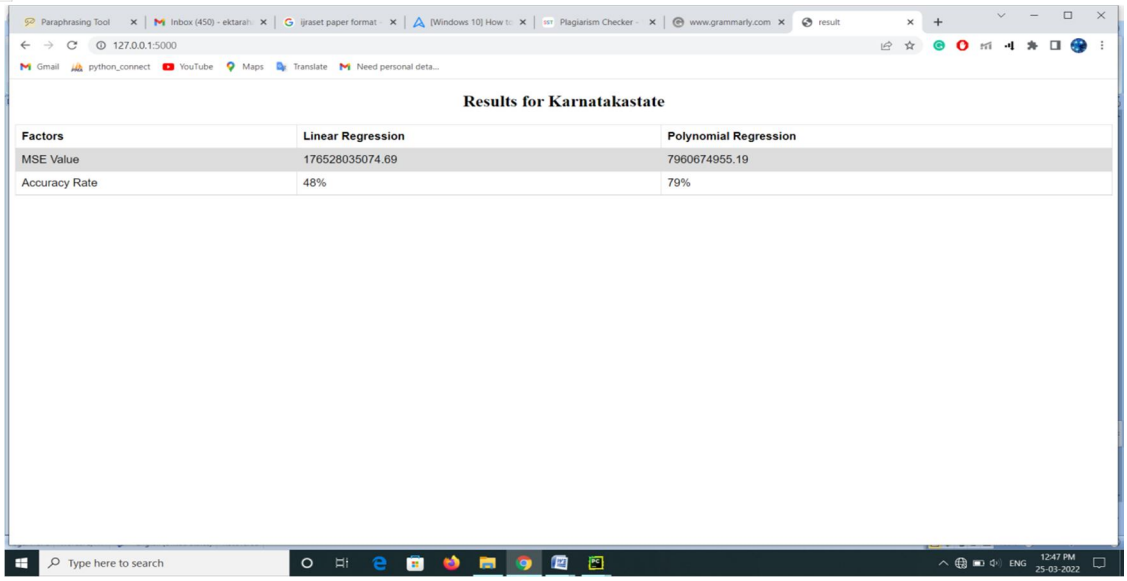


Fig: Web deployment of the project

VII. CONCLUSION

The Coronavirus pandemic is a worldwide pandemic. Understanding the spread of Coronavirus as per which region has biggest number of cases can be useful for the public authority for future.We have observed that the express that has most noteworthy thickness has the largest number of affirmed. As in metropolitan regions, the populace thickness is extremely high, and social separating is trying to keep up with; the job of government is urgent in battling the pandemic. By guaranteeing the wellbeing and cleanliness related offices, (giving sufficient clean water, satisfactory disinfection, and sewerage offices, cleaning the city, keeping up with isolation focuses and general medical services foundations, and so on), and further developing public circulation framework to guarantee least food supply, particularly among the metropolitan poor and other denied sub-gatherings, can assist with controlling the spread of Coronavirus infection.We have additionally separate between two AI calculations linear and polynomial and applying the calculations to the datasets and observed that polynomial relapse give the preferable outcome over linear.Our examination has a couple of impediments. To begin with, there is plausible of under-detailing positive and deadly cases because of an absence of testing or social shame. Subsequently our information gives the most safe approximations of the contamination proportion. Second, for most cases, the patients' degree of data (like age, sex, and comorbidity) is inaccessible. In this manner, we examined the area level determinants rather than individual-level determinants. Along these lines, our outcomes recognized the significant associates just at the area level. At last, we examined the quantity of affirmed cases for contamination proportion as opposed to the quantity of dynamic cases. The later considers the recuperation rate and relies upon the wellbeing administration accessible in an area. We involved the quantity of affirmed cases as the essential sign of the spread of the contamination. Regardless of these restrictions, the review's legitimacy lies in uniting spatial-segment weaknesses pervasive the country over during the pandemic time frame.

REFERENCES

[1] N. Darapaneni, P. Jain, R. Khattar, M. Chawla, R. Vaish and A. R. Paduri, "Examination and Forecast of Coronavirus Pandemic in India," 2020 second Global Gathering on Advances in Processing, Correspondence Control and Systems administration (ICACCCN), 2020, pp. 291-296, doi: 10.1109/ICACCCN51052.2020.9362817.

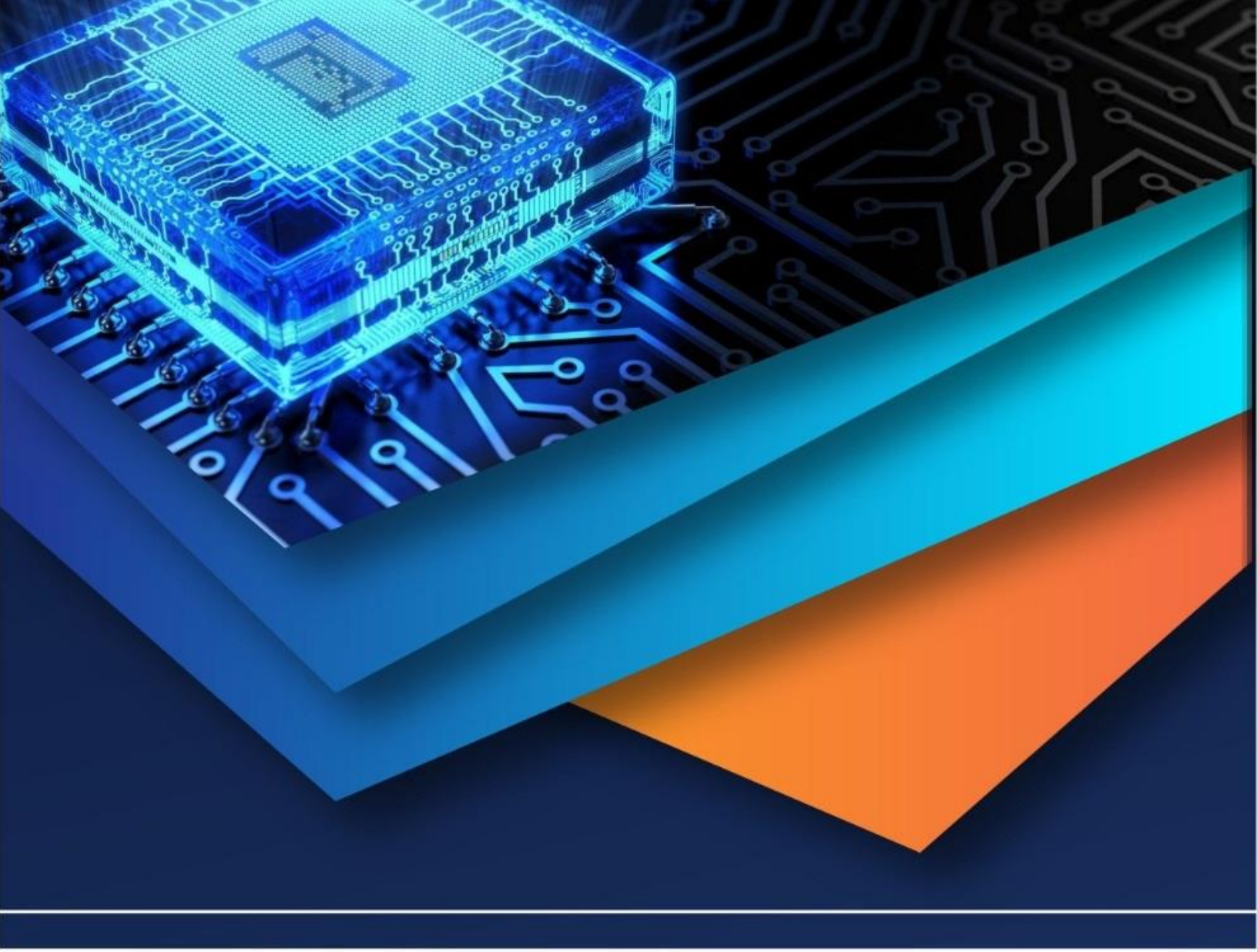
[2] Senapati, Apurbalal et al. "A unique framework for Covid case figure through piecewise backslide in India." Worldwide journal of information development : a power journal of Bharati Vidyapeeth's Underpinning of PC Applications and The leaders, 1-8. 10 Nov. 2020, doi:10.1007/s41870-020-00552-

[3] Z. Liu, J. Zuo, R. Lv, S. Liu and W. Wang, "Covid Pandemic (Coronavirus) Forecast and Pattern Investigation In light of Time Series," 2021 IEEE Worldwide Gathering on Man-made brainpower and Modern Plan (AIID), 2021, pp. 35-38, doi: 10.1109/AIID51893.2021.9456463

[4] Yadav, Ramjeet Singh. "Data assessment of Covid 2019 plague using PC based knowledge systems: a setting centered appraisal of India." Generally speaking journal of information movement : a power journal of Bharati Vidyapeeth's Supporting of PC Applications and The board, 1-10. 26 May. 2020, doi:10.1007/s41870-020-00484

[5] Shordia and Y. Pawar, "Breaking down and Guaging Coronavirus Flare-up in India," 2021 eleventh Worldwide Gathering on Distributed computing, Information Science and Designing (Intersection), 2021, pp. 1059-1066, doi: 10.1109/Confluence51648.2021.9377115

[6] R. Gupta, et. al., SEIR and Relapse Model based Coronavirus episode expectations in India, medRxiv 2020: https://doi.org/10.1101/2020.04.01.20049825



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY