

Automated Data Analysis and Prediction of Covid-19 Pandemic Using Machine Learning.

¹Mrs.Thangamma K C, ²Bhavana N, ³Gnaneshwari D S, ⁴Syeda Sukaina Fatima, ⁵Yashaswini S

¹Assistant Professor, ²Student, ³Student, ⁴Student, ⁵Student

¹²³⁴⁵Information Science and Engineering,

¹²³⁴⁵GSSS Institute of Engineering and Technology for Women, Mysuru, Karnataka, India

Abstract: The coronavirus disease originated from China is spread across worldwide. As per the recent count India has the second highest number of cases in the world. Machine learning tools and techniques are widely used to curb the disease. The paper firstly gives an overview of methodologies that can be adopted and the system proposed to design the framework. Our project aims in develop a robust automated model for exploratory data spread analysis and forecasting for COVID-19.

Index Terms - Artificial intelligence (AI), Machine learning (ML), COVID-19, Coronavirus.

I. INTRODUCTION

The Covid-19 pandemic has affected our lives through various factors causing deaths worldwide. In Wuhan of China had reported the first case since then the number of cases kept increasing and has now spread to most of the countries affecting lives of the people and thus reducing the population world-wide. This pandemic has led to a huge economic and social devastation. There are many people who are in danger of falling into extreme poverty. Large number of enterprises have been facing existence threat. Almost half of world's 3.3 billion workforce is at risk. There are about 219 countries and territories that have recorded COVID-19 cases. The necessity is to develop right and effective vaccine against COVID-19 virus. The AI, ML and many other technologies have their major role in countering to the novel coronavirus disease. The Artificial Intelligence and Machine learning techniques have contributed to rapid and vigorous research that has provided the capacity to understand virus and its transmission through various applications. Government's efforts in curbing the pandemic such as lockdowns, curfews and policies are helping people to cope with the COVID-19 situations. This paper provides a knowledge of frameworks that can be developed using ML techniques to examine the increased spread analysis of infectious disease so that we can predict the future pattern of the disease. We discuss about the possible techniques and provide a survey of most appropriate ML technique that can be used for prediction of the infectious disease.

II. LITERATURE REVIEW

The newborn coronavirus disease has affected all the parts of the world and has become a pandemic that nobody is unaware of even after a long time it is being difficult to tackle with. The technology of today's world has various things to offer. Many developers around the globe have mentioned various ways to tackle with the pandemic technically.

[1] "Machine learning prediction for covid-19 pandemic in India" The base paper has it that the future dynamics of the pandemic can be predicted by using different algorithms. The project uses mainly three algorithms-(1) The ANN algorithm (2) Linear regression and (3) Support vector machine regression. By predicting the covid -19 it helps the government to take advance measures. Instead of using one algorithm, the three algorithms combined give a reduced error and improves the accuracy. We have a similar technique but we use (1) ARIMA time series (2) Support vector machine and (3) Linear regression. We will also be giving the government insights as to what might happen in the future along with an analysis. We will also be recommending the best algorithm out of the three which has the least error using the root mean square method.[2] " Artificial intelligence(AI) and big data for coronavirus(COVID-19) pandemic : A survey on the state-of-the-arts ",In this paper, a survey is presented on the state-of-the-art solutions in the battle against the COVID-19 pandemic. The techniques used here are big data for spatial analysis methods and Geographic Information Systems (GIS) technology which would help collect the data from clinics, government hospitals etc. The technique used for classifying respiratory viruses is the reverse transcription polymerase chain reaction (RT-PCR) detection technique. We collect the W.H.O dataset and use that for the analysis.[3]" Artificial intelligence in the battle against coronavirus (COVID-19):A survey and future research directions" This paper is about directing the future by surveying the data. The survey paper mentions the deep learning algorithms for image processing. Radiology images such as chest X-ray and CT scans are high-dimensional data that require processing capabilities of deep learning methods in which CNN-based models are common and most suitable.[4] " Analysis of the distribution of COVID-19 in Italy using clustering algorithms " ,This paper analyzes the distribution of covid-19 in Italy using clustering methods. The paper mentions the use of K-means clustering and hierarchical clustering algorithms. This gives government insights on geographical areas where the corona affected patients are more and where they are less.

[5] "Regression analysis of covid-19 using machine learning algorithms", This paper uses polynomial regression algorithm and support vector machine to analyze the situation of the pandemic. It gives a clear picture of the present state of the spread as well as predicts the future spread possibility of the covid-19 pandemic. It concludes to say that the polynomial regression algorithm gives 93% accuracy which is better than the SVM algorithm. The reason for doing state wise spread analysis is also substantiated along with the facts that describe why India's mortality state is higher than other countries inspite of being the second most COVID-19

affected country in the world.[6] “A Novel parametric model for the prediction and analysis of the COVID-19 casualties”, This paper analyses the spread in turkey. It reviews the SIR model and its properties. A new model named the SPID model has been inspired by it and established. LS based parametric optimization approach is used to ascertain the unknown parameters. It concludes that the SPID model can be used which gives the accurate number of covid cases.

III. EXISTING SYSTEM

Existing System might not help to prepare against possible threats and consequences. The existing System predicts the extent of the pandemic till a limited period of time. The system uses different machine learning algorithms that would not provide accurate results. Most of the existing systems have analysis done only during the arise of coronavirus disease.

PROPOSED SYSTEM

The proposed system uses machine learning techniques, which are performed using the python library to predict the covid-19 outbreak in India. Dataset is collected from official websites. After collecting the data, formulative analysis is performed to get the data in date-time format. Then by performing the machine learning algorithms, prediction of covid is done.

System Modules:

- **Exploratory data analysis:** Here first we carry out full exploratory data analysis on the dataset, including a study of covid-19 growth rate. Under pre-processing here, we convert the data into date-time format and group the data under three categories: confirmed, recovered and deaths based on the date. We also add week column to perform weekly analysis further ahead.
- **Predictive analysis:** Here prediction of covid-19 is done using the polynomial regression. Here first we predict confirmed cases for near future using the above machine learning algorithm and plot the graphs.

IV. SYSTEM DESIGN

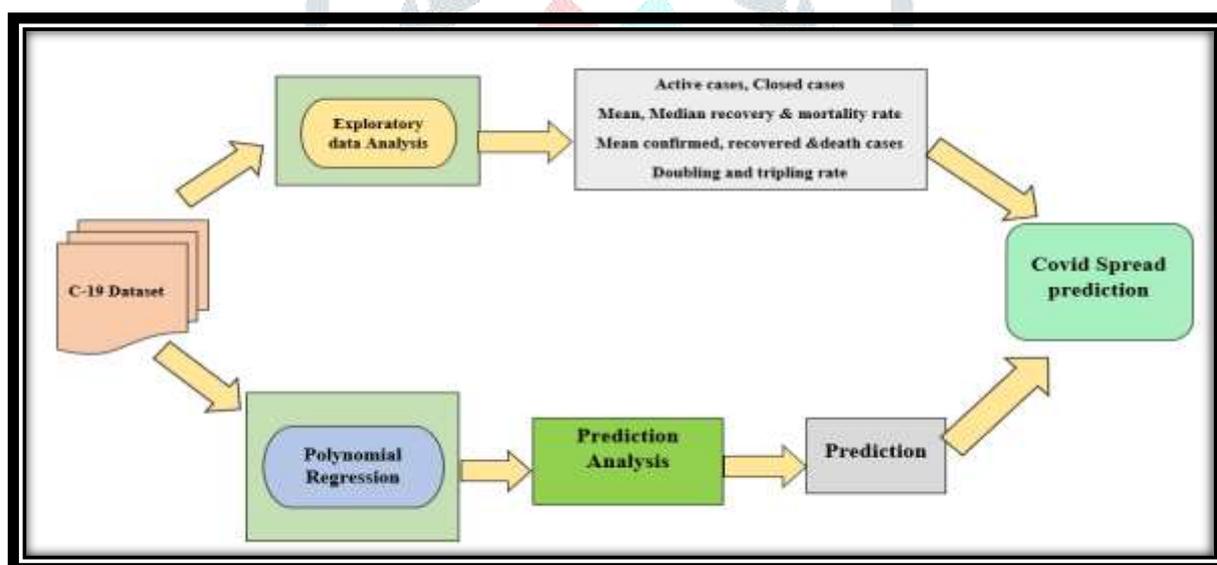


Fig 1: System Architecture of Automated data analysis and prediction of covid-19

The figure 1 depicts the work flow of automated data analysis and prediction of covid-19 pandemic using machine learning techniques. Here we first collect the dataset from different websites then carry out the exploratory data analysis of covid-19 and then use machine learning techniques for predictive analysis. Prediction of the covid-19 pandemic is done using the Polynomial regression algorithm.

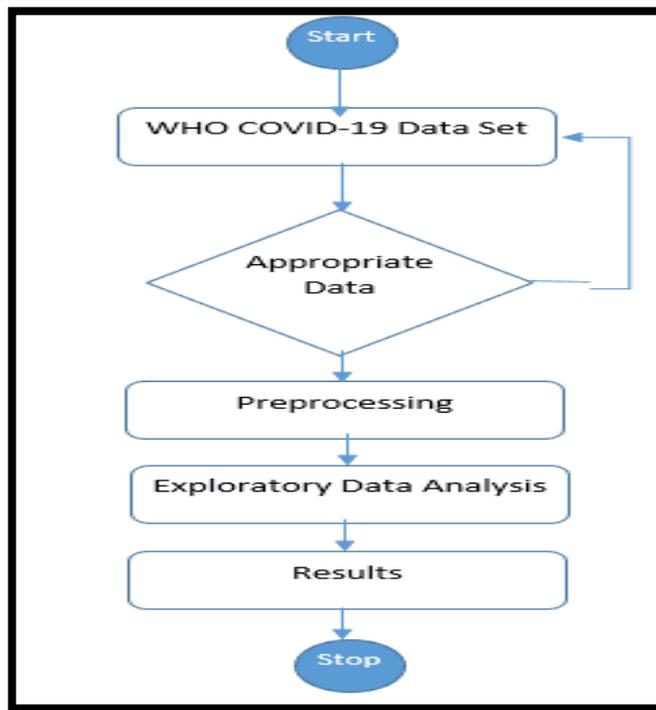


Fig 2: Dataflow diagram for exploratory data analysis.

Figure 2 depicts the dataflow diagram for the exploratory data analysis. The dataset is collected and checked whether appropriate. If the dataset is correct preprocessing is done and results are displayed.

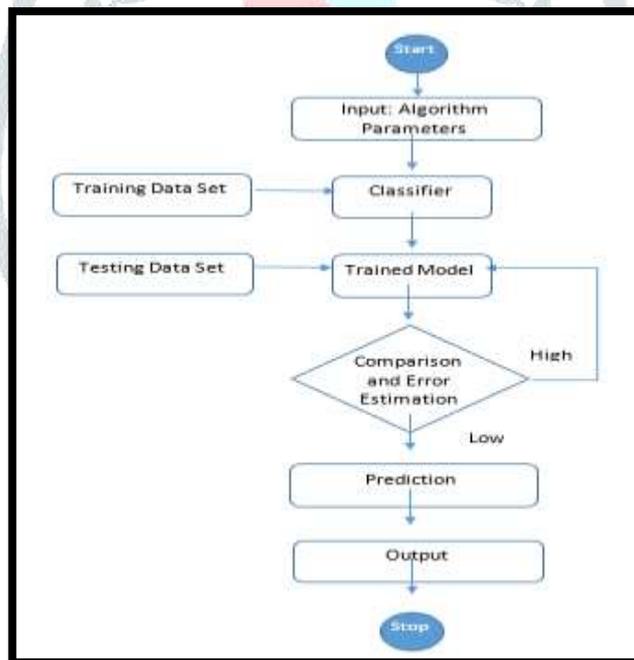


Fig 3: Dataflow diagram for Prediction

Figure 3 depicts the dataflow diagram for prediction. The dataset is given as input to the algorithms and training and testing of dataset is done. The comparison and estimation is performed for error estimation and prediction results are obtained.

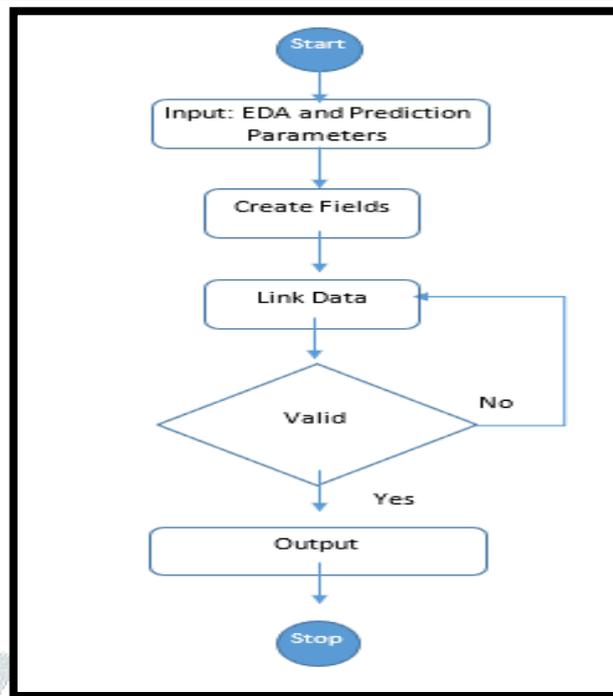


Fig 4: Dataflow diagram for web application

Figure 4.2.3 depicts the dataflow diagram for web application. Here the exploratory data analysis and prediction values are linked to the web application.

V. RESULTS

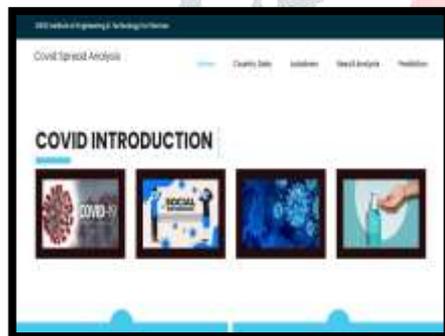


Fig 5: Home page

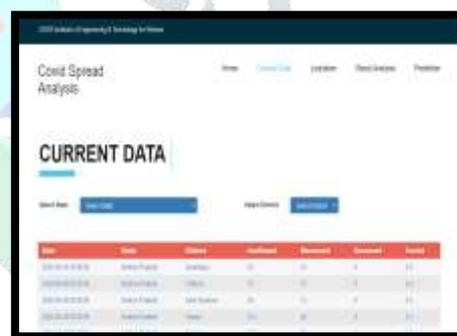


Fig 6: Current Data Of All States

This Figure 5 depicts the home page of the automated data analysis and prediction of covid-19 pandemic and Figure 6 current date of all the states and their districts that can be chosen from dropdown.



Fig 7: Current data of Karnataka

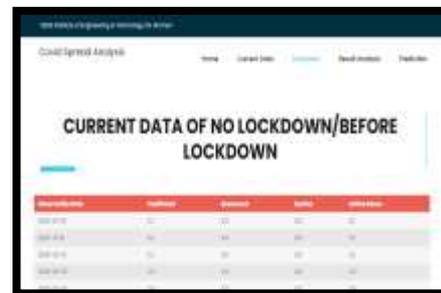


Fig 8: Lockdown/ Before Lockdown Data

This Figure 7 depicts the current data of Karnataka state and Fig 8 current data of covid-19 in no lockdown or before lockdown of India.

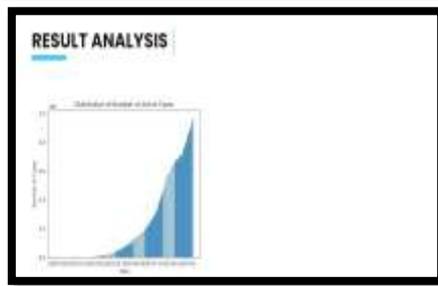


Fig 9: Result analysis

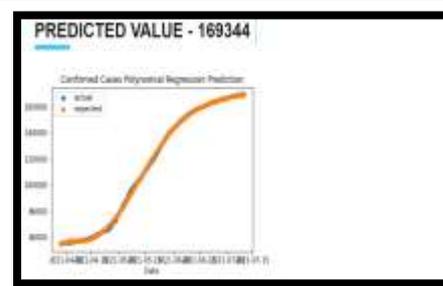


Fig 10: Polynomial regression Prediction

The Figure 9 depicts the result analysis of exploratory data analysis and Fig 10 prediction of covid cases considering the Mysuru district using polynomial regression.

VI. CONCLUSION

Covid-19 Pandemic has severely affected people's life in various ways resulting in drastic loss. There is no particular treatment for coronavirus to date. However, one can avoid the further spread of infection by maintaining social distancing from affected persons and getting vaccinated. It is of utmost importance to identify future infected cases and the virus spread rate for advance preparation in the healthcare services to avoid deaths. The paper provides a real-time prediction for shaping and tracking Covid-19 disease in Mysore, Koppal and Kolar districts reckoning the coronavirus disease asperity, predicting the extent of the pandemic with supporting government and health staffs to constitute strategy and competent verdicts towards the eradication of Covid-19 disease.

REFERENCES

- [1] "Machine learning prediction for covid-19 pandemic in India", Roseline Oluwaseun, Joseph Bamidele, IEEE,2020.
- [2] "Artificial Intelligence (AI) and big data for coronavirus (COVID-19) pandemic: A survey on the state-of-the-arts", Won Joo Hwaang, Ji Honn Joon Jae, IEEE, 2020.
- [3] "Artificial intelligence in the battle against coronavirus (COVID-19): A survey and future research directions", Mohammad Jamshidi ,Ali Lalbakshh,IEEE,2020.
- [4] "Analysis of the distribution of COVID-19 in Italy using clustering algorithms", Anastasiya Doroshenko, IEEE, 2020.
- [5] "Regression analysis of covid-19 using machine learning algorithms", Ekta Gambhir, Ritika Jain, Alankrit Gupta, ICOSEC,2020.
- [6] "A Novel parametric model for the prediction and analysis of the COVID-19 casualties", Onder tutsoy Sule Colak,Adam Polat,IEEE ,2020.
- [7] "COVID-19 future forecasting using supervised machine learning models" Furqan Rustam, Aijaz Ahmad Reshi, Arif Mehmood, Saleemullah, Wqqr Aslam, IEEE, 2020.
- [8] "Prediction and forecast for covid-19 outbreak in India based on epimediological models" Siddharth Singh, Piyush Raj, Raman Kumar, Rishu Chaujar, ICIRCA,2020.
- [9] "Covid-19 pandemic prediction using time series forecasting model", Naresh Kumar,Seba Susan,IEEE,2020.
- [10] "DeepcovidNet: An interpretable Deep learning model for predictive surveillance of covid-19 using heterogenous", Ankit Ramchandani, ChaoFan , and Ali Mostafavi, IEEE, 2020.