 **LEAP ACADEMY**

**TRIBE – I**

**TRIBE MEMBERS NAMES:**

1. **ABHIJEET MAHINDRAKAR**
2. **RITUPRIYA**
3. **SALOMY**

**UNDER THE GUIDANCE: DINESH BEHERE**

1. **INTRODUCTION**

The novel coronavirus (COVID-19) was widely reported to have first been detected in Wuhan (Hubei province, China) in December 2019. After the initial outbreak, COVID-19 continued to spread to all provinces in China and very quickly spread to other countries within and outside of Asia. At present, over 45 million cases of infected individuals have been confirmed in over 180 countries with in excess of 1 million deaths. Although the foundations of this disease are very similar to the severe acute respiratory syndrome (SARS) virus that took hold of Asia in 2003, it is shown to spread much more easily and there currently exists no vaccine.

Analyses on North and South America have also used similar classical methods, for example model the progression of the outbreak in the United States until the end of 2021 with the simple Susceptible-Infected-Recovered model, and predict epidemic trends in Brazil and Peru using a logistic growth model and machine learning techniques. However, other studies include: analysis of the spatial variability of the incidence in the United States using spatial lag and error models, and geographically weighted regression estimation of the number of deaths in the United States using a modified logistic fault-dependent detection model estimating prevalence and infection rates across different states in the United States using a sample selection model investigating the relationship between social media communication and the incidence in Colombia using non-linear regression models.

Focusing on Africa, simulate and predict the spread of the disease in South Africa, Egypt, Algeria, Nigeria, Senegal, and Kenya, using a modified Susceptible-Exposed-Infectious-Recovered model; apply a six-compartmental model to model the transmission in South Africa; predict the spread of the disease in West Africa using a deterministic Susceptible-Exposed-Infectious-Recovered model; implement Autoregressive Integrated Moving Average models to forecast the prevalence of COVID-19 in East Africa; predict the spread of the disease using travel history and personal contact in Nigeria through ordinary least squares regression; use logistic growth and Susceptible-Infected-Recovered models to generate real-time forecasts of daily confirmed cases in Saudi Arabia.

Aside from many of the classical models mentioned above, recent developments in the econometrics and statistics literature have led to a number of new models that could potentially be applied in the modelling of infectious diseases. These include (but are not limited to) mixed frequency analysis, model selection and combination, and dynamic time warping. Mixed frequency analysis is an iterative approach proposed for dealing with the joint dynamics of time series data which are sampled at different frequencies. In the economic literature, the common example is quarterly gross domestic product (GDP) and monthly inflation notes that studying the co-movements between mixed frequency data usually involves analysing the joint process sampled at a common low frequency; however, this can mix-specify the relationship. Propose vector autoregressive models for mixed frequency analysis that operate at the highest sampling frequency of all the time series in the model. These models allow for the modelling of the joint dynamics of the dependent and independent variables using time disaggregation, where the low frequency variables are interpolated and time-aggregated into a higher frequency. In the context of infectious diseases, such models could be beneficial for modelling the relationship between higher frequency data such as the number of daily cases or deaths and lower frequency data relating to, say, weekly cases or deaths, news and information about health prevention measures, etc. They note that there are many scenarios that generate multiple, interrelated time series, where the dependence has a significant impact on decisions, policies, and their outcomes. In addition, methods need to learn and integrate information about forecasters and models, bias, etc. and how they change over time, to improve their accuracy.

Decision and policy makers often use multiple sources, models, and forecasters to generate forecasts, in particular, probabilistic density forecasts. However, although complex estimation methods may have useful properties for policy makers, large standard deviations may be a result of the complexity of the data, model, etc., and it may be difficult to know the source. The aim is to use the dependencies between time series to improve forecasts over multiple horizons for policy decisions. For example, in the economic literature, setting interest rates based on utility or loss that account for inflation, real economy measures, employment, etc. BPS relates to a decision maker that accounts for multiple models as providers of “forecast data” to be used for prior-posterior updating. The decision maker learns over time about relationships between agents, forecasts, and dependencies, which are incorporated into the model, and dynamically calibrate, learn, and update weights for ranges of forecasts from dynamic models, with multiple lags and predictors. In epidemiology, BPS could potentially be used in a similar context to analyse the dependency between various interrelated time series such as daily cases and deaths, hospital capacity, number vaccinations, etc.

Different models and sources of data could then be combined and characterised in one single model improving the accuracy of forecasts. Dynamic time warping as noted by is a technique that has not been widely used outside of speech and gesture recognition. It can be used to identify the relation structure between two time series by describing their non-linear alignment with warping paths. The procedure involves a local cost measure characterising the sum of the differences between pairs of realisations of data at each time point, where an optimal warping path gives the lowest total cost. The optimal path is found under a variable lead-lag structure, where the most suitable lag can then be found. This then reveals and identifies the lead-lag effects between the time series data. Indeed, dynamic time warping has recently been used in the modelling of COVID-19 by use the method to determine the lead-lag relation between the cumulative numbers of daily cases of COVID-19 in various countries, in addition to forecasting the future incidence in selected countries. This allows for the classification of countries as being in the early, middle, and late stages of an outbreak.

Controlling an infectious disease such as COVID-19 is an important, time-critical but difficult issue. The health of the global population is, perhaps, the most important factor as research is directed towards vaccines and governments scramble to implement public health measures to reduce the spread of the disease. In most countries around the world, these measures have come in the form of local or national lockdowns where individuals are advised or required to remain at home unless they have good reason not to—e.g. for educational or medical purposes, or if they are unable to work from home. However, the implications of trying to control COVID-19 are being felt not only by the health sector, but also in areas such as the economy, environment, and society.

As the number of cases of infected individuals has risen rapidly, there has been an increase in pressure on medical services as healthcare providers seek to test and diagnose infected individuals, in addition to the normal load of medical services that are offered in general. In many cases, trying to control COVID-19 has led to a backlog for and deprivation of other medical procedures , with healthcare providers needing to find a balance between the two note that this conflict may change the nature of healthcare with public and private health sectors working together more often. The implementation of restrictions on the movement of individuals has also led to many suggesting that anxiety and distress may lead to increased psychiatric disorders. These may be related to suicidal behaviour and morbidity and may have a long-term negative impact on the mental health of individuals.

In addition to restrictions on the movement of individuals, governments have required most non-essential businesses to close. This has negatively impacted national economies with many businesses permanently closing leading to a significant increase in unemployment. Limits on travel have severely affected the tourism and travel industries, and countries and economies that are dependent on these for income. Whilst many of the implications of controlling COVID-19 on the economy are negative, there have been some positive changes as businesses adapt to the ‘new normal’. For example, the banking industry is dealing with increased credit risks, while the insurance industry is developing more digital products and pandemic-focused solutions. The automotive industry is expected to see profits reduced by approximately $100 billion, which may be offset by the development of software subscription services of modern vehicles. Some traditional office-based businesses have been able to reduce costs by shifting to remote working, while the restaurant industry has shifted towards takeaway and delivery services.

In terms of the environment, the limitations on businesses that have been able to continue operating throughout the epidemic has led to possible improvements in the environment—mainly from the reduction in pollution. However, societal issues have been exacerbated note that the reduction in the labour force that has resulted from controlling for COVID-19 has affected ethnic minorities and women most significantly. Furthermore, in many countries health services employ more women than men creating a dilemma for working mothers—either leave the labour force and provide childcare for their families or remain in employment and pay extra costs for childcare.

As a result of this on-going pandemic, new results and reports are being produced and published daily. Thus, our motivation stems from wanting to contribute to the statistical analysis of the incidence of COVID-19 in Italy and Spain, where the literature is limited.

The main contributions of this paper are:

1. To model the incidence of COVID-19 in Italy and Spain using simple mathematical models in epidemiology
2. To provide estimates of basic measures of the infectiousness and severity of COVID-19 in Italy and Spain

3) To investigate the predictive ability of simple mathematical models and provide simple forecasts for the future incidence of COVID-19 in Italy and Spain.

1. **ABSTRACT**

COVID-19, the pandemic has created a fearsome sensation the entire world. Although in India, many articles of research had examined the spread of this deadly disease, but the population of our country makes it more important to have a closer look at the present scenario of the individual states. The whole world is facing this pandemic for the past 15 months since December 2019, the month when the very first case of covid-19was discovered in China from where it got spread throughout the globe. In January 2020, the first case of covid-19 in India was witnessed. For this very scenario in this paper, we have analysed and crafted different sections for treatment methodologies and safety measures undertaken, detailed analysis of various States and Union Territories during the outbreak of Covis-19, count on the number of people infected, cured, and died along with the prediction of the upcoming situation of covid-19 in future months. Moreover several time series models had been studied among which Prophet model fits best to our purpose of study.

**3. LITERATURE SURVEY**

[1] B. Biswas, S. Debdas, S. Samanta, S. Chakraborty, A. Chhangani and S. Mohanty, "Covid-19 Data Analysis of India and Future Forecasting Using FbProphet," 2021 International Conference in Advances in Power, Signal, and Information Technology (APSIT), 2021, pp. 1-7, doi: 10.1109/APSIT52773.2021.9641221.

[2] N. Darapaneni, P. Jain, R. Khattar, M. Chawla, R. Vaish and A. R Paduri, "Analysis and Prediction of COVID-19 Pandemic in India," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2020, pp. 291-296, doi: 10.1109/ICACCCN51052.2020.9362817.

In this paper, we have analysed the COVID-19 progression in India and the three most affected Indian states (viz. Maharashtra, Tamil Nadu and Andhra Pradesh) as of 29-Aug-20 and developed a prediction model to forecast the behaviour of COVID-19 spread in the future months. We used time series data for India and applied the Susceptible-Infective-Removed (SIR) model and the FbProphet model to predict the peak invectives and peak infective date for India and the three most affected states. In this paper, we further performed the comparative analysis of the prediction results from SIR and FbProphet models. From this study, we concluded that with the assumption that a total 5% of India's population might be infected by the pandemic; the countrywide spread is forecasted to reach its peak by the end of Nov-20. And till the time there is no vaccination, for the states that have already reached their peak and with festivals around the corner, there are high chances of resurgence in the number of cases if the social distancing and other control measures are not followed diligently in the coming months.

[3] C. M. Kim, E. J. Hong and R. C. Park, "Chest X-Ray Outlier Detection Model Using Dimension Reduction and Edge Detection," in IEEE Access, vol. 9, pp. 86096-86106, 2021, doi: 10.1109/ACCESS.2021.3086103.

With the advancement of Artificial Intelligence technology, the development of various applied software and studies are actively conducted on detection, classification, and prediction through interdisciplinary convergence and integration. Among them, medical AI has been drawing huge interest and popularity in Computer-Aided Diagnosis, which collects human body signals to predict abnormal symptoms of health, and diagnoses diseases through medical images such as X-ray and CT. Since X-ray and CT in medicine use high-resolution images, they require high specification equipment and huge energy consumption due to high computation in learning and recognition, incurring huge costs to create an environment for operation. Thus, this paper proposes a chest X-ray outlier detection model using dimension reduction and edge detection to solve these issues. The proposed method scans an X-ray image using a window of a certain size, conducts difference imaging of adjacent segment-images, and extracts the edge information in a binary format through the AND operation. To convert the extracted edge, which is visual information, into a series of lines, it is computed in convolution with the detection filter that has a coefficient of 2n and the lines are divided into 16 types. By counting the converted data, a one-dimensional 16-size array per one segment-image is produced, and this reduced data is used as an input to the RNN-based learning model. In addition, the study conducted various experiments based on the COVID-chest X-ray dataset to evaluate the performance of the proposed model. According to the experiment results, the LFA-RNN showed the highest accuracy at 97.5% in the learning calculated through learning, followed by CRNN 96.1%, VGG 96.6%, AlexNet 94.1%, Conv1D 79.4%, and DNN 78.9%. In addition, LFA-RNN showed the lowest loss at about 0.0357.

[4] A. S, R. F. Johnson, R. k. N, M. T R and V. V, "Predicting the number of new cases of COVID-19 in India using Survival Analysis and LSTM," 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2021, pp. 1-4, doi: 10.1109/I-SMAC52330.2021.9640899.

COVID-19 has been the cause of death for thousands of people across the globe. The goal of this paper is to forecast the new COVID-19 cases in India. The other methods used to forecast COVID-19 cases fail to give results with good accuracy when they try to predict the new cases number for a long time period or when the count of daily cases reported is large since the population of a country is large. The proposed study overcomes the challenge by firstly customizing the dataset. Second, the survival analysis has been utilized to choose appropriate factors, and third, the data will be integrated into the Long Short-Term Memory Network (LSTM). With a mean absolute percentage error of 5.79 percent, data from the 30th of January, 2020, to the 16th of June, 2021, was used to determine the new cases number of every day for the next 21 days.

[5] K. Prathyusha, K. Helini, C. V. Raghavendran and N. Kumar Kurumeti, "COVID-19 in India: Lockdown analysis and future predictions using Regression models," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 899-904, doi: 10.1109/Confluence51648.2021.9377052.

The new virus named COVID-19 identified in Wuhan, China causes a severe impact on the respiratory system of the human. In considering its effect and spread in the community, the Government of India has imposed World’s biggest Lockdown from 25th March 2020. Later on, it was extended in another three phases as Lockdown 2.0, 3.0, and 4.0 with some relaxations in each Lockdown. In this paper, we have studied the COVID-19 patients’ data of Confirmed cases, Recovered cases, and Deaths based on before, after, and during lockdowns. The data analysis is done basing on the daily growth rate of confirmed cases, recovery rate, and fatality rate. We have applied Regression techniques viz., Linear Regression, Polynomial Regression of Machine Learning (ML) to predict the future spread of this virus in India. The Polynomial Regression has given accurate predictions comparing with the Linear Regression.

[6] S. Chordia and Y. Pawar, "Analysing and Forecasting COVID-19 Outbreak in India," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 1059-1066, doi: 10.1109/Confluence51648.2021.9377115.

The unprecedented outbreak of the COVID-19 virus has infected more than 50 million people all over the world in less than a year. More than 1 million people have lost their lives due to the on-going pandemic. The pandemic struck India on January 30, 2020, when the first positive case of COVID-19 was identified in Kerala. Today, India is one of the most adversely affected countries in the world. Hence, it is of utmost importance to analyse the trends in India and use the adopted knowledge to forecast the future course of outcomes. Along with the overall trend analysis in India, this study also takes into account 5 most affected states of the country: Maharashtra, Andhra Pradesh, Tamil Nadu, Karnataka and Uttar Pradesh as the subjects of the research. ARIMA and Prophet time series forecasting models have been used to make three types of predictions: confirmed cases, deaths and recovered cases in India as well as in the adopted states. The effectiveness of the forecasting models is evaluated based on metrics such as Root Mean Squared Error, Mean Absolute Error, Mean Absolute Percentage Error and Coefficient of Determination. The results suggest that the adopted models are promising mechanisms for forecasting COVID-19 trends. Our study also suggests that ARIMA model performs better than Prophet Model at this task of forecasting the outbreak. The forecasts can be useful in increasing the preparedness level of government authorities, health facilities and hospitals to combat against massive spread of the virus.

**4. OBJECTIVES**

1. The aim of this analysis and study the infection spreading through out , using open source data visiluatization which features to extract the data. The tool will help to extract huge data of Countries.
2. This project aim to get the information of covid cases acoss the world in the simple graphical formats.
3. This project also aim to get result data of infected patient,cured patient and different varient of Covid-19.
4. This project aim analyse data on the number of infected people in each Indian state (restricted to only those states with enough data for prediction).

1. It enables public health personnel to track the disease spread, vaccination, and adverse events across different sub-populations [. It also allows researchers to gain new insights, and to scrutinize the data to understand the rationale behind the policies put forth by the government.
2. Granular data is also more transparent and informative for the general public. Governments cannot give personalized health recommendations to each citizen. The government’s advice on mask mandate, lockdowns, and vaccination is designed as a standard recommendation for all the citizens. However, health is a personal matter.

**5. PROBLEM STATEMENT**

* The pandemic has already taken grip over peoples’ life. Since the start of the pandemic, some countries are facing problem of ever-increasing cases. Through the data analysis of cases one can analyse how countries all over the world are doing in terms of controlling the pandemic. Analysing data leads to adapt the prevention model of the countries that are doing great in terms of lowering the graph. Predictions are made with the dataset available to the individual/country/organisations, thus helping them to decide how far they are able to control the pandemic or up to how much extent they should guide preventive measures. Through this project, a step towards helping people to understand the spread and predict the cases in their country is done. This project also gives an insight of how a country is doing in terms of limiting the spread.
* Deficiency of Data over death and Recovered from Covid-19 across the world.
* Lack of Data available to covid-19 detection and Cure.
* Data over the Senior Citizen,adult and children who are infected and cured around the world.
* This project will supervised learning technique is used for detection of Covid-19 and analysis of the consequence of it.

**6. EXISTING SYSTEM**

Based on the aggregated, anonymous data, Corona Surveys employs several methods to produce estimates of the number of COVID-19 cases in all geographical areas for which sufficient data are available, comparing these estimates with those provided by the official authorities. The estimation methods are:

• **cCFR-based:** This method is based on estimating the corrected case fatality ratio (cCFR), from the official numbers of cumulative cases and fatalities, and taking an estimation of the approximate number of cases with known outcomes into consideration. It is also assumed that a reliable value of the traditional case fatality ratio (CFR\*CFR\*) is available (We use CFR\*=1.38%CFR\*=1.38% with a 95%95% confidence interval of 1.23%1.23% and 1.53%1.53%, as described in. Then, the number of cases is estimated by multiplying the official figure of cumulative cases in a region D by the ratio cCFR (D)/CFR\*cCFR (D)/CFR\*, where cCFR (D) cCFR (D) is the cCFR estimated for D.

• **CCFR-fatalities:** This method divides the official number of fatalities on a given day d by CFR\*CFR\*, and assigns the resulting number of cases to day d−Pd−P (P is the median number of days from symptom onset to death). Here used P=13P=13, following the values reported by the Centre’s for disease Control and Prevention.

• **UMD-Symptom-Survey:** This method uses the responses to direct questions about symptoms from the University of Maryland COVID-19 World Survey to estimate active cases. In particular, it counts the number of responses that declare fever, and cough or difficulty breathing. This survey collects more than 100,000100,000 individual responses daily.

• **UMD-Symptom-Survey-Indirect:** This method estimates active cases applying the NSUM method to the responses of an indirect question from the University of Maryland COVID-19 World Survey. In this estimation method the Reach is obtained from the Corona Surveys data, while the Number of Cases are the cases reported by answering YES to the question 1) “Do you personally know anyone in your local community who is sick with a fever and either a cough or difficulty breathing?” and answering the question 2) “How many people do you know with these symptoms?”

• **300Responses:** This method uses a weighted average of 300 filtered Corona Surveys responses for a given geographical area. Filtering consists in discarding answers that report an unusually large reach (entries larger than 1.5 times the interquartile range above the upper quartile) or an unusually large number of cases (over 1/31/3 of cases in the reach).

• **Estimates-W:** This method uses a weighted average of Corona Surveys responses from the last W days, using the same filtering criteria as 300Responses.

**7. PROPOSED SYSTEM**

* Prosposed project will be dealing different Covid cases across India.This Cases are concern of Government and different organization dealing with it.The data will be segregated and Process the data obtained from it.
* One of the main features of this project is to visualize the data using open source data visualization Tool.
* Various studies have been conducted on the COVID-19 with different data visualization tools to analysis the spread to COVID.
* This data obtained from the output will be used by state government to Organize various policy for people who suffered during panic Directly or Indirectly.

**8. REQUIREMENT SPECIFICATIONS**

8.1 Hardware Requirements

* Processor : Intel i3 or Above or AMD processor
* RAM: Min 4GB and above.
* Hard Drive: Min 250GB and above
* GPU (optional)

8.2 Software Requirements

* Google Colab or Jupiter Book
* Kaggle Library
* Dataset related to Covid-19
* Libraries Used :

1) Pandas: Working with data files

2) Numpy: for Scientific Calculation

3) Matplotlib: Basic Visualization

4) Plotly: Advance Visualization

5) Folium: Open Source Data Visualization Tool.

**9. METHODOLOGY**

Our methodology consists of three main experiments to evaluate the performance of the models and assess the influence of the different stages of the process. Each experiment follows the workflow. The difference between experiments is the dataset used. In all instances, the same images for COVID-19 positive cases were used. Meanwhile, three different datasets for negative cases will be used. In that order, Process 1 and 2 consists of evaluating positive vs. negative cases datasets, and Process 3 involves Pre-COVID era images (images from 2015-2017).

9.1. Datasets

A total of 13 Excel datasets were used in different stages:

9.2 COVID-19 classification datasets

This dataset consists of CXRs from different individuals with COVID-19, 1341 CXRs from healthy individuals, and 1345 CXRs from individuals with other types of viral pneumonia. All the images are in the Portable Network Graphics (PNG) file format, and with a resolution of either 1024-by-1024 pixels or 256-by-256 pixels. It must be noted that the dataset is divided into 3575 training and 311 test images, as outlined in Table [1](https://www.hindawi.com/journals/ijbi/2021/8828404/tab1/). In the training phase, the dataset was prepared and verified as reliable by reviewing it with chest specialists. In addition, cases of viral pneumonia should be free from any instances of COVID-19. Before passing the images into a pertained model for feature extraction, we resized all images to a size of pixels. All images were normalized according to the pertained model standards. Figure [2](https://www.hindawi.com/journals/ijbi/2021/8828404/fig2/) shows examples of CXR images within the training set that were used in this study.

### 9.3. Classify a Dataset Mathematically.

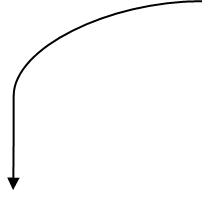
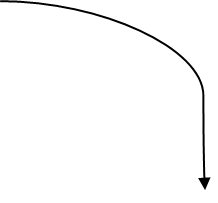
In the mathematical modelling of infectious diseases, there exist many compartmental models that can be used to describe the spread of a disease within a population. One of the simplest models is the SIR (Susceptible-Infectious-Recovered) model proposed, in which the population is split into three groups or compartments: those who are susceptible (S) but not yet infected with the disease; those who are infectious (I); those who have recovered (R) and are immune to the disease or who have deceased. The SIR model has been extensively researched and applied in practice, thus it would not be practical to mention and cover all of the literature. However, some of the most prominent literature covers areas such as the stability and optimality of the simple SIR model , pulse vaccination strategy in the SIR model; applications of the SIR in the modelling of infectious diseases.

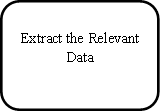
With regards to COVID-19, many have applied the basic SIR model (or slightly modified versions) to model the outbreak. Some particular examples include (but are not limited to):who estimate the overall symptomatic case fatality risk of COVID-19 in Wuhan and use the SIR model to generate simulations of the COVID-19 outbreak in Wuhan, who apply a modified SIR model to identify contagion, recovery, and death rates of COVID-19 in Italy; who combine the SIR model with probabilistic and statistical methods to estimate the true number of infected individuals in France; who use a number of methods including the SIR model to estimate the basic and controlled reproduction numbers for the COVID-19 outbreak in Wuhan, China; who show that the basic SIR model performs better than extended versions in modelling confirmed cases of COVID-19 and present predictions for cases after the lockdown of Wuhan, China, who model the temporal dynamics of COVID-19 in China, Italy, and France, and find that although the rate of recovery appears to be similar in the three countries, infection and death rates are more variable; who simulate the outbreak in Wuhan, China, using an extended SIR model and investigate the age distribution of cases; who study the number of infections and deaths from COVID-19 in Sweden using the SIR model; who use the SIR model, with an additional parameter for social distancing, to model and forecast the early stages of the COVID-19 outbreak in India.

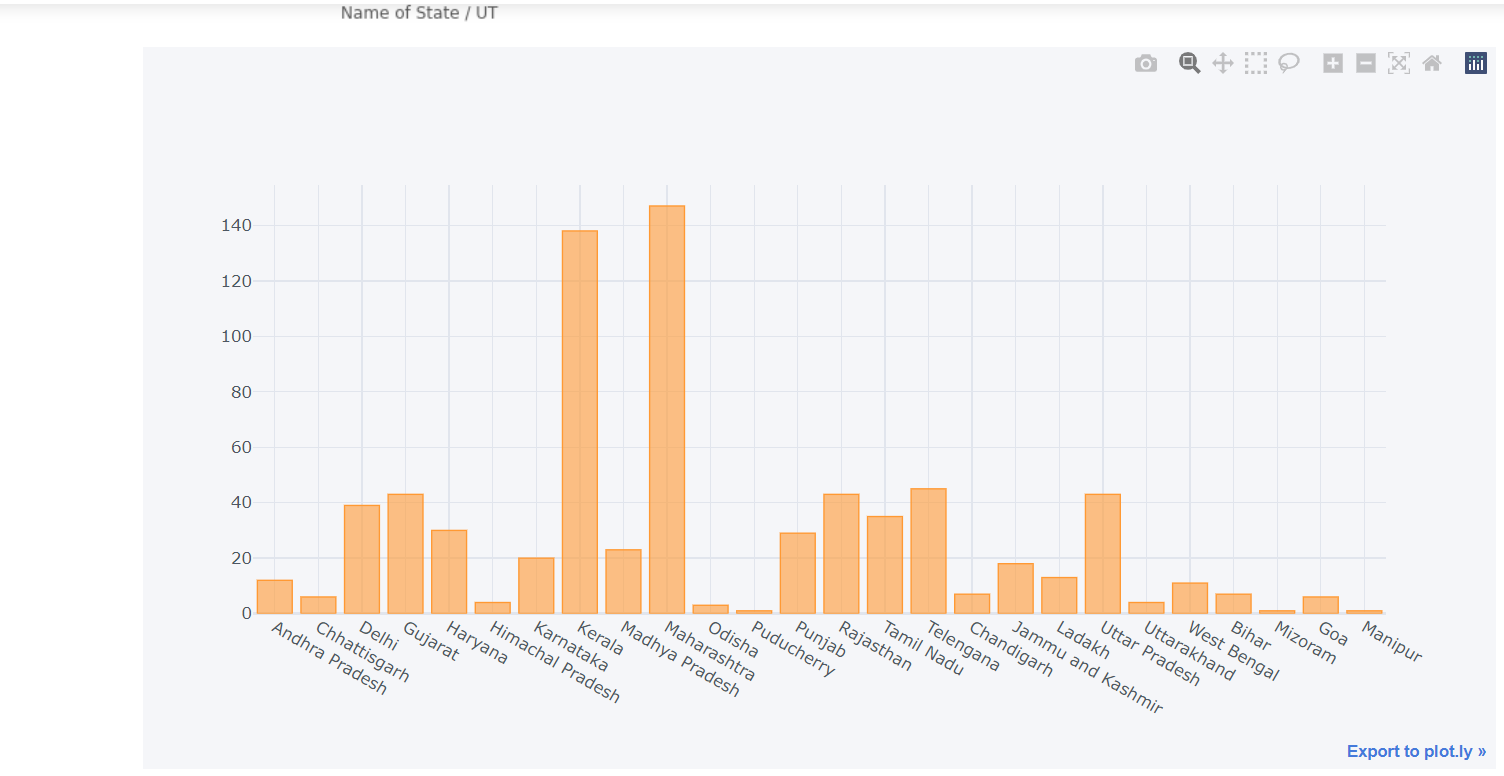




Prediction Analysis of using various Regression model

9.1 Data-flow diagram of COVID-19 Analysis of India.

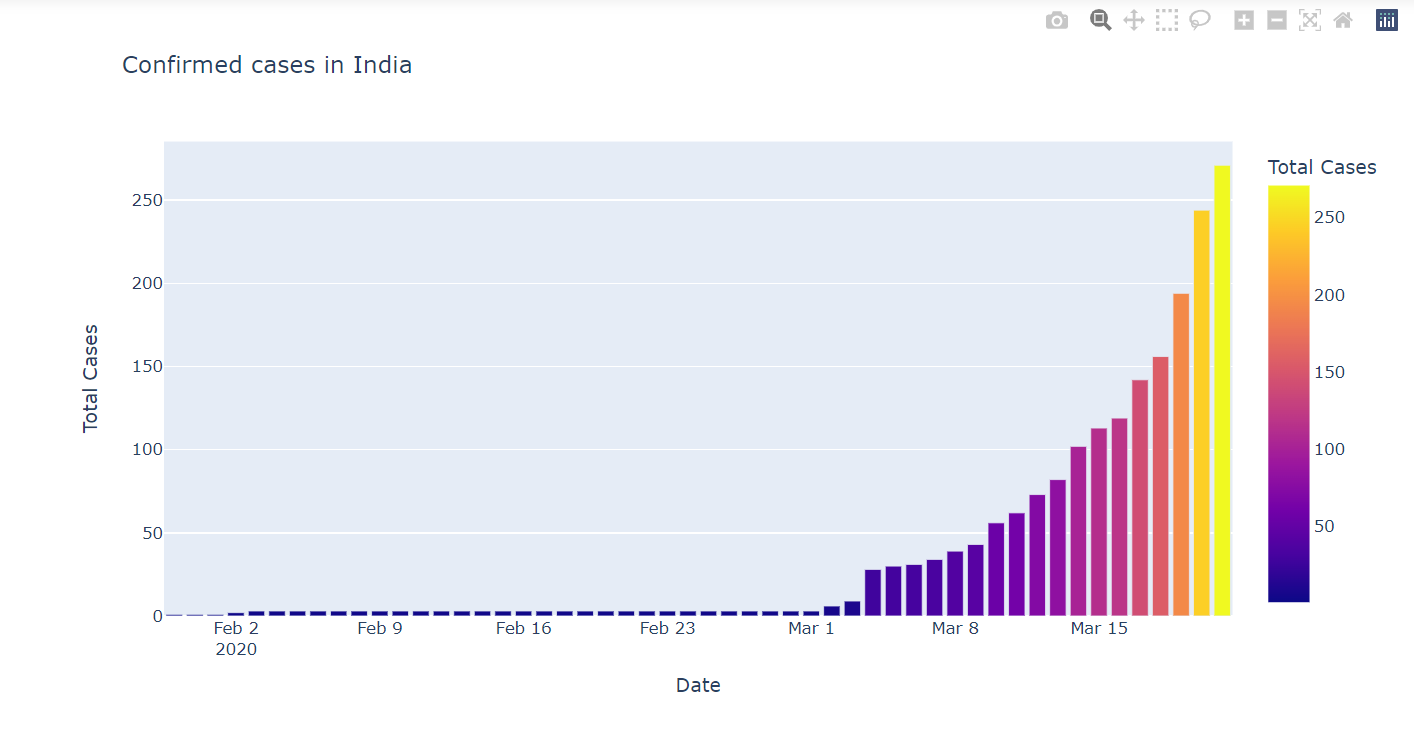
From the above diagram Covid -19 Data of India, which are unorganized will be taken into consideration for further processing, this further will be pre-processed and organized, and the real time data will be extracted using time series, this data will be arranged, according to recovered and the death.



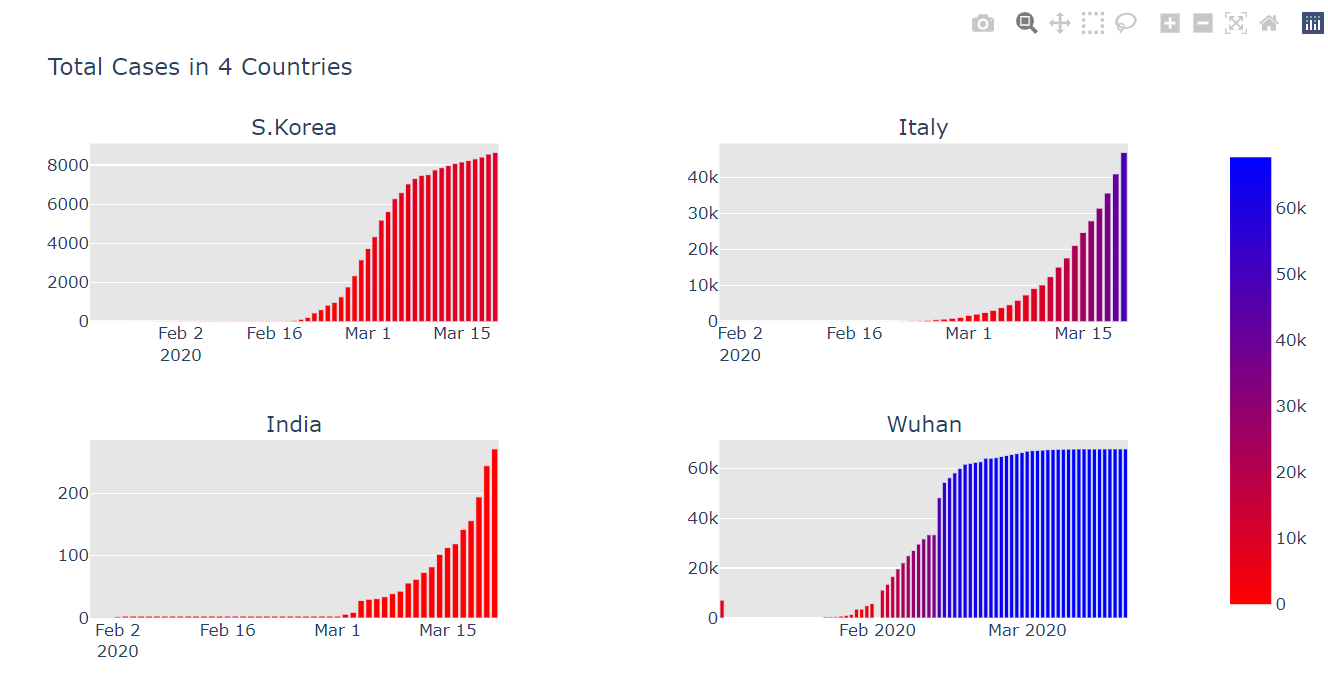
9.2 Cases across the different states

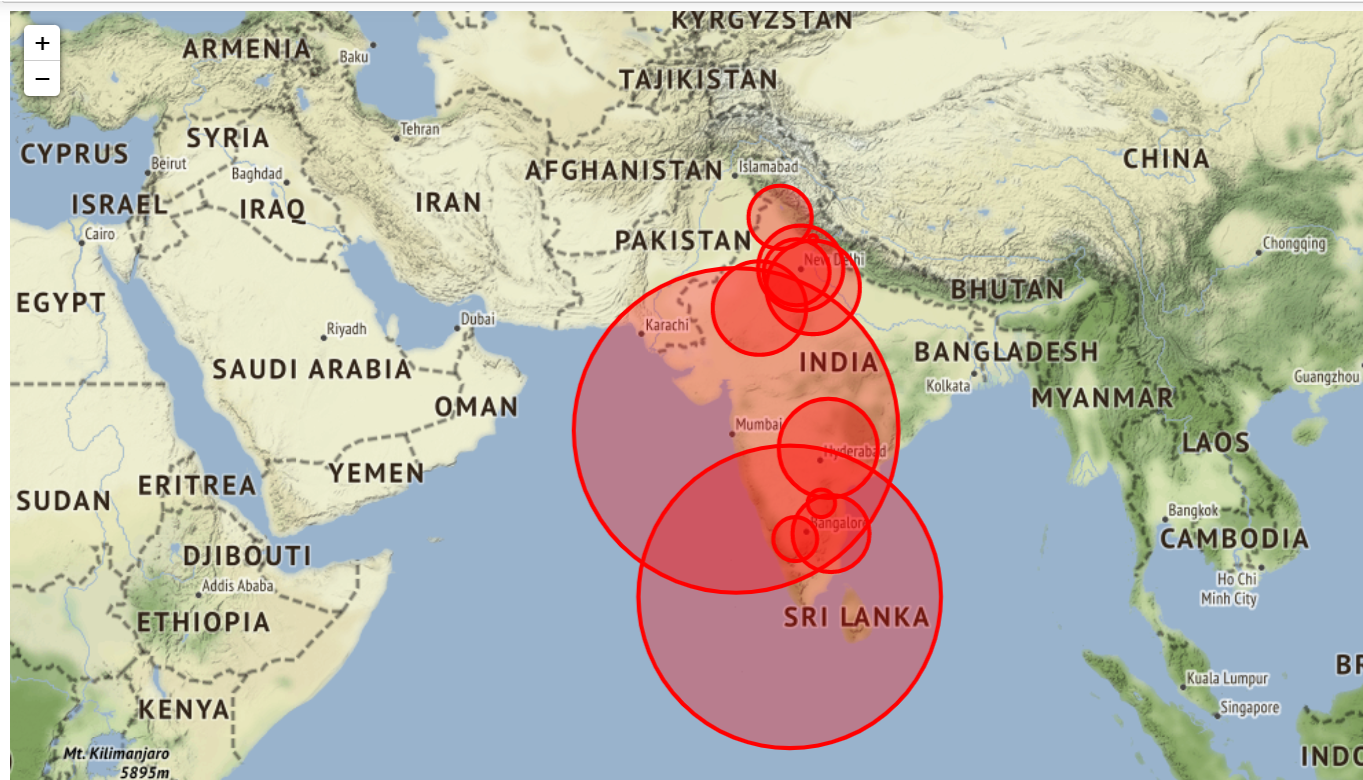
This data will be put together and the relevant data will be extracted from it. The generated data will be in numerous amounts. To enhance this data, this data will visualized and further analysed to Graphs and Charts.

9.3 Confirmed Cases in India



9.4 Case in World compared to India





9.5 Epic Center of Higest covid cases across India

**10.TECHNIQUE OR ALGORITHM**

* At first the covid dataset will be uploaded to the system,this dataset will be having different logs of covid cases.
* Then,this data will be organized and processed for futher analysis.
* The anaylised data will evaluated and calculated the numeric values.
* The Evaluated data will be processed and visualisied in demographic format.
* The demograph and chart obtained from the dataset will be shown statewise.
* The processed data will show confirmed and negative case through graphical representation.

**11.ADVANTAGES AND DISADVANTAGES**

**Advantages :**

* Easy to analysis the data on using the open source data visualization tool.
* Getting the information and represent Demograph of huge amount of data becomes easy.
* Easy to analysis the infection spread and confirmed cases across the country.
* Analysing the recovered patient across the country becomes easy.

**Disadvantages:**

* Dataset are not easily available.
* Complex process for Data visualization.

**CONCLUSION**

Through this project, the analysis on COVID-19 data has been performed successfully. The analysis on this pandemic spread has been done and compared between different countries. The analysis of confirmed cases, active cases, recovered cases and deaths are done separately to give a clear look on how the virus is spreading, which countries are getting affected mostly and how different countries are recovering. A separate analysis on cases of INDIA has been done and predictions of different cases both around the world and INDIA have been done. At last, the accuracy check using different metrics is performed over all the analysis done in this project.

**REFERENCES**

[1] Liu C, Zhou Q, Li Y, Garner LV, Watkins SP, Carter LJ, et al. Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. ACS Central Science. 2020; 6(3):315–331. Pmid: 32226821

[2] Chen X, Han W, Wang G, Zhao X. Application prospect of polysaccharides in the development of anti-novel coronavirus drugs and vaccines. International Journal of Biological Macromolecules. 2020; 164:331–343. pmid: 32679328

[3] Kim E, Erdos G, Huang S, Kenniston TW, Balmert SC, Carey CD, et al. Micro needle array delivered recombinant coronavirus vaccines: Immunogenicity and rapid translational development. EBioMedicine. 2020; 55:102743. Pmid: 32249203.

[4] de León UAP, Pérez ÁG, Avila-Vales E. An SEIARD epidemic model for COVID-19 in Mexico: mathematical analysis and state-level forecast. Chaos, Solitons & Fractals. 2020; 140:110165.

[5] Mena RH, Velasco-Hernandez JX, Mantilla-Beniers NB, Carranco-Sapiéns GA, Benet L, Boyer D, et al. Using posterior predictive distributions to analyse epidemic models: COVID-19 in Mexico City. Physical biology. 2020;17(6):065001. Pmid: 32959788

[6] Vorra B, Fernandez MR, Vela-Pérez M, Ramos A. Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. Communications in nonlinear science and numerical simulation. 2020; 88:105303. Pmid: 32355435

[7] Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science. 2020; 368(6489):395–400. Pmid: 32144116

[8] Drew DA, Nguyen LH, Steves CJ, Menni C, Freydin M, Varsavsky T, et al. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. Science. 2020;368(6497):1362–1367. pmid:32371477

[9] SZens M, Brammertz A, Herpich J, Südkamp N, Hinterseer M. App-based tracking of self-reported COVID-19 symptoms: analysis of questionnaire data. Journal of medical Internet research. 2020;22(9):e21956. pmid:32791493

[10] K. S. and D. Radha, "Analysis of COVID-19 and Pneumonia Detection in Chest X-Ray Images using Deep Learning," *2021 International Conference on Communication, Control and Information Sciences (ICCISc)*, 2021, pp. 1-6, doi: 10.1109/ICCISc52257.2021.9484888.