COVID-19 CASE ANALYSIS

A. NAGARAJU, SALA GEETHA BHAVANI
1.Assistant Professor in DEPARTMENT OF MASTER OF COMPUTER SCIENCE, BHIMAVARAM-534202.
Email.id:-nagaraju.dnr345@gmailcom
1.PG STUDENT, D.N.R. COLLEGE,P.G. COURSES (AUTONOMOUS), BHIMAVARAM-534202.
Email id:- salageetha545@gmail.com

ABSTRACT

COVID-19 pandemic is one of the prevalent challenges mankind has ever faced and there is a lot of uncertainty prevailing over the future with respect to it. In this situation, machine learning algorithms can be useful for real-time analysis and prediction of the trends of infections. The objective of our project is to analyze the COVID-19 and vaccination trend globally and individually, and forecast the trend of the pandemic in the near future. Considering the variation of the scenario with time, it has been observed to analyze the data with the time series analysis models. This analysis has been conducted using six time series forecasting methods, viz. AR, MA, ARIMA, Holt's Linear Trend, Holt's Winter Seasonal and FB Prophet. It is of utmost importance to identify the future infected cases, the virus spread rate and time required to vaccinate the total population for advance preparation in the healthcare services to avoid deaths, and social entities to tackle this pandemic across the world. The datasets are taken from Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU), Dataset on Novel Corona Virus Disease 2019 in India and COVID-19 World Vaccination Progress

.1. INTRODUCTION

The COVID-19 pandemic has had a significant global impact, necessitating the need for data analysis and predictions to understand the spread and trajectory of the virus. The COVID-19 Case Study and Predictions project seeks to address this need by providing a comprehensive analysis of COVID-19 data and utilizing machine learning models for predictions.

The outbreak of the new disease in Wuhan, Chinawas caused by novel Coronavirus (2019-nCoV) [1]. This diseaseis a form of pneumonia. Coronavirus belongs to the *Orthocoronavirinae*subfamily. The first case was observed atthe Chinese Center for Disease Control and Prevention (CDC) on 12 December 2019 and was considered as a non-SARS novel coronavirus [2]. The family to which Coronavirus belongs is *Coronaviridae*which consists of a large,single RNA strand of plus sign [3]. Viruses of these family show the symptoms of common cold, diarrhea in humanbeings. In the year 2003, it was seen the outbreak of coronavirus i.e. severe acute respiratory syndrome coronavirus(SARS-CoV) [4]. In December 2019 at Wuhan, China's symptoms closely resembled the same as pneumonia [5].Several cases of approximately 1974 were confirmed in China according to thecouncil information office inBeijing, China's capital on 26th January,2020. Virus started spreading inmany other countries likethevery firstcase after China was reported in Thailand, Japan and two cases were also seen in Korea on 16 January 2020. Recentresearches have shown some evidence of the origin of the virus from the bat and it was also seen that transmission

ofthevirusistakingplacefromhumantohuman.Thesituationstarted gettingworstfrom19January2020 daybyday,

sototakesomeseriousactionforthecontrolandpreventionfromthedisease.WorldHealthOrganisation(WHO)on 30 January 2020 declared that Coronavirus Disease was an outbreak emergency of international concern after theattack of H1N1 in 2009, the emergence of Ebola virus in 2014, polio in 2014 and Zika virus in 2016[6] [29].Finally, on 11 February 2020, World Health Organization (WHO) gave the name of the novel disease which wascaused by the corona virus as Corona Virus Disease- 19 (COVID-19) [7] [32]. Record maintenance on 24 February2020 showed that more than 78, 000 patients were suffering from COVID-19 throughout many countries. Themaximum patients were from China according to the World Health Organization (WHO) which were approximately77,000 and 2500 death [8]. According to the World Health Organization (WHO) the rest of the countries reported2000 confirmed cases and 300 deaths as on 7 March 2020. In Wuhan, China lockdown orders of all the trains, fightsand public transport were passed on 23 January 2020. The exact origin of COVID-19 was not reported but throughdifferent researches, it was seen that coronavirus possibly has originated from the bat. According to the Centers forDisease Control and Prevention (CDC), the novel disease COVID-19 was transmitted from person to person throughdroplets,and thesymptomsseen werefever,shortnessofbreath,and cough which wasseenafter 14 days[9].

2.LITERATURESURVEYANDRELATEDWORK

According to the research paper [1], the authors,
R. Wang, G. Hu, C. Jiang, H. Lu and Y. Zhang, have compared the prediction of patterns by using 3 methods and comparing their graphs with each other. These models are the conventional logical regression model, the Particle Swarm Optimization SIR model and the Lowest Square approach SIR model. The chart ultimately shows some patients with a novel form of X-axis coronary pneumonia, and Y-axis date. By seeing the three patterns we come to know that the data is plotted in the form of a curve.
“The public figures of daily updated confirmed instances of Covid-19 from University John Hopkins were analysed in this study article [2] proposed by V.Z.Marmarelis.[2]. RM as described by Riccati Equation, is the main modelling element for the method. The public figures of daily updated confirmed instances of Covid-19 from University John Hopkins were analysed in this study article [2] proposed by V.Z.Marmarelisetal. [2]. RM, as described by Riccati Equation, is the main modelling element for the method. Further by applying the equation we find 5 different parameters and their dependence on the no. of cases increasing day by day”.

Everyone analysed knowledge on coronary disease and sustainable therapy utilising research articles from Gerry Wolfe*, Ashraf elnashar*, Will Schreiber* Izzat Alsmadi*. " Guided by COVID-19 Literary Clustering of the Datasets from Kaggle based on COVID-19. [3] The data were further divided into four: (1) Mobility social distances, (2) Health and COVID; (3) Economic impact; and (4) Vulnerable population, and were utilised in a second dataset from MTI. The document has been analysed and text has been processed in order to produce tokens for clustering and the use of the K-Median method to label data to assist extract and analysecategorised data.

According to Tuli,[4] the epidemic may be tracked extremely efficiently via Shrestha et al Machine Learning (ML) and Cloud Computing, anticipate an outbreak of the illness, and create appropriate policies to regulate its expansion. Then given the array, face extraction and

collection is done. They have proposed a Machine Learning model that can be run continuously on Cloud Data Centers (CDCs) for accurate spread prediction and proactive development of strategic response by the government and citizens. The dataset used by them in this case study,World in Data by Hannah Ritchie. They have also used a cloud framework and azure instances for real time analysis of data The research paper [5] Francisco Nauber,Bernardo Gois et al. have emphasised the rising popularity of epidemic behaviour prediction research due to their capacity to anticipate the natural course of viruses. This study presents several predictor approaches with machine training, logistic regression, filters, and epidemiological models in order to explain COVID-19's behaviour.

The research paper [6], the authors Yazeed Zoabi, Shira Deri-Rozov and Noam Shomron have acknowledged that accurate SARS-CoV-2 screening allows for fast and efficient COVID-19 diagnosis and reduces the strain on health care systems. Prediction models using many characteristics have been created to assess the likelihood of infection. The model projected 0.90 auROC in the forward-looking test set (area under the receiver operating characteristic curve).

The research paper [7], authors Enis Karaarslan and DoğanAydın mentioned thatThe incident at COVID-19 showed that the world was unwilling to disseminate the virus so rapidly. One crucial factor in mitigating the detrimental impacts of an epidemic or pandemic is the effective use of information technology. They suggested a management epidemic system (EMS), which relies on the unfettered and timely flow of information between states and organisations. They have been using an MPISA paradigm, which allows different platforms to be integrated and gives the solution for issues of scalability and interoperability.

[8] This paper Describes the use of a new epidemiological compartiment-based model for the estimation of the propagation of the coronavirus CO VID-19, that is, SEIAR(Susceptible Exposed Asymptomatic Infectious Recovered). This is accomplished through the heuristic approach of differential
evolution. In this way the day(s) when that number reaches its maximum, the associated value and the future evolution of its spread may be evaluated in approximate order for different situations.

The [9] authors Ayyoubzadeh S et al haveUsedcomputerised data mining technologies for improved insights on the outbreak of COVID-19 in each country and globally for the management of the health catastrophe. Google Trends website collected data. For estimating the number of positive COVID-19 instances, linear regression and long-term memory (LSTM) models were utilised.

3.EXISTING SYSTEM

COVID-19 showed that the world was unwilling to disseminate the virus so rapidly. One crucial factor in mitigating the detrimental impacts of an epidemic or pandemic is the effective use of information technology. They suggested a management epidemic system (EMS), which relies on the unfettered and timely flow of information between states and organisations. They have been using an MPISA paradigm, which allows different platforms to be integrated and gives the solution for issues of scalability and interoperability.

DISADVANTAGES:

- Limited only for state
- Research works have been carried out without considering the specific area

4. PROPOSED SYSTEM

The proposed COVID-19 Case Study and Predictions system is built using Python and several libraries commonly used in data analysis and machine learning. The key components and features of the system include:

ADVANTAGES:

- We prepare our data library taken by a Kaggle dataset at first, then select the feature quantities,
- Each data is composed of a large number of attributes , and in an effort to extract sufficient amount of meaningful color information.

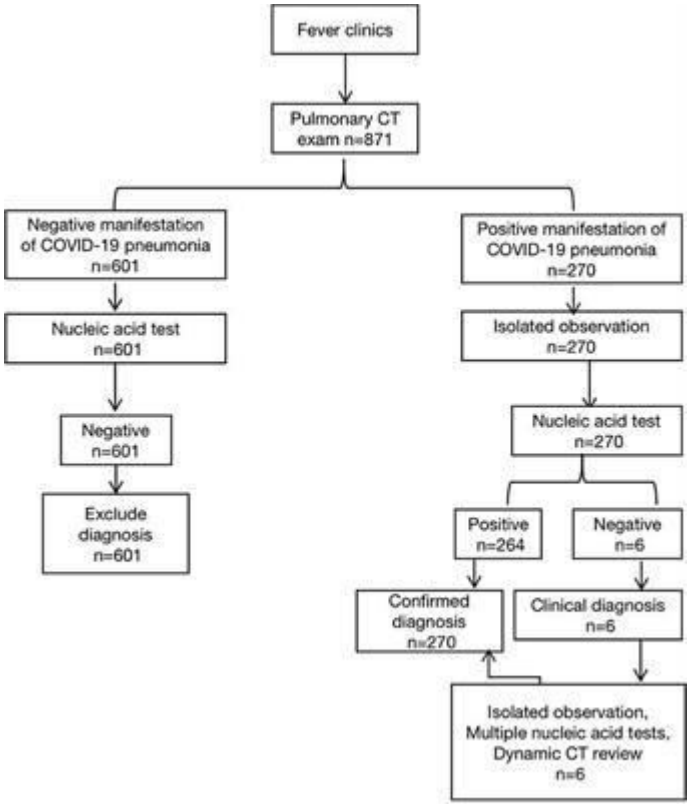

FIG 1 – SYSTEM ARCHITECTURE

5.METHODOLOGIES

MODULES

USER

ADMIN

DATA PREPROCESS

MACHINE LEARNING

MODULES DESCRIPTION:

We are using Machine Learning to give predictions on the basis of data taken from government website[11], and then we clean the data by using excel cleaning methods and give prediction by using the algorithm with highest accuracy to predict COVID -ve or +ve on basis on 5 major symptoms.

The process can be explain in following points -

1. First, Take the dataset, remove redundant data and organise the data according to our needs.
2. Second, Load the dataset on the Jupyter Notebook and apply data visualization techniques to understand the data better.
3. Third, then we calculate accuracy for various algorithms and plot a graph on the basis of accuracy of various algorithms.
4. Finally, using the accuracy graph we finally use the algorithm with best accuracy in this case (Decision Tree Classifier) to predict the person is either-ve or +ve on the basis of symptoms.

Description of the Process

We are building our own COVID Prediction System using Jupyter Notebook.

We can describe the process in following steps :

Step 1: Cleaning the dataset

The very first step in our project is to get a reliable and authentic dataset for the prediction and analysis.

Our search for dataset ended on [11] which is govt website which has provided dataset for free use and is absolutely authentic.

Then next thing we did was to clean the dataset and remove unwanted columns from dataset for faster computation.

Here, we use the dataset and check the consistency of the dataset by checking the values out of the dataset randomly.

Then we do data visualization for better understanding of data by the use of various plots, graph and heatmaps.

All this graphs and plots gets us an insight into huge datasets easily.

Machine Learning:

We build several detection models for different states of corona discharge by applying four types of machine learning algorithms to extract the information characteristics of visible images. The four types of machine-learning algorithms are SVM, KNN, SLP, and DT algorithms. SVM is a generalized linear classifier for binary classification of data using supervised learning and kernel methods. It can be used to classify data nonlinearly, and is one of the common kernel-learning methods. The KNN classification algorithm is one of the simplest methods in data mining classification technology. The core idea of KNN is that if the majority of the k-most adjacent samples of a certain sample X in the feature space belong to a certain category, then sample X also belongs to this category and is assigned to the characteristics of the samples in this category. SLP is a type of simple, one-layer, feed-forward artificial neural network. DT is a method to approximate the value of the discrete function. After data are processed, readable rules in the form of a DT are generated by an induction algorithm. Then, when a new data point comes in, it is classified by following the tree structure from top to bottom. All of our algorithms would output one single class per image for prediction.

Step 3: Computing Accuracy

In this step we compute accuracy of all the algorithms by checking the four algorithms mentioned here: Logistic Regression, KNN, Random Forest Classifier, Decision tree Algorithm , we selected these algorithms on the basis of their qualities of regression & classification.

Step 4: Predicting Covid +ve or -ve

In the last step, all we need to do is plot a graph of accuracy of all the algorithms and use the algorithm with best accuracy to predict whether a person has corona or not.

We take input of 5 symptoms in binary values and using our predictor we predict the person is positive or negative on the basis of these 5 symptoms.

6. RESULTS AND SCREEEN SHOTS:

sample Dataset

	Sno	Date	Time	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Deaths	Confirmed
0	1	2020-01-30	6:00 PM	Kerala	1	0	0	0	1
1	2	2020-01-31	6:00 PM	Kerala	1	0	0	0	1
2	3	2020-02-01	6:00 PM	Kerala	2	0	0	0	2
3	4	2020-02-02	6:00 PM	Kerala	3	0	0	0	3
4	5	2020-02-03	6:00 PM	Kerala	3	0	0	0	3
...	...	...	...	...	...	...	...	...	...
18105	18106	2021-08-11	8:00 AM	Telangana	-	-	638410	3831	650353
18106	18107	2021-08-11	8:00 AM	Tripura	-	-	77811	773	80660
18107	18108	2021-08-11	8:00 AM	Uttarakhand	-	-	334650	7368	342462
18108	18109	2021-08-11	8:00 AM	Uttar Pradesh	-	-	1685492	22775	1708812
18109	18110	2021-08-11	8:00 AM	West Bengal	-	-	1506532	18252	1534999

18110 rows × 9 columns

```
Sno                          18110 non-null int64
Date                         18110 non-null object
Time                         18110 non-null object
State/UnionTerritory         18110 non-null object
ConfirmedIndianNational      18110 non-null object
ConfirmedForeignNational     18110 non-null object
Cured                        18110 non-null int64
Deaths                       18110 non-null int64
Confirmed                    18110 non-null int64
dtypes: int64(4), object(5)
memory usage: 1.2+ MB
```

Fig:- 1. attributes of the dataset

	Sno	Cured	Deaths	Confirmed
count	18110.000000	1.811000e+04	18110.000000	1.811000e+04
mean	9055.500000	2.786375e+05	4052.402264	3.010314e+05
std	5228.051023	6.148909e+05	10919.076411	6.561489e+05
min	1.000000	0.000000e+00	0.000000	0.000000e+00
25%	4528.250000	3.360250e+03	32.000000	4.376750e+03
50%	9055.500000	3.336400e+04	588.000000	3.977350e+04
75%	13582.750000	2.788698e+05	3643.750000	3.001498e+05
max	18110.000000	6.159676e+06	134201.000000	6.363442e+06

Fig :- 2 statical Analysis of data

	Sno	Date	State/UnionTerritory	Cured	Deaths	Confirmed
0	1	30-01-2020	Kerala	0	0	1
1	2	31-01-2020	Kerala	0	0	1
2	3	01-02-2020	Kerala	0	0	2
3	4	02-02-2020	Kerala	0	0	3
4	5	03-02-2020	Kerala	0	0	3
...	...	...	...	...	...	...
18105	18106	11-08-2021	Telangana	638410	3831	650353
18106	18107	11-08-2021	Tripura	77811	773	80660
18107	18108	11-08-2021	Uttarakhand	334650	7368	342462
18108	18109	11-08-2021	Uttar Pradesh	1685492	22775	1708812
18109	18110	11-08-2021	West Bengal	1506532	18252	1534999

Fig :- 3. confirmed death cases state wise

State/UnionTerritory	Confirmed	Cured	Deaths
Andaman and Nicobar Islands	7548	7412	129
Andhra Pradesh	1985182	1952736	13564
Arunachal Pradesh	50605	47821	248
Assam	576149	559684	5420
Bihar	725279	715352	9646
Cases being reassigned to states	9265	0	0
Chandigarh	61992	61150	811
Chhattisgarh	1003356	988189	13544
Dadra and Nagar Haveli and Daman and Diu	10654	10646	4
Daman & Diu	2	0	0
Delhi	1436852	1411280	25068
Goa	172085	167978	3164
Gujarat	825085	814802	10077
Haryana	770114	759790	9652
Himachal Pradesh	208616	202761	3537

Fig:- 4. statical analysis stat wise

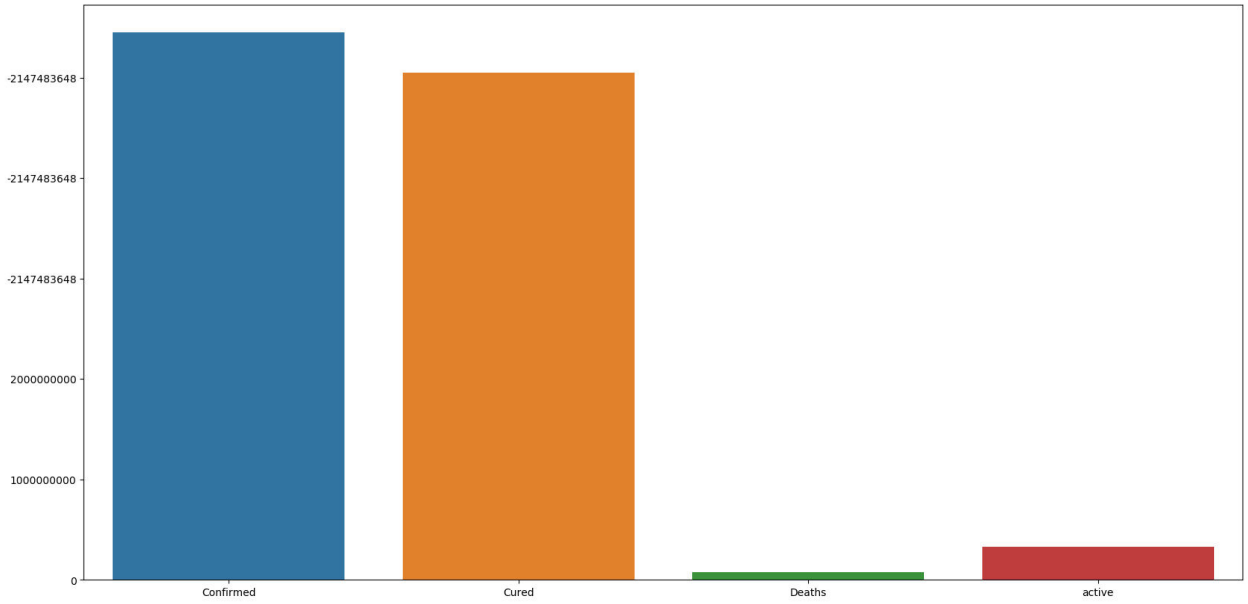

Fig :-5. statical analysis graph

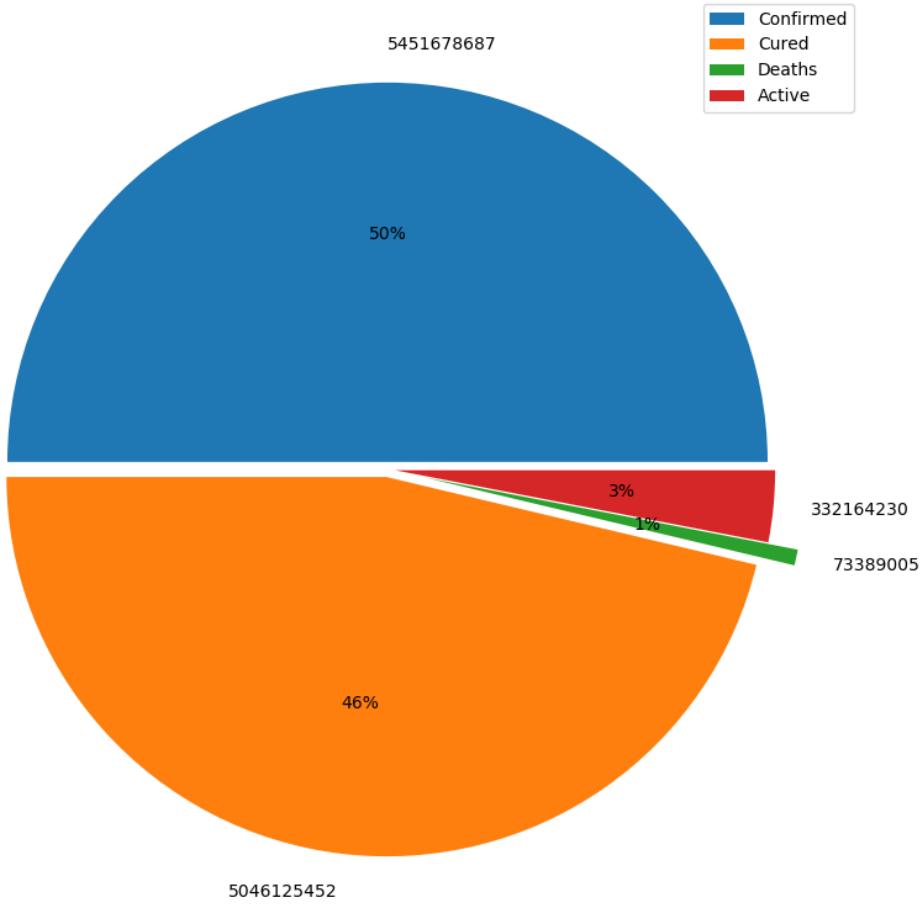

Fig: - 6 statical analysis pie graph conferment and death cases

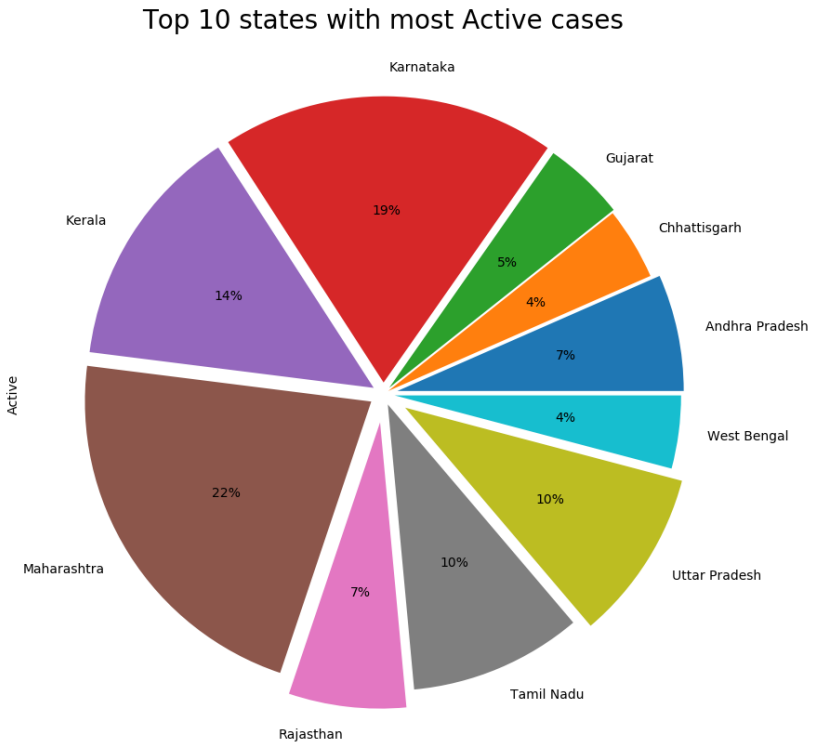

Fig :-7 Top data active cases state wise analysis

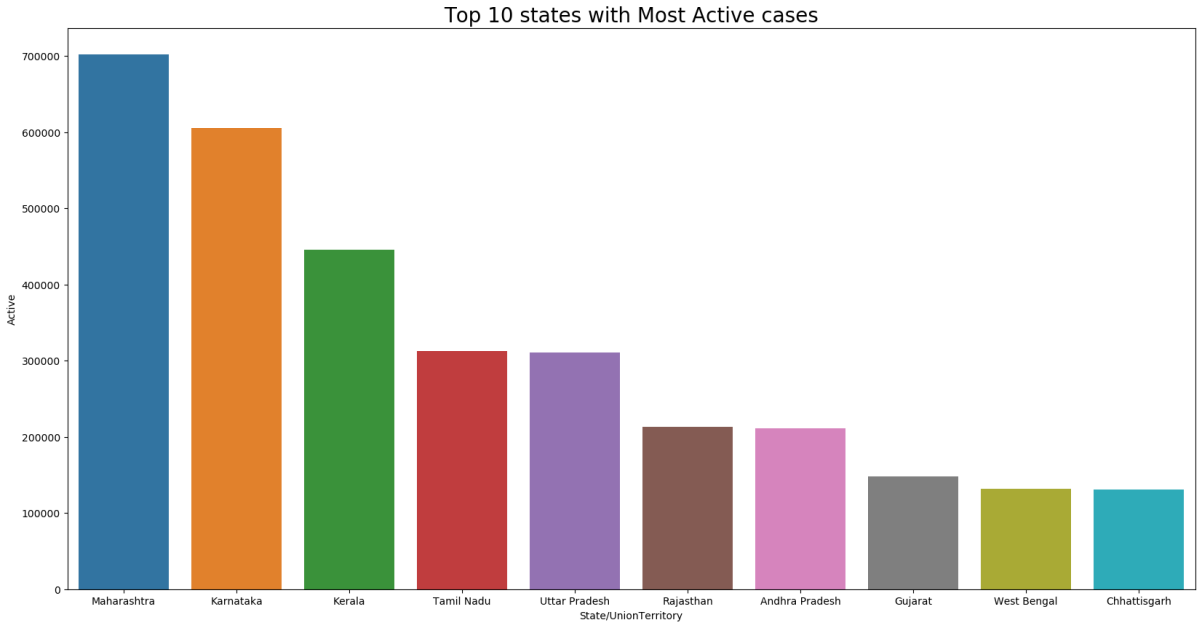

Fig :- 8 top 10 most active cases

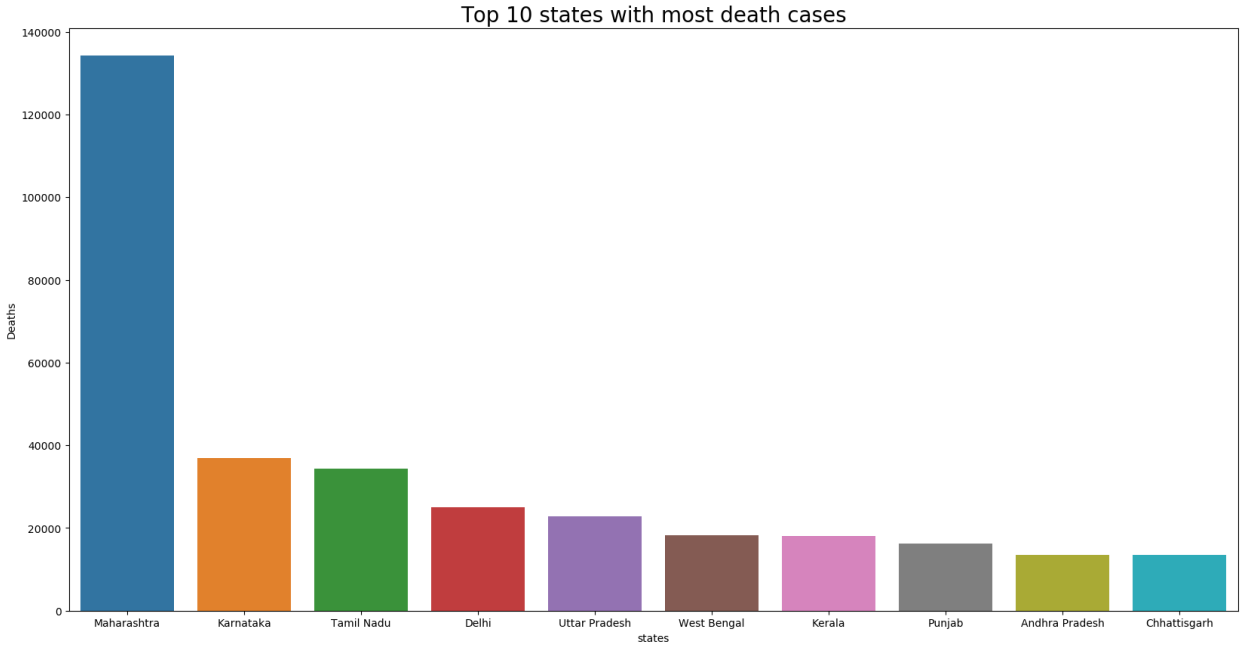

Fig:- 9 top 10 death cases state wise

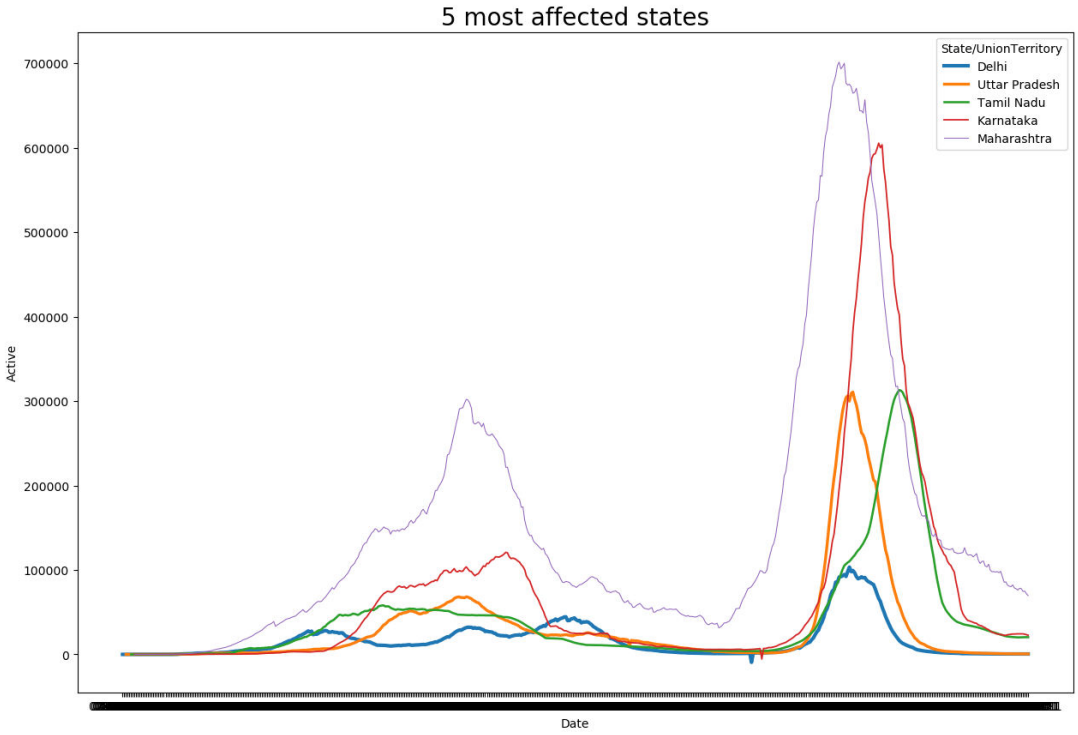

Fig: 10 top 5 most effected cases

7. CONCLUSION AND FUTURE SCOPE

CONCLUSION

The Covid - 19 Pandemic is a huge struggle for allof us. The project we are making will seek to findthe answers to the most pertinent questions as towhat is it that makes the covid 19 such a tragedyand what all people are the ones who are most affected by it. It will seek to find the appropriateresponse which can be mounted by the authoritiesconcerned and we can reach to a place of properdiscussion about the problem and solve it in thebest possible manner out there. It will also lead toa solution to any medical condition we might encounter later on in our lives where we can applydata sciences for medical diagnostics. This projectsaves on the already limited resources that Indiahave and prevents the spread as people can use itto get an idea that they should go and get tested .It also helps unhealthy and infected people toisolate themselves. Using this system we caneffectivelyandefficientlymitigatetheburdenonour healthcare system which is completelystressed out.

The term "feature scope" refers to the extent or range of functionality that a specific feature or component of a software project is expected to encompass. It defines the boundaries and capabilities of a particular feature within a larger software application or system. Understanding and defining the feature scope is crucial in software development.

By clearly defining the feature scope, you can effectively manage expectations, reduce misunderstandings, and increase the likelihood of successful feature development within your software project.

FUTURE SCOPE

Corona discharge only forms when the electric field (potential gradient) at the surface of the conductor exceeds a critical value, the dielectric strength or disruptive potential gradient of the fluid. In air at atmospheric pressure, it is roughly 30 kilovolts per centimeter, but this decreases with pressure, so corona is more of a problem at high altitudes. A corona discharge is an electrical discharge possible because of the ionization of air surrounding a conductor that is electrically charged. The corona treatment is frequently used for polypropylene, PVC, PET, polyethylene, metalized surfaces, paper, and paperboard stock. Electric cables, automotive components, 3D parts, medical devices, pipes & tubes, board &foam,domestic appliances, extruded profiles are some components that are processed with corona.

8.REFERENCES

[1] D. T. Markus and T. S. Lee, “A method of direct corona imaging on a dielectric transparency,” IEEE Trans. Ind. Appl., vol. 32, no. 4, pp. 832–836, Jul. 1996.

[2] Y. Arao, T. Fujii, and M. Yatsuzuka, “Image evaluation of atmospheric corona discharge applied for road tunnel high-velocity electrostatic precipitator,” IEEE Trans. Plasma Sci., vol. 33, no. 2, pp. 312–313, Apr. 2005.

[3] P. H. Ceccato, O. Guaitella, M. Le Gloahec, and A. Rousseau, “Timeresolved nanosecond imaging of the propagation of a corona-like plasma discharge in water at positive applied voltage polarity,” J. Phys. D: Appl. Phys., vol. 43, no. 17, 2010, Art. no. 175202.

[4] S. N. Abolmasov, K. Tachibana, and T. Shirafuji, “Mechanisms of pattern formation in dielectric barrier discharges,” IEEE Trans. Plasma Sci., vol. 39, no. 11, pp. 2090–2091, Nov. 2011.

[5] M. Liao, J. Huang, G. Ge, X. Duan, Z. Huang, and J. Zou, "Vacuum arc commutation characteristics of the DC microgrid hybrid circuit breakers," IEEE Trans. Plasma Sci., vol. 45, no. 8, pp. 2172–2178, Aug. 2017.

[6] D. B.Watson, S. K. Kho, K. A. Samuels, L. Ma, and J. B. Chiu, "Impulse flashover trajectory in air in nonuniform fields," IEEE Trans. Electr. Insul., vol. 28, no. 2, pp. 200–208, Apr. 1993.

[7] D. Amarasinghe, U. Sonnadara, M. Berg, and V. Cooray, "Correlation between brightness and channel currents of electrical discharges," IEEE Trans. Dielectrics Electr. Insul., vol. 14, no. 5, pp. 1154–1160, Oct. 2007.

[8] J. M. K. MacAlpine, D. H. Qiu, and Z. Y. Li, "An analysis of spark paths in air using 3-dimensional image processing," IEEE Trans. Dielectrics Electr. Insul., vol. 6, no. 3, pp. 331–336, Jun. 1999.

[9] D. H. Qiu, J. M. K. MacAlpine, and Z. Y. Li, "An incremental analysis of spark paths in air using 3-dimensional image processing," IEEE Trans. Dielectrics Electr. Insul., vol. 7, no. 6, pp. 758–763, Dec. 2000.

[10] L. Kebbabi and A. Beroual, "Fractal analysis of creeping discharge patterns propagating at solid/liquid interfaces: Influence of the nature and geometry of solid insulators," J. Phys. D: Appl. Phys., vol. 39, no. 1, pp. 177–183, Jan. 2006.