How to Cite:
Thorat, A. ., Patankar, A., & Doke, P. (2022). Use of supervised machine learning algorithms for state wise COVID-19 forecasting. *International Journal of Health Sciences*, *6*(S5), 6432–6443. https://doi.org/10.53730/ijhs.v6nS5.10131

Use of supervised machine learning algorithmsfor state wise COVID-19 forecasting

Abhishek Thorat
Bachelor's Degree in Electronics and Telecommunication Engineering, and Master's Degree Student at Alard College of Engineering, Pune, Savitribai Phule Pune University.
Corresponding author email : thoratabhishek735@gmail.com

Dr. Abhijit Patankar
PhD in Computer Science and Engineering, Masters and Bachelor's Degree in Computer Engineering, and Associate Professor at Department of Information Technology in Dr. D. Y. Patil Institute of Technology, Akurdi, Pune, Savitribai Phule Pune University.
Email: ajpatankar@dypcoeakurdi.ac.in

Prof. Priyadarshani Doke
Master's Degree in Computer Engineering, and Assistant professor at Department of Computer Engineering in Alard College of Engineering and Management, Pune, Savitribai Phule Pune University.
Email: priyadoke1@gmail.com

***Abstract*---** The COVID Techniques for predicting have demonstrated their value in anticipating perioperative outcomes for the objective of improve future decision-making activities. The designs have existed for a long time. Utilized in a large number of possible uses where unfavorable variables for a danger required be identifying and prioritizing. To take care of forecasting challenges, a large number of prediction approaches are widely utilized. This study demonstrates the model's ability to estimate how many patients will be COVID-19 is a virus that affects people. Currently assumed to be a possibility danger to the human race. In this case, study; four conventional forecasting models were put to use to foresee the hazardous elements of COVID-19: LR, least LASSO, SVM, and ES.Each of the models makes 3 sorts of forecasts for the next ten days: the no. of freshly infected cases, the no.of newly infected cases, the no. of newly infected cases, the no. of newly fatalities, and the no. of recoveries. The study's findings show that using these strategies in the present situation COVID-19 pandemic scenario is a promising mechanism. The findings reveal that the ES outperforms all other models, LR is then followed by LR and LASSO, which excel at forecasting additional cases have

been confirmed death rates, and recovery rates, respectively, while SVM performs badly in all given the scenarios for prediction provided information.

***Keywords*---**COVID-19, Machine Learning, LSTM, Health.

1 Introduction

Over the last decade, ML has established itself as a significant topic of research by tackling plenty of extremely complicated and complex real-world issues. Healthcare, autonomous vehicles (AV), business applications, NLP, intelligent robotics, gaming, climate modeling, voice, and image processing were among the applications covered. In contrast to traditional algorithms, which follow programming instructions based on decision statements like if-else, ML algorithms & learning is often dependent on trial and error.

Various regression and NN models have a wide range of applicability when it comes to forecasting the future health of individuals with a given ailment. There has been numerous research employing ML approaches to forecast various diseases, such as coronary artery disease, cardiovascular disease prediction, and breast cancer prediction. The research is centered on COVID-19 confirmed live forecasting cases, as well as the forecasting of a COVID-19 outbreak and early reaction. COVID-19 is currently posing a significant danger to human life around the world. The virus was originally discovered in late 2019 in the Chinese city of Wuhan, when a large no. of patients suffered pneumonia-like symptoms. Every day, thousands of new persons are reported to be positive from all over the world. Close physical contact, respiratory droplets, and touching contaminated surfaces are the most common ways for the virus to spread.

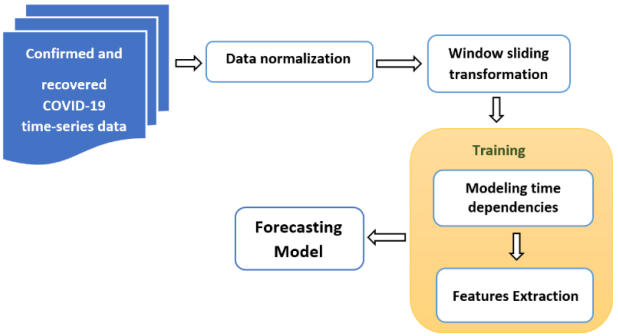


Figure 1. Flow chart of COVID-19 cases

2 Materials and Methods

Literature Survey

FURQAN RUSTAM et.al [1] The study findings show that using these strategies in the present situation COVID-19 pandemic scenario is a promising mechanism. The results show that the ES outperforms all other models, followed by LR and LASSO, which excel at forecasting life, new incidences rates, and recovery rates, respectively, When SVM is in action badly in all given the scenarios for prediction provided information.

Mujeeb Ur Rehman et.al [2] The proposed diagnostic procedure considers a variety of factors symptoms, such as B. flu symptoms, sore throat, immune status, diarrhea, voice type, body temperature, joint pain, dry cough, vomiting, difficulty breathing, headache and chest discomfort. The proposed method can predict the likelihood of infection caused by the COVID19 virus based on these symptoms, which are modeled as features of ML .The results suggest that the proposed technique can predict the presence of COVID19 more than 97 percent accurately. [5][13]

Yanping Zhang, zhangyp et.al [3] The 2019 new corona virus diseases outbreak that began in Wuhan, Hubei Province, China, has quickly expanded across the country. The Chinese IDIS is a database that keeps track of was used to collect all COVID-19 cases reported through February 11, 2020.

Dr. Vakula Rani J &, Aishwarya Jakka2 [4] The COVID19 pandemic has impacted everyone's emotional, physical and financial well-being, impacted the economy and changed the way people live. The severity and contagion of the epidemic have put pressure on several of the world's fastest growing economies. Due to the increasing diversity of cases and the associated impact on healthcare providers and government, estimating the no. of people infected cases of COVID19 could be helpful in determining future hospital resource needs.

Saksham Gera et.al [6] The COVID – 19 (Novel Corona Virus) Pandemic has spread the globe, wreaking damage on the environment. People’s lives. In India, the study looks at the current pattern or trend in COVID – 19 transmissions. Furthermore, a ML-based forecasting system has demonstrated its value in improving management skills in the course of action that will be taken in the future. This study illustrates the capacity of various Models using ML to forecast the no. of patients who will be admitted to the hospital would be afflicted by nCov, a global threat.

Ovi Sarkar et.al [7] SARSCoV2 (Ncorona virus) research analyzed the current status of the world's coronavirus pandemic and Bangladesh, as well as its impact and future prospects. For the purpose of reproduce the confirmed, retrieved and transient situations; the author used various data representations and ML calculations. The authors of the study hope it will help scientists, researchers and ordinary people predict and analyze the consequences of the pandemic.

Shreyansh Chordia & Yogini Pawar et.al [8] India is one of the hardest hit countries in the world. It is therefore crucial to study India's trends and apply what we have learned to gauge the future course of events. In addition to a general analysis of trends in India, this report focuses on the five hardest-hit states in the country: Maharashtra, Andhra Pradesh, Tamil Nadu, Karnataka and Uttar Pradesh.

Narayana Darapaneni et.al [9] In their paper, the authors examine the spread of COVID19 in India and the 3 most affected Indian states as of August 29, 2020 and create a predictive model to estimate the spread of COVID19 within the next few months. They analyzed time-series data from India to use the Susceptible Infectious Removed (SIR) model and the Fb Prophet for India and the 3 hardest-hit states, a model for predicting the future infection peak and date is being developed.

Ashish U Mandayam1 et.al [10] Predictive analytics has grown in popularity a necessary component of future predictions as ML science advances. As the authors deal with the COVID19 pandemic, it would be beneficial to estimate the amount of positive cases that will occur in the future for the purpose of plan ahead improve measures and controls. In the COVID19 time series dataset, they used two supervised learning models to forecast the future. A comparison of LR vs. SVR is used to examine prediction performance. Since the data were nearly linear, they used these two models.

Senthil kumar Mohan et.al [11] The Covid19 epidemic has established itself as one of the most worrying global health disasters of the 21st century, underscoring the critical need for robust prognostic approaches for disease detection, mitigation and prevention, among others.

Abdelkader Dairi et.al [12] The impact of ML, a key component of artificial intelligence, on previous epidemics provides a new avenue for combating the unique Coronavirus pandemic. Accurate short-term forecasting of COVID-19 spread is critical for bettering management of the hospital overpopulation problem and allowing proper resource optimization.[15]

Proposed System

The study is on COVID-19 predictions, which is a new corona virus. The COVID-19 has proven to be a clear and present danger to human life. This work seeks to do future mortality forecasting to make a contribution to the control of the pandemic crisis. We have used LSTM algorithm for data prediction purpose. The system takes continuous real time data from the data sets and LSTM algorithm performs the prediction of data. Results are visible in Graph format along with the numerical data in the table. By visualizing the above data the hospitals and medical organizations can easily arrange the vaccines and medical equipments for future crisis

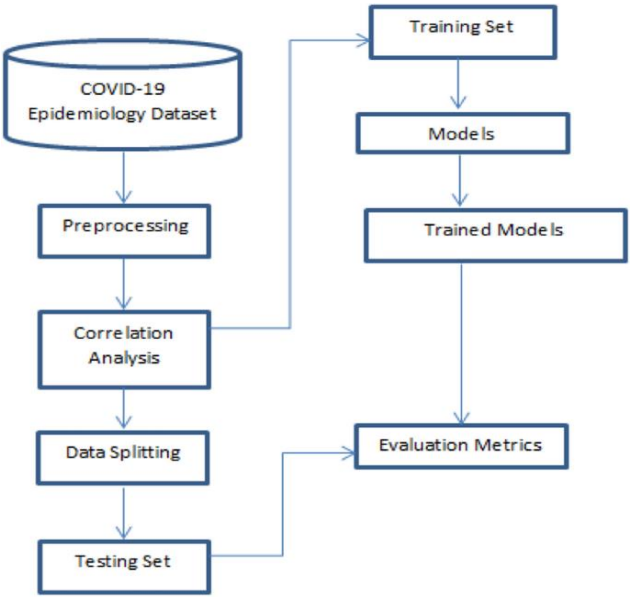


Figure 2. Architecture diagram of proposed system

Methodology

Data preprocessing:
Preprocessing is the process of transforming unprocessed data into usable information. a format suitable for ML. From a data scientist, more precise results can be obtained. a ML model when the data is structured and clean. Data formatting, cleansing, and sampling are all a portion of process.

Data formatting:-
When data is gathered from a wide range of sources by different people, it is referred to as multi-source data persons, data formatting becomes increasingly important. A data scientist's initial responsibility is to standardize record formats. A professional examines the variables that stand for each characteristic to see if they are all recorded in the same way. Variables include product and service names, pricing, date formats, and addresses. The notion of data consistency also applies to numeric ranges as characteristics.

Data sampling: -
For analysis, large datasets necessitate greater time and computing resources. When a dataset is too huge, data sampling is the best option. This strategy is used by a data scientist to choose a smaller but representative data sample for analysis construct and test models faster while still producing reliable results.

Featuraization:-
Featuraization is a technique for converting text, graph, and time-series data into numerical vectors. Featuraization is not to be confused with feature engineering. Feature engineering is simply the process of altering numerical characteristics in such a way that ML models can function properly.

Data splitting:-
ML datasets should be divided into three subsets: training, test, and validation sets.

Training set: -
To learn, a data analyst uses a learning algorithm and define a model's ideal parameters, which it must learn from data.

Test set: -
A test set is required for evaluating the trained model's generalization capacity. The latter refers to a model's ability to spot patterns in newly discovered information after it has been trained on it. To avoid model over fitting, which leads to the inability to generalize mentioned earlier, it's critical to use different subsets for training and testing.

Modeling:-
During this step, a data scientist trains a wide range of models in order to determine which one makes the most accurate predictions. Training as a model: - A data scientist can begin model training after preprocessing the acquired data and dividing it into three subsets. The algorithm is "fed" with training data during this process. An algorithm will evaluate data and generate a model capable of detecting a target value (attribute) in new data — the answer you're looking for using predictive analysis. Model development is the goal of model training.

LSTM

A recurring NN is a kind of long-term memory. The result of in the phase, the prior step is used as input. Step in RNN. Hochreiter and Schmidhuber developed the LSTM. They cannot predict words stored in long-term memory, but can make more precise forecasts based on current data. RNN does not perform efficiently as the gap length increases. By default, the LSTM can store information for a long time. It is used for processing, forecasting and classifying time series data.

Structure of LSTM:

The LSTM has a chain structure that consists of four NN and several memory blocks known as cells.

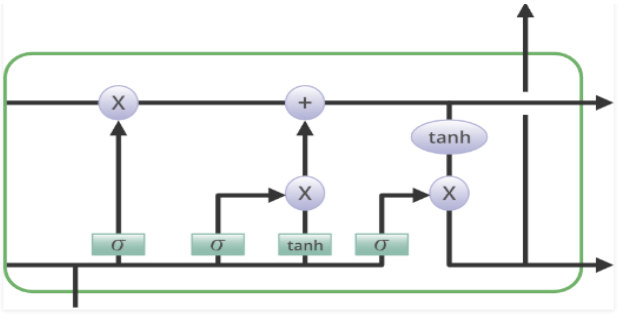


Figure 3. *LSTM Structure*

LSTM Working

Because significant occurrences in a time series may have unpredictable lags, LSTM networks are well-suited to categorizing, analyzing, and making predictions based on time series data. LSTMs were developed to address the issue of vanishing gradients. Might occur when training traditional RNNs. RNNs that learn to predict the future based on a series of events varying lengths use LSTM cells. The information transfer rate to the cell is controlled by the LSTM's carefully built gate structure. It has the ability to selectively allow information to pass. A sigmoid function and a point wise multiplication process are among them. The sigmoid layer returns a value between 0 and 1, indicating how many passes each part has gone through.

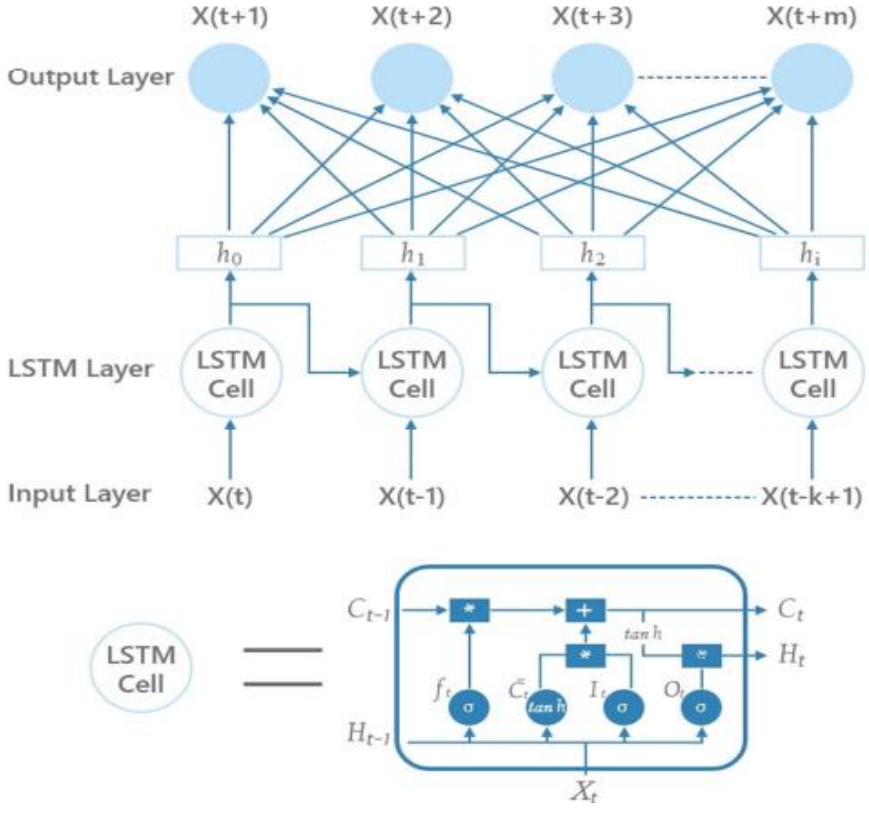


Figure 4: LSTM Working Structure

Where,

X_t = Input Vector	H_t = Current Cell Output
C_{t-1} = Previous Cell Memory	f_t = Forget gate
H_{t-1} = Previous Cell Output	$\bar{C}_t$ = Intermediate Cell state
C_t = Current Cell Memory	I_t = Input gate
	O_t = Output gate

In an LSTM cell, there are three gates: a forget gate, an input gate, and an output gate. This source is used to modify all LSTM cell pictures.

3 Results and Discussions

The home page of our system, Covid-19 future forecasting system, will be the first phase. We will choose a state, such as Maharashtra or India, and our system will produce future covid-19 instances and vaccination status based on that pick.

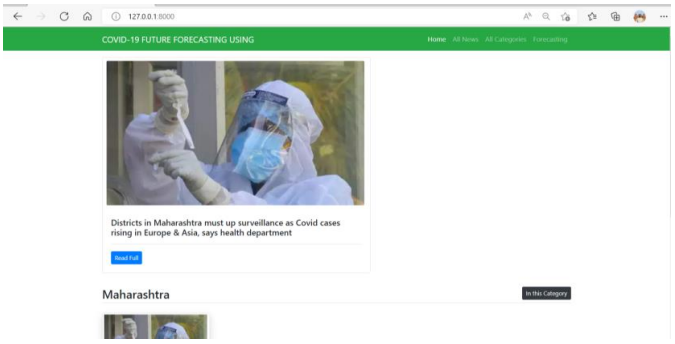
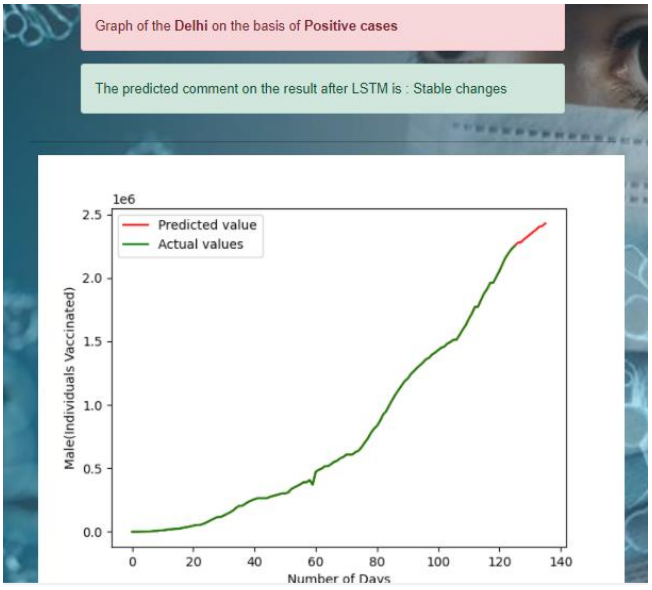


Figure 4. Homepage

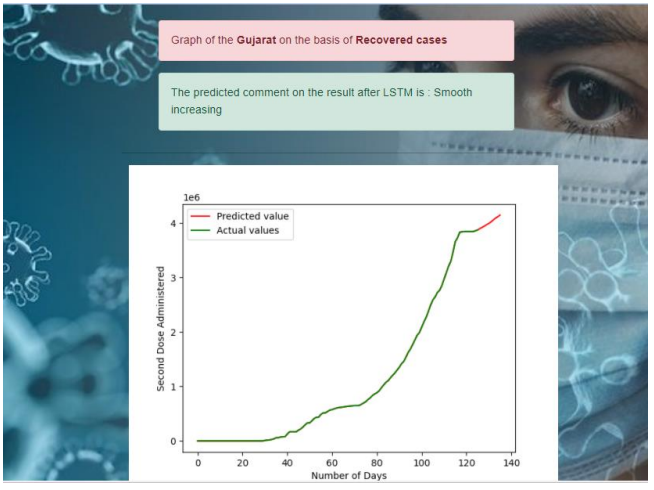
Figure 5. Select state

After selecting the state there will be a dropdown that which kind of prediction the user wants either it is of Upcoming cases, Recovered cases, Death Rates, Vaccine status over the regions



Next predicted day	Predicted result of Positive cases
1	2227935.0
2	2251807.0
3	2277217.0
4	2279825.0
5	2302607.0
6	2322448.0
7	2341196.0
8	2362008.0
9	2379448.0
10	2402674.0

Figure 6. Graph and Table view prediction for state Delhi



Next predicted day	Predicted result of Recovered cases
1	3856708.0
2	3872425.0
3	3898122.0
4	3920047.0
5	3942765.0
6	3968728.0
7	3992404.0
8	4020380.0
9	4056409.0
10	4090143.0

Figure 7. Graph and Table view prediction for state Gujarat.

Confusion Matrix

	Class1	Class2
Class 1	**1336**	**9**
Class2	**4**	**1337**

Figure 8: Confusion Matrix

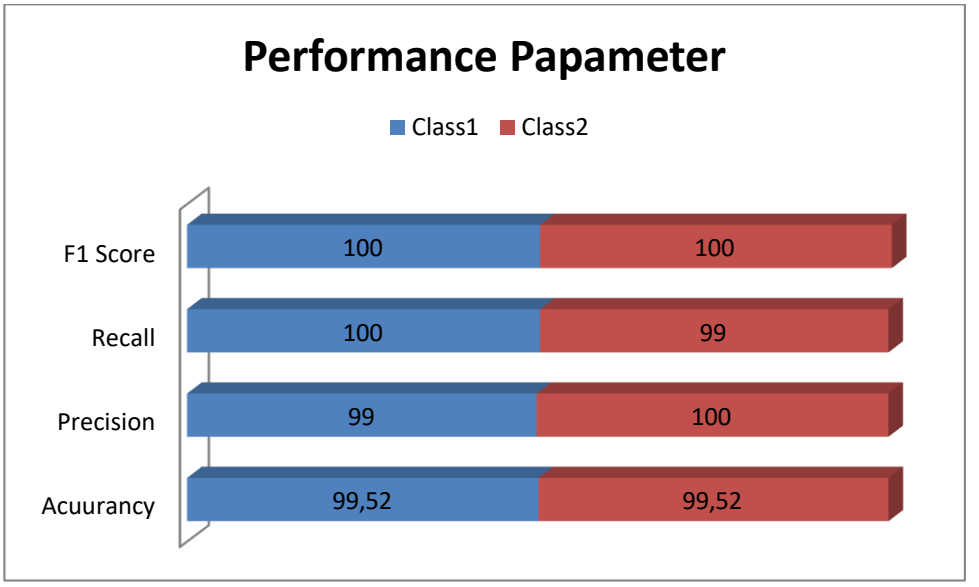


Figure 9: Performance Parameter

The confusion matrix Class 1, Class 2 training modules can be seen in the diagram above. In Class 1, the input photos are 1345, and we achieved accuracy of 99.52 % and precision of 99 % while training the classifier as a train with the supplied input database. Because the 1345 classifier failed to classify 5 photos as an output form of a class 1, recall was reduced to 100%, and F1 score was also reduced to 100%.

In Class 2, the input photos are 1341, and we achieved accuracy of 99.52 % and precision of 100 % while training the classifier as a train with the given input database. As a result of the 1341 classifier failing to detect 5 photos as an output form of a Class 2, recall has been reduced to 99 %, and F1 score has been reduced to 100 %. We can conclude that our system's performance is better with 99.52% after looking at the above performance parameters.

4 Conclusion

The no. of possible positive cases of COVID19 in India for the next 30 days was estimated using an information-based assessment/prediction technique. Using and fitting the curve, the no. of cases retrieved, the daily transient ES in specific cases, and expired cases were also determined. Although this strategy often requires adequate information to support it, this technique can be utilized in early stages of plague transmission to more accurately predict pest transmission indicators for the time being, to provide intermediary control at all levels of offices and the implementation of the strategy offers programmers momentary crises. For mixed boundaries and districts, the expected results from three different numerical models are different. The reasonable influence of the logistic model can be the best of the three models on average this exam will be carefully revised later in the course; next, we intend to study figure theory using the revived dataset and analyzes using the most comprehensive and appropriate ML systems. We

achieved accuracy of 99.52 % using LSTM algorithm. When we compared our proposed work to current schemes, we discovered that our 99.52 percent accuracy was substantially higher than what most other ML algorithms had accomplished.

Acknowledgments
We are grateful to two anonymous reviewers for their valuable comments on the earlier version of this paper.

References

FURQAN RUSTAM 1, AIJAZ AHMAD RESHI 2, (Member, IEEE), ARIF MEHMOOD 3, SALEEM ULLAH 1, BYUNG-WON ON4, WAQAR ASLAM 3, (Member, IEEE), AND GYU SANG CHOI 5 "COVID-19 Future Forecasting Using Supervised Machine Learning Models "Received May 4, 2020, accepted May 13, 2020, date of publication May 25, 2020, date of current version June 10, 2020.

Mujeeb Ur Rehman 1, Arslan Shafique 1, Sohail Khalid 1, Maha Driss 2, 3 and Saeed Rubaiee 4 "Future Forecasting of COVID-19: A Supervised Learning Approach "Sensors 2021, 21, 3322. https:// doi.org/10.3390/s21103322 Received: 5 April 2021 Accepted: 6 May 2021 Published: 11 May 2021

Yanping Zhang, zhangyp "The Epidemiological Characteristics of an Outbreak of 2019 Novel Corona virus Diseases (COVID-19) — China, 2020 "Submitted: February 14, 2020; accepted: February 14, 2020

Dr. Vakula Rani J#1 & Aishwarya Jakka#2 "Forecasting COVID-19 cases in India Using Machine Learning Models "Authorized licensed use limited to: IEEE Xplore. Downloaded on January 24, 2022 at 11:34:26 UTC from IEEE Explore.

Saud Shaikh, 2Jaini Gala, 1Aishita Jain, 1Sunny Advani, 1Sagar Jaidhara, 1Dr. Mani Roja Edinburgh "Analysis and Prediction of COVID-19 using Regression Models and Time Series Forecasting "2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) | 978-1-6654-1451-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/Confluence51648.2021.9377137

Saksham Gera, Mr Mridul, Mr. Kireet Joshi "Regression Analysis And Future Forecasting of COVID-19 Using Machine Learning's Algorithm. "2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) | 978-1-6654-1451-7/20/$31.00 ©2021

Ovi Sarkar, Md Faysal Ahamed , Pallab Chowdhury "Forecasting & Severity Analysis of COVID-19 Using Machine Learning Approach with Advanced Data Visualization "2020 23rd International Conference on Computer and Information Technology (ICCIT), 19-21 December, 2020

Shreyansh Chordia & Yogini Pawar "Analyzing and Forecasting COVID-19 Outbreak in India "2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) | 978-1-6654-1451-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/Confluence51648.2021.937711

Narayana Darapaneni , Praphul Jain , Rohit Khattar , Manish Chawla, Rijy Vaish "Analysis and Prediction of COVID-19 Pandemic in India "2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) |978-1-7281-8337-4/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICACCCN51052.2020.9362817

Ashish U Mandayam1 , Rakshith.A.C2, Siddesha S3, S K Niranjan4, "Prediction

of Covid-19 pandemic based on Regression “Authorized licensed use limited to: IEEE Xplore. Downloaded on January 24, 2022 at 11:39:32 UTC from IEEE Explore.

Senthilkumar Mohan 1 John A 2 Ahed Abugabah 3 Adimoolam M4 Shubham Kumar Singh5 Ali kashif Bashir 6, 7 Louis Sanzogni.” An approach to forecast impact of Covid-19 using supervised machine learning model” Received: 3 November 2020 Revised: 15 February 2021 accepted: 22 February 2021

Abdelkader Dairi, Fouzi Harrou, Abdelhafid Zeroual, Mohamad Mazen Hittawe, Ying Sun, Comparative study of machine learning methods for COVID-19 transmission forecasting, Journal of Biomedical Informatics, Volume 118,2021.

Widjaja, G. (2021). Impact of human resource management on health workers during pandemics COVID-19: systematic review. International Journal of Health & Medical Sciences, 4(1), 61-68. https://doi.org/10.31295/ijhms.v4n1.850

Mujeeb Ur Rehman 1, Arslan Shafique 1, Sohail Khalid 1, Maha Driss 2, 3 and Saeed Rubaiee 4.” Future Forecasting of COVID-19: A Supervised Learning Approach” https:// doi.org/10.3390/s21103322 Received: 5 April 2021.Accepted: 6 May 2021 published: 11 May 2021

Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2021). The COVID-19 pandemic. International Journal of Health Sciences, 5(2), vi-ix. https://doi.org/10.53730/ijhs.v5n2.2937

G.PRATHIBA PRIYADARSHINI1, G.SRAVANTHI2, D.VENI3, G.ANUSHA4, B.RANI5, D.KEERTHI6.” COVID-19 FUTURE FORECASTING USING SUPERVISED ML” Vol 12, Issue3, 2021

M. Akshitha 1 , R Jegadeesan2 , G. Akshaya 3 P. Akhila c 4 , M.Pavan Kalyan 5 , G.Sindhusha6 . “Covid-19 Future Forecasting Using Supervised Machine Learning Models” IEEE 2021.