2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)
March 17–18, 2021, Amity University Dubai, UAE

2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) | 978-1-6654-2921-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCIKE51210.2021.9410777

# COMPARATIVE INTREPRETATION OF MACHINE LEARNING ALGORITHMS IN PREDICTING THE CARDIOVASCULAR DEATH RATE FOR COVID-19 DATA

D.R. Krithika,

Research Scholar,

Department of Computer Science,

Vels Institute of Science Technology and Advanced Studies,
Chennai, India,
krithikabanu@gmail.com.

Dr.K.Rohini

Associate Professor,

Department of Computer Science,

Vels Institute of Science Technology and Advanced studies,
Chennai, India,
rrohini16@gmail.com.

***Abstract:***

Every year 31% of people die from cardiovascular disease worldwide. The big data analytics technique is very useful to Identify Heart disease and COVID-19. To control the COVID-19 spread around the world and many of the companies adapting this technology and also remote places patient reports doctors view easily to analyze health condition of the patient using IOT based big data. In 2019 COVID-19 (Novel coronavirus Disease) was recognized. COVID-19 signs of CT scan include pleural thickening and vascular enlargement. Nucleic acid detection and epidemiological tracing are using Chest CT scans counteract. To understanding of the disease COVIDE-19 the Researchers are using ML, AI (Artificial Intelligence) and natural language processing. We are using big data analytics to track the spread of this coronavirus. In this paper we discuss about Comparison of Tools in Big data Analytics using machine learning Algorithm.

***Keywords*:** Big data Analytics, ML-Machine Learning, AI-Artificial Intelligence, IOT- Internet of Things.

## I. INTRODUCTION:

Every data is valuable when it gives useful information. So the modification of meaningless data into the format of valuable data presents in big data analytics. One part of regression analysis is Linear regression analysis to find the target variable in the effect of input variable. Linear, Multiple, Logistic, Polynomial are the types of Regression Analysis. Tableau, Python, R, Zoho are few big data analytical tools.

## II. LITERATURE REVIEW:

The paper proposed about coronavirus spread all over the world and huge number of patients infected. The diagnosis purpose needed machine learning algorithm and cloud, it reduces clinician's large effort and also chest computed tomography is used for diagnosis. The learning algorithm used [1]. The paper discusses about the situational information is more important for COVID-19 epidemic researchers and social media publishing information strategies. The regression model was used [2]. This paper discussed about coronavirus caused severe respiratory disorder, lung infections and massive death toll. AI (artificial intelligence) and CT scan plays important role of diagnosis COVID-19 disease [3]. This paper discusses about big data implementation widely spread around the world [4]. In this paper discussed about Blockchain, artificial intelligence, 5G, Internet of things helps to find impact of coronavirus. These technologies useful to prevent the community spread [5]. In this paper discovered AI (artificial intelligent) was analyzed to chest x-ray and implemented Deep Learning [6]. The paper discussed about Integration of Artificial intelligent with CT scan and X-ray, fighting against coronavirus deep learning is the dominant approach [7]. In this paper discussed about COVID-19 affected consumers daily lives and also e-commerce expanded to launch new products [8]. The paper proposed about X-Ray image segmentation in improved marine predators algorithm detection model based [9]. The paper discussed about Hybrid artificial intelligence model used to predict COVID-19. China starting stage of COVID-19 treated traditional epidemic model and then improved susceptible–infected model Implemented next stage only they applied hybrid AI model Correlation Analysis to find

978-1-6654-2921-4/21/$31.00 ©2021 IEEE

Cumulative daily confirmed cases and infection rate [10]. This paper proposed the covid-19 management imposed challenges [11]. This paper proposed Chest X-Ray is strong detection of coronavirus and deep learning used [12]. The paper discussed Machine learning interprets result and resurgences of coronavirus help of cloud [13]. In this paper proposed about 3D chest CT images to find covid-19 and using deep learning [14]. The paper discussed about localization and classification in deep learning of coronavirus. Lung[25] ultrasonography is analyzed in DL Techniques [15]. In this paper discussed coronavirus-2019 health monitoring of wearable sensors [16][24]. This paper discussed about machine learning and IOT in e-healthcare to predict vascular diseases and proposed diagnosis of heart disease applied integrated prediction algorithm [17]. The paper discussed about using IT technology to making prevention activities. Face image capture for color comparison and facial region extracted to diagnose the cardiovascular disease [18]. In this paper discussed about healthcare sector the impact of big data and various tools handling in Hadoop ecosystem [19]. The paper discussed with feasible findings of open research challenges [20]. The paper discusses compared biggest challenges of big data and data analytics and analyzed [21].

## III THE SOURCES OF DATA

The qualitative and quantitative of data literally it gives something. Data will be collected two ways primary and secondary.

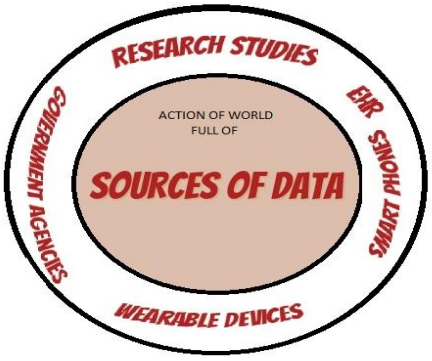


Fig1. Data Sources in various sectors

The action of world is full of data from research studies, electronic health record (EHR), smart phones, wearable devices, government agencies, public data etc. Smart watch to detect and diagnosis heart disease.

## IV ALGORITHM AND TOOLS IN BIGDATA

Machine learning algorithm[24] is categorized by supervised, unsupervised and reinforcement. Linear Regression [22] is best to fit line by minimizing errors. In this paper regression is implemented to best fit of data. Most of the prediction work used in regression analysis and using this regression analysis mathematically predict the values independent variable based upon dependent variable. One dependent and one independent variable must present in Regression analysis. Compared bigdata analytical Tools are Tableau and Python.

## V. TOOLS AND ALGORITHMS COMPARISION

In this paper comparison purpose tableau and python analyzed and implementation part two set of data used. One set of data is covid19 data and another set of data cardiovascular data from UCI machine learning repository. Each data set minimum we can take two variables, dependent and independent variable plots two dimensional scatter plot allows to regression analysis. The straight line of linear regression analysis best fit of data. The next step of data collection dataset data analyzed and wrangling the data. The process of wrangling the data is removing null values.coviddt.isnull(),coviddt.isnull.sum(),,sns.heatmap(coviddt.isnull(), yticklabels == false, cmap="viridis")

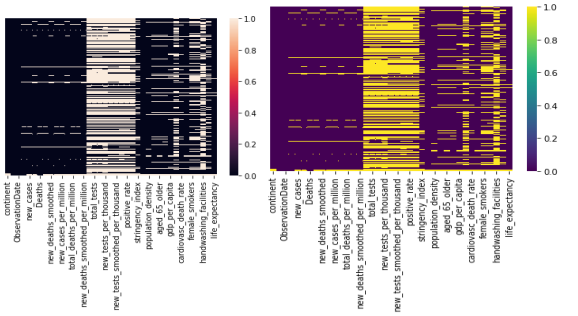

Fig2. Data Preprocessing Fig. 3 Processed Data

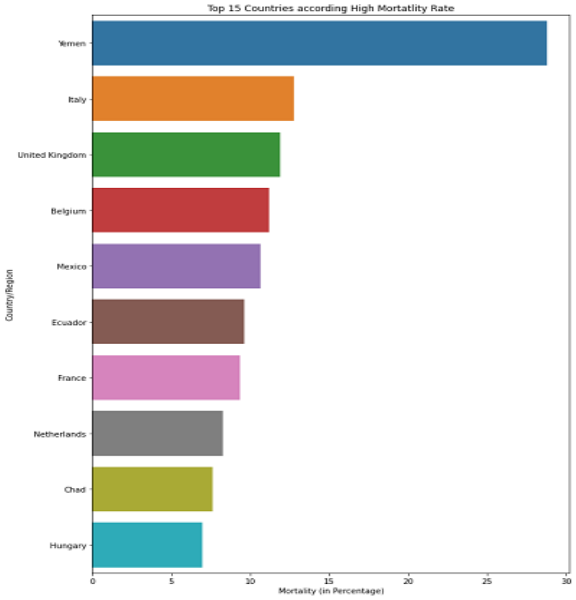


Fig.4 High Mortality Rate Top 15 Countries

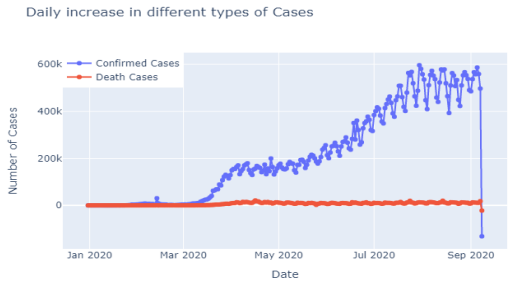


Fig.5 Daily increase of confirmed and Death Cases

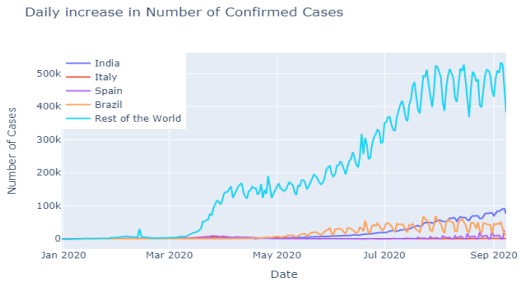


Fig9 Number of Confirmed cases daily increase

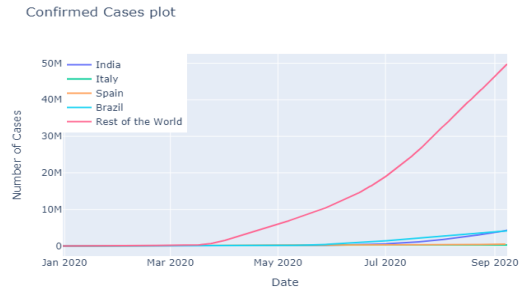


Fig.6 Plots for Confirmed cases

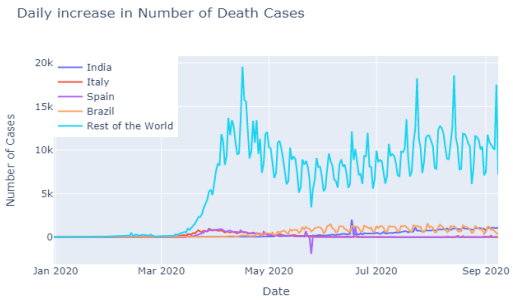


Fig10 Number of death cases daily increase

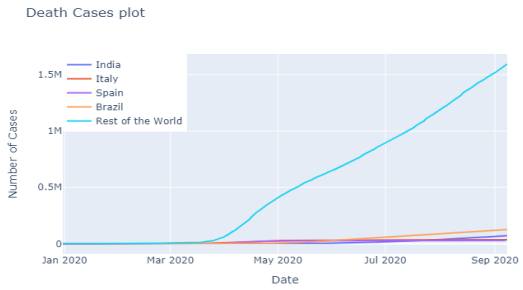


Fig.7 Plots for Death Cases

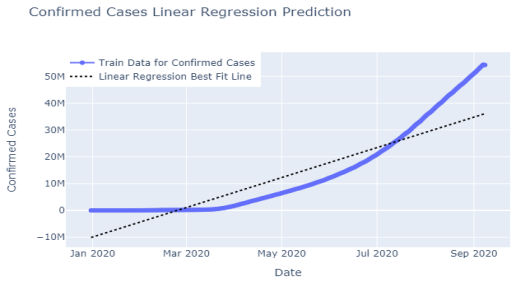


Fig11 Prediction of confirmed cases in Linear Regression

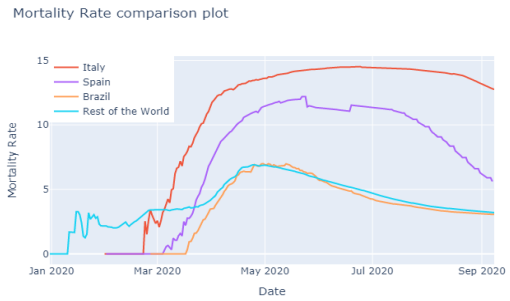


Fig8 Comparison plot for Mortality Rate

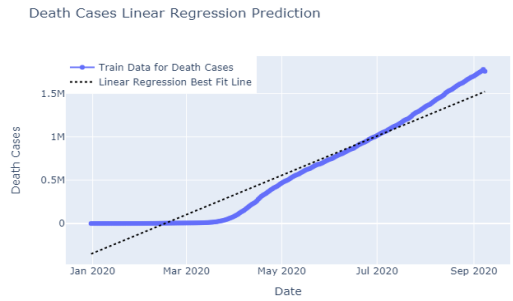


Fig12 Prediction of Death cases in Linear Regression

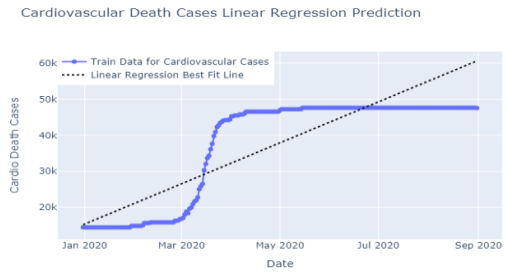


Fig13 Prediction of Cardiovascular Deaths in Linear Regression

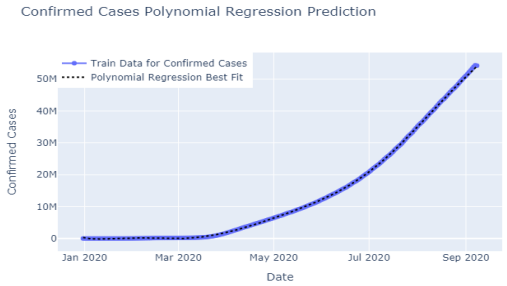


Fig14 Prediction of confirmed cases in Polynomial Regression

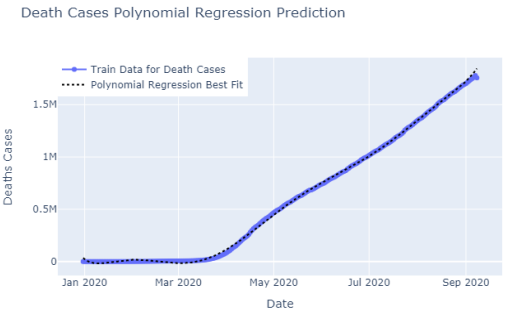


Fig15 Prediction of Death cases in Polynomial Regression

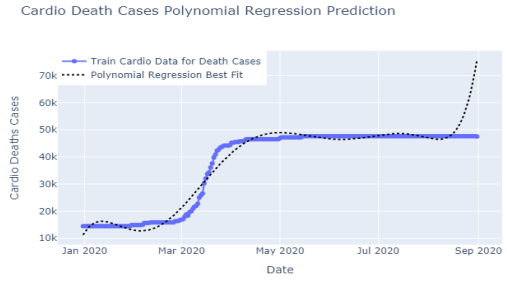


Fig16 Prediction of Cardiovascular death cases in Polinomial Regression

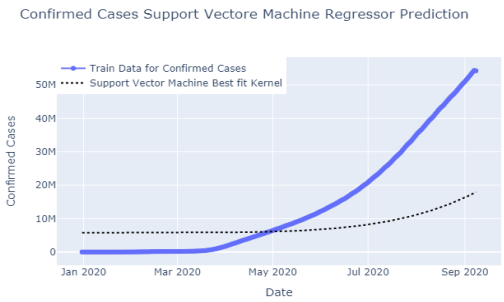


Fig17 Prediction of Confirmed Cases in SVM

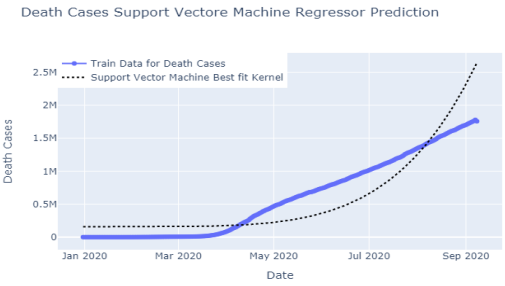


Fig18 Prediction of Death Cases in SVM

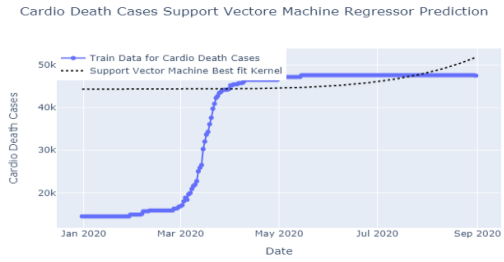


Fig19 Prediction of Cardio Death Cases in SVM

Covid Data Results shows Fig2 to fig19 using python tool. Applied linear Regression Algorithm, Polynomial Regression Algorithm and Support Vector Machine. Confirmed, Death and Cardiovascular Death Cases existing data using to predicted Future cases. Table1 is the Result of Confirmed cases prediction in linear, Polynomial and SVM Algorithms and Table2 is the Result of Cardiovascular Death cases prediction in linear polynomial and SVM Algorithms. The prediction value is nearer to actual value in linear Regression.

| Prediction | | | | |
|---|---|---|---|---|
| | **Dates** | **Linear** | **Polynomial** | **SVM** |
| 0 | 2020-09-09 | 36243266.759389 | 54465320.092138 | 18156866.690962 |
| 1 | 2020-09-10 | 36426403.314953 | 54944137.869115 | 18403911.824438 |
| 2 | 2020-09-11 | 36609539.870516 | 55428373.553113 | 18654878.217346 |
| 3 | 2020-09-12 | 36792676.426080 | 55919265.825033 | 18909812.366113 |
| 4 | 2020-09-13 | 36975812.981643 | 56418164.585828 | 19168761.133273 |
| 5 | 2020-09-14 | 37158949.537207 | 56926536.412008 | 19431771.748910 |
| 6 | 2020-09-15 | 37342086.092770 | 57445970.174627 | 19698891.812088 |
| 7 | 2020-09-16 | 37525222.648333 | 57978182.824842 | 19970169.292289 |
| 8 | 2020-09-17 | 37708359.203897 | 58525025.348592 | 20245652.530851 |
| 9 | 2020-09-18 | 37891495.759460 | 59088488.893682 | 20525390.242399 |
| 10 | 2020-09-19 | 38074632.315024 | 59670711.071844 | 20809431.516284 |
| 11 | 2020-09-20 | 38257768.870587 | 60273982.438863 | 21097825.818018 |
| 12 | 2020-09-21 | 38440905.426151 | 60900753.155836 | 21390622.990709 |
| 13 | 2020-09-22 | 38624041.981714 | 61553639.834547 | 21687873.256497 |
| 14 | 2020-09-23 | 38807178.537278 | 62235432.569631 | 21989627.217992 |

Table1. Confirmed cases prediction

| Prediction | | | | |
|---|---|---|---|---|
| | **Dates** | **Linear** | **Polynomial** | **SVM** |
| 0 | 2020-09-01 | 60770.394957 | 79382.415529 | 51959.917539 |
| 1 | 2020-09-02 | 60956.335185 | 83572.287941 | 52116.549841 |
| 2 | 2020-09-03 | 61142.275414 | 88154.484637 | 52275.749821 |
| 3 | 2020-09-04 | 61328.215642 | 93155.314999 | 52437.548921 |
| 4 | 2020-09-05 | 61514.155870 | 98602.346981 | 52601.978835 |
| 5 | 2020-09-06 | 61700.096098 | 104524.449051 | 52769.071517 |
| 6 | 2020-09-07 | 61886.036326 | 110951.833074 | 52938.859178 |
| 7 | 2020-09-08 | 62071.976554 | 117916.098118 | 53111.374285 |
| 8 | 2020-09-09 | 62257.916782 | 125450.275210 | 53286.649568 |
| 9 | 2020-09-10 | 62443.857010 | 133588.873063 | 53464.718017 |
| 10 | 2020-09-11 | 62629.797239 | 142367.924759 | 53645.612883 |
| 11 | 2020-09-12 | 62815.737467 | 151825.035431 | 53829.367681 |
| 12 | 2020-09-13 | 63001.677695 | 161999.430938 | 54016.016188 |
| 13 | 2020-09-14 | 63187.617923 | 172932.007543 | 54205.592448 |
| 14 | 2020-09-15 | 63373.558151 | 184665.382625 | 54398.130769 |
| 15 | 2020-09-16 | 63559.498379 | 197243.946419 | 54593.665729 |
| 16 | 2020-09-17 | 63745.438607 | 210713.914801 | 54792.232170 |

Table2. Cardiovascular Death cases prediction

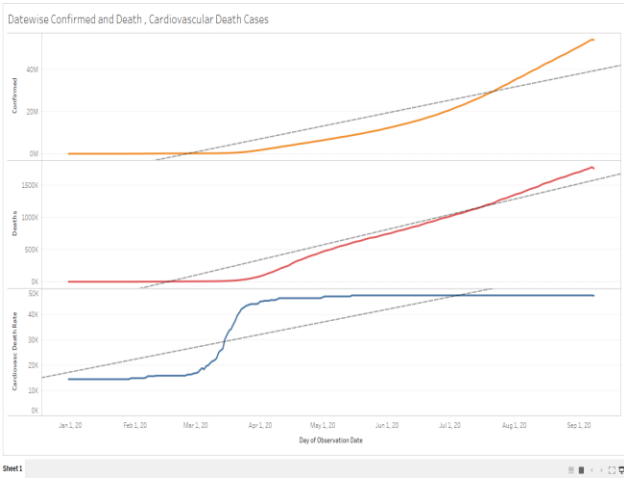


Fig20 Linear Regression Confirmed, Death and Cardiovascular Death Cases in Tableau Tool

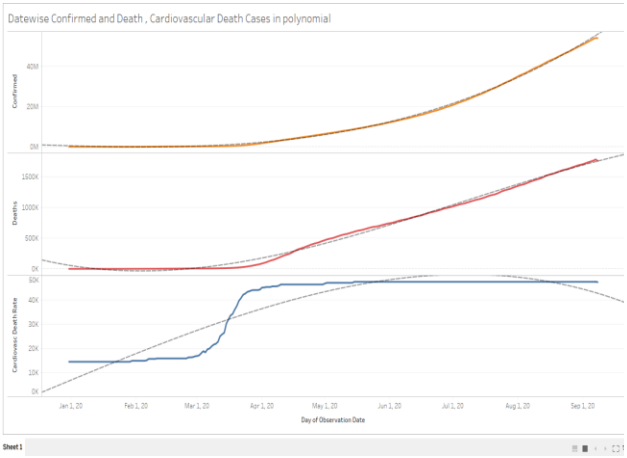


Fig21 Tableau Tool Confirmed, Death and Cardiovascular Death Cases in Polynomial

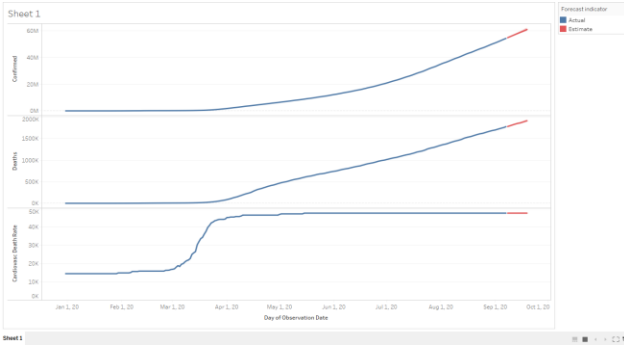


Fig22 Confirmed, Death and Cardiovascular Death Cases Forecast in Tableau

Fig20 shows linear regression best fit of confirmed cases, death Cases, cardiovascular death cases and fig21 show these three cases in polynomial view. The forecasting view of confirmed, death, cardiovascular death cases view in fig22. Visualization part of this tableau tool drag and drop the data within a few seconds shown results.

The dimension

Observation Date Day ranges from December 31, 2019 to September 8, 2020 on this sheet.

The Shelves

Rows : Confirmed, Deaths, Cardiovasc Death Rate

Columns: Day of Observation Date

Confirmed cases in LinearTrend Lines

sum squared error : 1.09085e+16

mean squared error : 4.34603e+13

R-Squared: 0.835198

Death cases in Linear Trend Lines

sum squared error : 5.73405e+12

mean squared error : 2.28448e+10

R-Squared: 0.933285

Cardio Death cases in Linear Trend Lines

sum squared error : 1.4806e+10

mean squared error : 5.89879e+07

R-Squared: 0.708004

P-Value: < 0.0001

## VI. CONCLUSION:

The style of living is changed by data, so we cannot ignore bigdata in everyday life it takes a part direct or indirect manner. Bigdata analytics machine learning algorithm is helps to manage data, especially heart disease and coronavirus patient reports and entire healthcare records to be maintained by this technology. The major cause of covid19 most of the death is diabetic and cardiovascular patients. Linear Regression, Polynomial Regression and SVM Algorithms applied for prediction process. Linear

results near to the actual cases. The comparison of python and Tableau gives best results, Implementation Part Python is very effective tool and the same time visualization part Tableau tool is made easy within a minute we got results. In this work the data set that has been implemented explores cardiovascular death rate is higher than other disease like diabetes prevalence. Henceforth my futuristic focus proceeds towards cardiovascular disease.

## REFERENCES:

**[1] Hengyuan Kang et al, Diagnosis of Coronavirus Disease 2019 (COVID-19) with Structured Latent Multi-View Representation Learning , Published in: 2020 IEEE 10.1109/TMI.2020.2992546.**

**[2] Lifang Li ; Qingpeng Zhang et al, Characterizing the Propagation of Situational Information in Social Media During COVID-19 Epidemic: A Case Study on Weibo Published in: 2020 IEEE , Accession Number 19509183 DOI: 10.1109/TCSS.2020.2980007**

**[3] Di Dong et al, The role of imaging in the detection and management of COVID-19: a review , Published in: 2020Publisher: IEEE, DOI: 10.1109/RBME.2020.2990959**

**[4] Wullianallur Raghupathi et al, Big data analytics in healthcare: promise and potential, Published in: 2014.**

**[5] Vinay Chamola et al, A Comprehensive Review of the COVID-19 Pandemic and the Role of IoT, Drones, AI, Blockchain and 5G in Managing its Impact , Published in: 2020 Publisher: IEEE, DOI: 10.1109/ACCESS.2020.2992341.**

**[6] Yujin Ohet al, Deep Learning COVID-19 Features on CXR using Limited Training Data Sets , Published in: 2020 Publisher: IEEE, DOI: 10.1109/TMI.2020.2993291.**

**[7] Feng Shi et al, A Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation and Diagnosis for COVID-19 , Published in: 2020 Publisher: IEEE, DOI: 10.1109/RBME.2020.2987975**

**[8] Rae Yule Kim, The Impact of COVID-19 on Consumers: Preparing for Digital Sales , Published in: 2020 Publisher: IEEE, DOI: 10.1109/EMR.2020.2990115.**

**[9] Mohamed Abdel-Basset et al, A Hybrid COVID-19 Detection Model Using an Improved Marine Predators Algorithm and a Ranking-Based Diversity Reduction Strategy , Published in: 2020 Publisher: IEEE, DOI: 10.1109/ACCESS.2020.2990893**

**[10] Nanning Zhenget al, Predicting COVID-19 in China Using Hybrid AI Model , Published in: 2020 Publisher: IEEE, DOI: 10.1109/TCYB.2020.2990162.**

**[11] Jingwei Zhang et al, Navigating the Pandemic Response Life Cycle: Molecular Diagnostics and Immunoassays in the Context of COVID-19 Management , Published in: 2020 Publisher: IEEE DOI: 10.1109/RBME.2020.2991444.**

**[12] Abdul Waheed et al, CovidGAN: Data Augmentation using Auxiliary Classifier GAN for Improved COVID-19 Detection, Published in: 2020 Publisher: IEEE, DOI: 10.1109/ACCESS.2020.2994762.[13] R.F. Sear et al, Quantifying COVID-19 content in the online health opinion war using machine learning, Published in: 2020 Publisher: IEEE, DOI: 10.1109/ACCESS.2020.2993967.**

**[14] Jun Wang et al, Prior-Attention Residual Learning for More Discriminative COVID-19 Screening in CT Images, Published in: 2020 Publisher: IEEE, DOI: 10.1109/TMI.2020.2994908.**

**[15] Subhankar Roy et al Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound, Published in: 2020 Publisher: IEEE, DOI: 10.1109/TMI.2020.2994459.**

**[16] Shweta Malhotra et al, Bigdata analysis and comparison of bigdata analytic approaches, Published in: 2017 International Conference on Computing, Communication and Automation (ICCCA), Publisher: IEEE, DOI: 10.1109/CCAA.2017.8229821, Accession Number: 17449090**

**[17] M. Ambigavathi ; D. Sridharan, Big Data Analytics in Healthcare, Published in: 2018 Tenth International Conference on Advanced Computing (ICoAC), Publisher: IEEE, DOI: 10.1109/ICoAC44903.2018.8939061, Accession Number: 19229027.**

**[18] Sunil Kumar ; Maninder Singh Big data analytics for healthcare industry: impact, applications, and tools, Published in: Big Data Mining and Analytics ( Volume: 2 , Issue: 1 , March 2019 ), Publisher: IEEEDOI: 10.26599/BDMA.2018.9020031, Accession Number: 18795700.**

**[19] Xiaoming Yuan et al , A High Accuracy Integrated Bagging-Fuzzy-GBDT Prediction Algorithm for Heart Disease Diagnosis, , Published in: 2019, Publisher: IEEE.**

**[20] Bong-hyun Kim et al, A Proposal of Heart Diseases Diagnosis Method Using Analysis of Face Color, Published in: 2008, Publisher: IEEE, DOI 10.1109/ALPIT.2008.27.**

**[21] Aleksandr A. Samoilov et al, Application of Heart Rate Variability Methods to the TT-intervals and ST-Segments Durations, Published in: 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), Publisher: IEEE, DOI: 10.1109/EIConRus.2018.8317486, Accession Number: 17618396.**

**[22] D. R. Krithika and K. Rohini, "Blockchain with Bigdata Analytics" ICTIDS 2019, Vol 118, page: 403 published Journal :springer.**

**[23] Corona virus data collected from https://github.com/owid/covid19data/tree/master/public/data and cardiovascular data from UCI machine learning repository.**

**[24] D. R. Krithika and K. Rohini, " A Survey on Challenging Capabilities of Big Data Analytics in Healthcare" International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE) Volume 4, Issue 1, DOI: 10.29027/IJIRASE.v4.i1.2020. 593-597, July 2020**

**[25] Rohini, K., and G. Suseendran. "Aggregated K means clustering and decision tree algorithm for spirometry data." *Indian Journal of Science and Technology* 9, no. 44 (2016): 1-6.**