GLOBAL ATLAS FOR PANDEMIC DATA ANALYSIS

[1]SAILAJA, [2]T.SPANDANA, [3]B.MANASA
[1,2,3]Assistant Professor, Keshav Memorial Institute of Technology, Hyderabad

ABSTRACT:
The pandemic situation has brought a huge loss around the globe. It has been a very tedious task to identify the Covid pandemic initially. Though some symptoms are considered, it is still difficult to analyze asymptotic patient effected by Corona. In this paper, we are discussing how we can detect the count of corona cases around the globe. Many countries have reached a high number of cases. The countries around the globe are now not able to even to declare the number of cases in their respective countries. A data analysis is done to find the number of cases in different countries in the world. It also gives a plot of which country has the maximum cases and being effected. The paper contains mainly five sections, which deal with the technology and the process of how the data is analyzed to know the most effected country around the world. The data regarding number of cases around the world are to analyzed by the data analytics. The application of python language for the objective to find number cases in different countries are discussed.

KEYWORDS: CORONA, PANDEMICS, Data Analytics, Machine Learning

1. INTRODUCTION:

Coronaviruses (CoVs) represent a major group of viruses mostly affecting human beings through zoonotic transmission. In the past two decades, this is the third instance of the emergence of a novel coronavirus, after severe acute respiratory syndrome (SARS) in 2003 and Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012[1],[2]. The repeated emergence and global scale of transmission, significant number of deaths, infection and mortality of care providers and healthcare workers (HCWs), and higher risk of death in vulnerable or susceptible groups, have been the major causes of concern. Integrated early warning and response systems are an effective way to raise a timely alarm about these emerging and re-emerging pathogens, but few tools are available to enable pre-emptive prediction of such diseases. The Global Virome Project has been initiated with the objective of creating a global atlas of pathogenic viruses, with the specific objective of identifying spill-over events [3],[4]. The project has not been without its critics, and is not yet close to providing evidence that can be translated into preparedness action [5]. This underscores the importance of preparedness of the health system to deal with dangerous pathogens and better control of endemic infections. The process of naming the novel coronavirus (2019-nCoV) which emerged in Wuhan, China, in December 2019, has created some controversies [6]. In this review, the WHO convention of referring to the disease condition as novel coronavirus disease (COVID-19) has been followed [7]. The virus will be referred to as SARS-related CoV-2, or SARS-CoV-2[8]. COVID-19 has been labelled as a public health emergency of international concern (PHEIC)[9,] and the epidemic curves are still on the rise[10]. Here, we summarize the clinical and public health aspects of COVID-19 and SARS-CoV-2, and the lessons gleaned from the global responses so far. As more data continue to emerge, the epidemiology of the disease will come into sharper focus.

2. DATA ANALYTICS:

Data is fuel for most of the companies irrespective of the domain they are operating e.g. banking, insurance, IT, manufacturing etc. Understanding data and leveraging it has become a significant part of the processes impacting business more than ever. Therefore, understanding data is a critical business requirement nowadays. But before understanding and analyzing data we must make sure that data is correct and well-structured else we will fail to fulfill the objective. People are aware of the term data analytics. Understanding this term has a large impact on the business. This gives a clear picture of what data is to be needed for the business. Many people have used the words of data

science and data analytics interchangeably. Its very necessary to know the difference between these terms as they effect business a lot. Data analytics include some steps as follows.

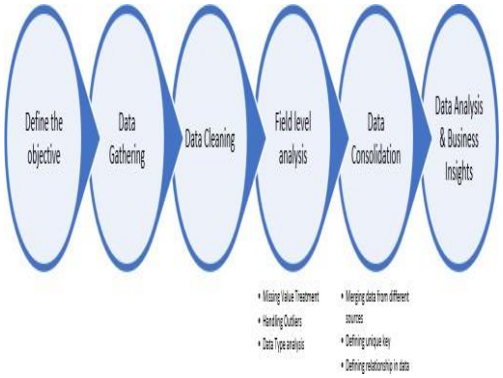

FIG1: STEPS OF DATA ANALYTICS

3. DATASCIENCE:

It is a broad term for number of models and varieties to gather information. Data Science mainly deals with the scientific methods and many other math tools including statistics to analyze and manipulate data.It includes number of tools which are being used to get information from the data.The practice of Data Science can be used to connect information and many data point connections. These data points seem to be very crucial for any business. New patterns and insights of data can be researched and analyzed in data science Many foresights using analytics can be sorted out using different models.

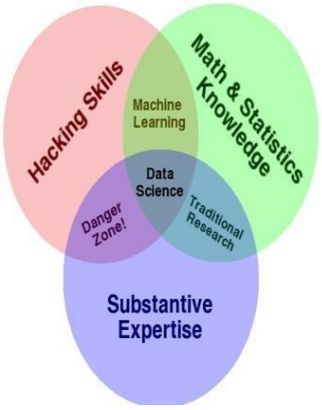

FIG2: DATASCIENCE vs DATANALYTICS

4. MACHINE LEARNING:

Machine learning can be defined as the practice of using algorithms to extract data, learn from it, and then forecast future trends for that topic. Traditional machine learning software is comprised of statistical analysis and predictive analysis that is used to spot patterns and catch hidden insights based on perceived data. Many machine learning algorithms gather behavioral information for every user on the social platform. Based on one's past behavior, the algorithm predicts interests and recommends articles and notifications on the news feed. Similarly, when Amazon recommends products, or when Netflix recommends movies based on past behaviors, machine learning is at work.In ML data is being categorized into Supervised, Unsupervised types. The data, which is collected, and used for a particular analysis is first categorized into any one of these. The type of algorithms used are to be decided according to the output and analysis expected. Examples of Supervised Learning: Regression, Decision Tree, Random Forest, KNN, Logistic Regression etc. Examples of Unsupervised Learning: Apriori algorithm, K-means.

4.1 COMMON ML ALGORITHMS:

4.1.1 Linear Regression

It is used to estimate real values (cost of houses, number of calls, total sales etc.) based on continuous variable(s). Here, we establish relationship between independent and dependent variables by fitting a best line. This best fit line is known as regression line and represented by a linear

equation Y= a *X + b[1].The best way to understand linear regression is to relive this experience of childhood. Let us say, you ask a child in fifth grade to arrange people in his class by increasing order of weight, without asking them their weights! What do you think the child will do? He / she would likely look (visually analyze) at the height and build of people and arrange them using a combination of these visible parameters. This is linear regression in real life! The child has actually figured out that height and build would be correlated to the weight by a relationship, which looks like the equation above.

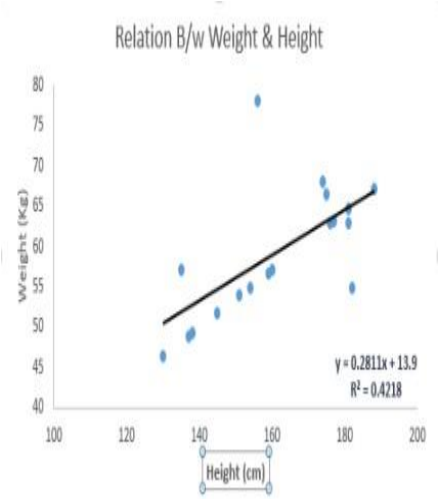

FIG3: LINEAR REGRESSION

4.1.2. Logistic Regression

It is a classification not a regression algorithm. It is used to estimate discrete values (Binary values like 0/1, yes/no, true/false) based on given set of independent variable(s). In simple words, it predicts the probability of occurrence of an event by fitting data to a logit function[4]. Hence, it is also known as logit regression. Since, it predicts the probability, its output values lies between 0 and 1 (as expected).

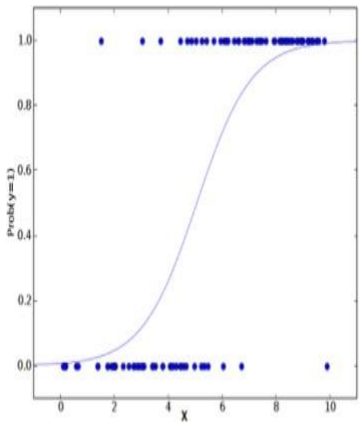

FIG4: LOGISTIC REGREWSSION

4.1.3. Decision Tree

This is one of an algorithm and used quite frequently. It is a type of supervised learning algorithm that is mostly used for classification problems. Surprisingly, it works for both categorical and continuous dependent variables[5]. In this algorithm, we split the population into two or more homogeneous sets. This is done based on most significant attributes/ independent variables to make as distinct groups as possible.

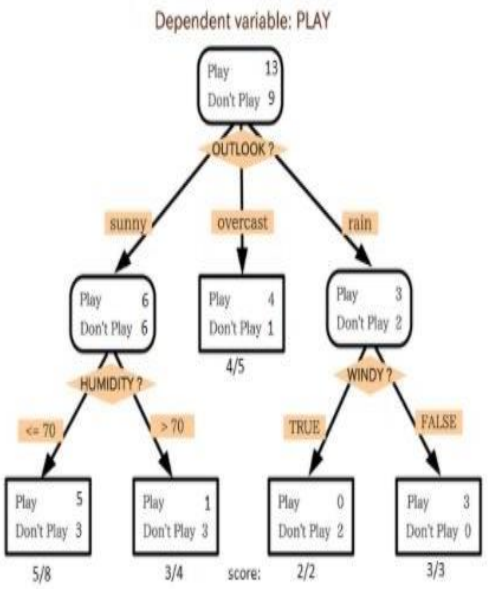

FIG5: DECISION TREE

5.PYTHON

Python is a general-purpose, high-level programming language which is widely used in the recent times [1][2][3]. Its design philosophy emphasizes code readability, and its syntax allows programmers to express concepts in fewer lines of code than would be possible in languages such as C [4]. The language constructs enable the user to write clear programs on both a small and large scale [5]. The most important feature in Python being it supports multiple programming paradigms, including object-oriented, imperative and functional programming or procedural styles. Python supports a dynamic type system and automatic memory management and has a large and comprehensive standard library. Python interpreters are available for many operating systems[10]. Python's status as the fastest-growing programming language is being fueled by a sharp uptick in its use for data science. Among the Python tagged questions, the fastest growing tag is related to pandas, a data analytics software library for Python[13],[14]. Only introduced in 2011, it now accounts for almost 1% of Stack Overflow question views. However, the second most visited tag by Python visitors is JavaScript, likely reflecting the healthy use of Python by web developers.

6. METHODOLOGY:

The pandemic situations around the globe has been increasing day by day. It is very essential to know how pandemics are making strong effect on various countries. It is also a sign for other countries nearby to be careful. Any virus or disease is not known until it starts effecting the human race.in such cases it even becomes very difficult to find the exact symptoms of a disease like Corona. Even though COVID is one year old, still there are no exact symptoms to be expected by doctors and medical experts.

In such cases, it is very essential to at least keep track of all corners of the world. This can warn the human race to be cautious and at least make people to take their own precautions[16]. In our methodology we have used python code with ML algorithms to get track of the number of cases of the present pandemic that is the covid.Though SARS and MERS were handled by humans,COVID has become unstoppable.SARS and MERS were from the same family of COVID which attacked African Countries and were controlled later. Considering the data, we have taken a linear regression from the data of how the disease is changing its form from one country to the other. Data analytics have been applied to present of a graph of the present situation. This graph helps the experts and researches to find the maximum number of cases in the whole world. It also indicators the number of cases in each and every country.

7. CODE:

```
def get_info(country_name):
```

```
    url="https://www.worldometers.info/coronavirus/country/" + country_name + "/" data = requests.get(url) soup = BS(data.text, 'html.parser') cases = soup.find_all("div", class_ = "maincounter-number")
total = cases[0].text
total = total[1 : len(total) - 2]
recovered = cases[2].text
    recovered = recovered[1 : len(recovered) - 1]
      deaths = cases[1].text
      deaths = deaths[1 : len(deaths) - 1]
      ans ={'Total Cases' : total, 'Recovered Cases' : recovered,'Total Deaths' : deaths}
    return ans
   country_name = "us"
  us = get_info(country_name)
print("Cases in United States")
for i, j in us.items():
    print(i + " : " + j)
print("..........")
country_name = "india"
  india = get_info(country_name)
  print("Cases in India")
for i, j in india.items():
    print(i + " : " + j)
```

8. RESULTS:

In this study we have come across data which is being analysed by different ML techniques.The results of global data analysis of corona is being observed in a document type.Though it has been the output expected,a visualization part would make things clear

RESULT1:

Search: india

Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	Active Cases	Serious, Critical	Tot Cases/ 1M pop	Deaths/ 1M pop	Total Tests	Tests/ 1M pop	Population
India	6,737,544	+55,471	104,179	+579	5,718,802	914,563	8,944	4,870	75	81,071,797	58,596	1,383,567,415
Total:	35,923,257	+201,311	1,051,555	+3,069	27,002,454	7,869,248	67,004	4,608.6	134.9			

RESULT2:

Search: spain

#	Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	Active Cases	Serious, Critical	Tot Cases/ 1M pop	Deaths/ 1M pop	Total Tests	Tests/ 1M pop	Population
1	Spain	865,631	+12,793	32,486	+261	N/A	N/A	1,544	18,512	695	13,689,776	292,769	46,759,607
	Total:	35,923,257	+201,311	1,051,555	+3,069	27,002,454	7,869,248	67,004	4,608.6	134.9			

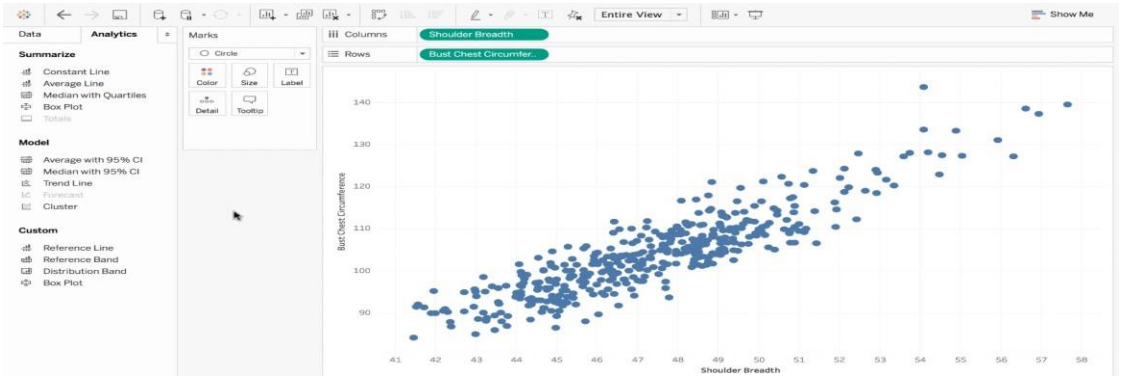

FIG6: VISUALIZATION OF THE DATA USING TABLEAU

9. CONCLUSION:

The technology has been developing day by day but still its not enough to identify pandemics like corona. The application of technologies applied in medical field give tremendous results. In this paper we have analyzed the covid data by the help of ML techniques. The ML including others models like RNN, CNN can be applied for such pandemic detections. The use of deep learning techniques can help the medical diagnosis. Thus many more technologies must be researched and applied. Though some of them are used, still there is no advancement for such pandemic detection neither their symtoms.The research on such technologies must be improved to protect the mankind.

10. REFERENCES:

[1]. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008 Feb 21;451(7181):990–3. https://doi.org/10.1038/nature06536,doi:http://dx.doi.org/10.1038/Nature 06536 PMID: 18288193
[2]. The control of neglected zoonotic diseases: a route to poverty alleviation.Geneva: World Health Organization; 2005.
[3]. Morse SS, Mazet JA, Woolhouse M, Parrish CR, Carroll D, Karesh WB, et al.Prediction and prevention of the next pandemic zoonosis. Lancet. 2012 Dec1;380(9857):1956–65. doi: http://dx.doi.org/10.1016/S0140-6736(12)61684-5 PMID: 23200504
[4]. Carroll D, Daszak P, Wolfe ND, Gao GF, Morel CM, Morzaria S, et al. The GlobalVirome Project. Science. 2018 Feb 23;359(6378):872–4. doi: http://dx.doi.org/10.1126/science.aap7463 PMID: 29472471
[5]. Woolhouse MEJ, Gowtage-Sequeria S. Host range and emerging andreemerging pathogens. Emerg Infect Dis. 2005 Dec;11(12):1842– 7.
[6]. Leroy EM, Kumulungui B, Pourrut X, Rouquet P, Hassanin A, Yaba P, et al.Fruit bats as reservoirs of Ebola virus. Nature. 2005 Dec 1;438(7068):575 6.https://doi.org/10.1038/438575adoi: http://dx.doi.org/10.1038/438575a PMID: 16319873
[7]. Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P.Host and viral traits predict zoonotic spillover from mammals. Nature. 201706 29;546(7660):646–50. https://doi.org/10.1038/nature22975doi:http://dx.doi.org/10.1038/ nature22975 PMID: 28636590
[8].PREDICT [internet]. Davis: The Regents of the University of California, Daviscampus; 2017. Available from: http://www.predict.global [cited 2017 Oct 15].
[9]. PREDICT Consortium. Reducing pandemic risk, promoting global health.Davis: The Regents of the University of California, Davis campus; 2014.Available from: http://www.vetmed.ucdavis.edu/ohi/local_resources/pdfs/ predict-global-flyer.pdf
[10]. Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P.Host and viral traits predict zoonotic spillover from mammals. Nature. 201706 29;546(7660):646–50. https://doi.org/10.1038/nature22975doi: http://dx.doi.org/10.1038/nature22975 PMID: 28636590
[11] TIOBE Software Index (2011). "TIOBE Programming Community Index Python". 1
[12] "Programming Language Trends - O'Reilly Radar". Radar.oreilly.com. 2 August 2006.
[13]. "The RedMonk Programming Language Rankings: January 2011 – tecosystems". Redmonk.com.
[14]. Summerfield, Mark. Rapid GUI Programming with Python and Qt.
[15]. Kuhlman, Dave. "A Python Book: Beginning Python, Advanced Python, and Python Exercises".
[16]. https://stackoverflow.com
[17]. https://github.com