COVID VACCINES ANALYSIS

ABSTRACT : The Covid-19 pandemic has Shaken the world completely. No one knew
What was coming and everyone was running Helter-skelter. The governments were paralyzed And the infrastructure required to deal with This problem was absent completely. The Genome sequence was out. But what the disease Entailed and what it will lead out was just Anyone's imagination. Till today as we write There are multiple dimensions of it that lay Unexplored and need a deep exploration to be
Found out. Our Project seeks to uncover the Mystery using the application of data sciences To solve it. We seek to use data sciences to help Authorities and also to give the medical field The insight that data can provide to them to Deal with the pandemic better. Data science is The application of data science algorithms and Machine learning to train the models to find Patterns. Patterns reveal what the common Issues are and common symptoms and Everything that is common comes out in a Visual representation. It's these representations Which make complex things easy and digestible To people from non tech backgrounds. Use of data science in such a pandemic will lea

public figures Of daily confirmed Of ftOiii John Hopkins were in this study article by VZ.Marmarelis.121_ RM as described

element for the method public figures Of daily updated confirmed instances Of Cov1d.19 from John this study article by etal. 121. RM. as described by Ricc:ati Equation. is

Further by applying the equation wc find 5 different parameters and then dependence on the no, Of caws day by day".

Everyone analysed knowledge on disease and therapy utilmng articles from Gerry Wolfe•, Ashraf elnashar•.
Will Schreiber• Alsnudi•, Guided by Literary Of from Kaggle based on

COVID•19, data were divided into) Mobility social (2) and (OVID; Economic impact. and (4) Vulnerable population. and woe utilised in a second dataset from MTV document has been analysed and has processed 10 produce and the use of the K-Medlan method to label data to data,

According to Tuli.141 the epidemic may be tracked extremely via Shrestha et al Machine l.carmng (Ml.) and Cloud Computing. anticipate an outbreak of the illness, and create appropriate policies to regulate its expansion given the array, face extraction Mid collection done They have proposed a Machine t.earn•ng model that can be run continuously on Cloud Data Centers (CDC') for accurate spread prediction and proactive development of

strategic respnse by tlw• government and citizens, The dataset used hy them In this case study.World in Data by Hannah Ritchie Iliey have also a cloud framework and azure instances for real analysis of dau The research paper (S) Francisco Gois et al. have entphasised the rising. Of epidemic due to their to the natural Of Viruses, study presents several predictor amroaches With machine

epidemiological in order to explain COVID-19's

palyr 161, the authors Yan-ed Zoabi, Shira Deri•Rozov and Noam Shomron have that accurate SA allows fast and diagnosis reduces the strain health care

characteristics have been created to likelihood of infection The rnodel 0 90 auROC •n Kuward•looking Orca under the operattng curve):.

paper 171. authors Enis and IkWanAy•dtn mentioned incident at COVID-19 showed that the world was unwilling io Virus so One crucial fictor in mn.gaung the detrunental impacts of an epidemic or pandemic is enectwe use of information technology suggested managcrncnt epidenuc system (EMS), which relies on the unfettered and timely flow of information between states and organisations, They have been using an MPISA parad•gm.which allows different platforms to tv integrated and gives gives the solution for issues of scalability interoperability.

and

[8]This palxr lhscritvs the use of a new cpidcnuologlcal comparumcnl•based numlel f'N the estimation of the propagation of the coronavirus CO VID.19. that is, SEIA R(Suscepuble Exposed Asympionuiic Infectious Recovered). This is xcomplished through the heuristic approxh of differential evolution. In this way the day(s) that numtxr reaches its rnaxvmum. the associated value and the future evolutim of its spread may be evaluated in approximate order for different

Ibe authors Ayyoubzadeh S et al haveUsed computenscd data numng technologies for improved insights on m.libre.ak Of in exh Country and globally for

management Of the health Trends website collected data For estimating the number Of COVID—19 linear reeression and long.tem (ISTM) models were

study by Kwe\ha Rashid.Hcamn N Abduljabbar and Bilal shows that in research. may bc proved to be deternumstic. transforming into clear findings and predictgons, outcontes Of supervised learning algorithms are better than those Of Of uncontrolled learning algonthrns. assistance for the Of standard diagnostic procedures like lgM. lgG, X•ray chest. and RT•PCR be seen as an intelligence and deep learning CNN Algorithms to this study

Xceptlon. Incept10nV3. IncepuonResNctV2, VGGNet, NASNet.

---

3. IMPLEMENTATION

3. I Methodology

We are using Machine Learning to give predictions on the basis of data taken from government websitell ll. 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 S maJor symptoms.

The process can be explain in following m)ints

1. First. Take the dataset. remove redundant data and organise the data io our
2. Second. Load the dataset cm the Jupyier Notebcok and apply data visualintion techniques to understand the data better. 3. Third, then wc calculate accuracy for various algorithms and plot graph on the basis o f 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 •vc or •eve on basis Of symptoms.

3,2 Description Ofthe Process

We building our own COVID Prediction System using jupyter Notebook.

We can describe the process in following steps:

Step l: Cleaning the dataset

TheVery first Step in our is to get a reliable and authenuc dataset the prediction and analysis,

Our search for dataset ended on I I I I which is govt website which has provided for free use and is absolutely authentic,

•nien next thing we did was to eiean the dataset and remove unwanted columns from dataset for foster computation

Step 2: Data Visualization

Here, we use the dataset and cheek the conststem•y of the dataset by checking the values out or the dataset randomly

Ttxn wc do data visualization for better understanding of data by the use of vanous plots. graph and heaimaps. All this and plots gets us an insight into huge datasets easily

Step J: Computing Aeeuraey

In this step we

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

Random Forest Classmer, free Algorithm . we selected these algorithms on the basis Of their qualities Of regression &

classification.

In the last Step. all we need to is p graph Of accurxy Of the algori and use the algorithm with acvura predict whether a 2 person has co or not

We take mptn Of S symptoms in binary values and

using our predictor we predict the person is positive or negative on the basis of these 5 symptoms.

3.3 Algorithm

1. Logistic Regression

Logistic Regression is a Classification mcukl. which tries to classify the data based on the probabllity Of it

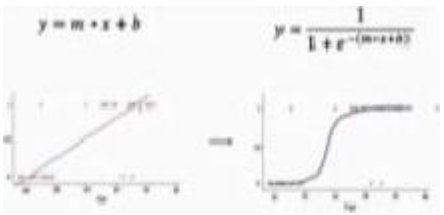
occurring This algortthm is used in multiple places where class I fieatl:on is we have used it to classify if the patient is susceptible infectedby eovidor not

This is one Of the classification methods which we have used It uses Sigmoid function to classify the data

$$sigmoid(z) = \frac{1}{1 + e^{-z}}$$

e = Euler's number = 2.71828

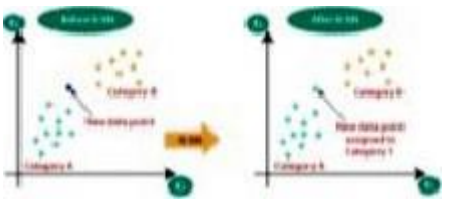
Sigmoid function converts input into range 0 to 1



KNN

KNN is a supervised machine learn mg algorithm KNN forms groups based on the enterias and decides for theineorn•ng data where to put in in which category

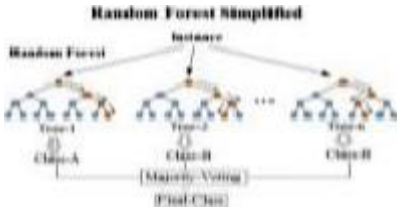
It can be used ror regression and for classification too, but mostly for the classification only its used



Random Forest Classifier

Random forest is a super-vised learn mg algorithm nforest*it builds is an ensemble Ofdecision trees.

usually trained with the "bagging" method The general idea of bagging method is that a combination or learning models increases the overall result, put simply: random forest builds multiple decision und merges them together to get J more accurate and stable prediction, One big advantage of random forest is that i! used both classification and regression problems, which form the majority Of Current nvochine learning systems



4, Decision tree Al gorilhi'ii

a.Oeeision Tree i' a supervised

•earning algorithm

b.Tluo 'K)des which are decision

node and leaf node are the ones making the decision

C . Repeated if clauses arg a' work when deciding the classificaticm for the algorithm

4. SYSTEM REQUIREMENTS

4.1 General Description

Data Analytics on Covid.19. as the name suggests is data analytics on the data such as tlE people infected.what thor

age is .whai are the sources that they ha ve been infected history Of previous chrome diseases and wc wish to obtain almost all the meaningful insights that wv can get using various data science and machine learning techniques and by leu»king at unose insights we can arrive at or basically predict the iUture trends or other crucial infitrmat.on It requires w•uvc internet connection because tlu• pro'ect uses various Machine I.earnmg model' depending on how we want to wain our data Ille various tools and library that wc intend to use are With tlw• intention that using them wc can gel the "best of the waste" and provide some services to the society.llence we look forward to *'hteve what we have intended and hope the analysis turns out to be a success.

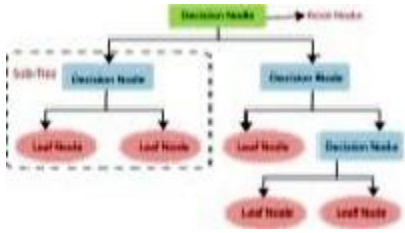
4.2 IIARDWARE REQUIREMENT S i , High Resolution Camera

2. RAM: 4 GB
3. Processor: intei if or nigher

4.2 GB Graohics Card

4.3 SOFTWARE REQUIREMENTS

1. Windows 7 or
2. Text Editor
3. python 3.9.O



4.3.1 Non-functional and functional requirements

System functional requirement defines the operations and services to be provided by the system

:•

1. Using **Jupyter Notebook** the csv file is

5. Jupyter Notebook

Using manipulated for gettung meaningful insights.

Tln•y are clarified by the following points.

1. RELIABILITY • The insights that we a.mjng to obtam should be highly reliable wIth mm. mum faults or nuscak•uianons Every paranxter of the dataset ruent.oned and observed properly and the "bights that we arrive at. are cross checked from practical,'prevlous observations.

4. Open CV

2. OpenRenne f'" data scrubbing.

Numu.Eandas.Mat01utli b fev data and visualisation Vor modellng the data we need decent knowledge o f Of Python. Tratmng the dataset 6.

•nterpetjng the data.

Non-functional Any features or qualities of the system capable Of evaluating its operation are the requirements

2. SCALABII.ITV • Sime new records are added to our dataset on daily basis otü model should be scalable to adopt the dynamic nature of dataset

J. SECT RITY . prowct is mainly dependent on the database from an open source data repository .there IS a high chance of data loss due to hackers or attackers So our system should s«ured

4. system requires g€»d maintainability from our

dataset Smce there might days "hen there is surge in numtk•t of dally cases abruptly and we need to such data

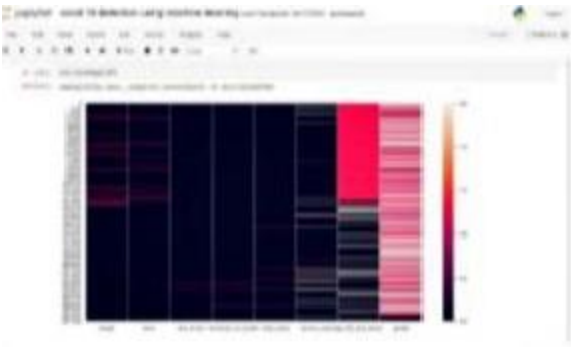
4_32 USER REQUikEMENTS l. data analysis system shall input andaccurately compare the With the previously stored data

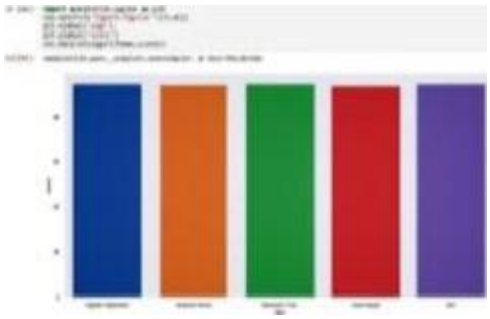
2. the input the 'Vobabdity of havine or is as a percentage.

3. A front•crul intcrfa•e ror taking the symptoms paralMers from the patients IS

4, user's parameters are compared against the test on Which the modelhas been traineduser shall keep his/herconnected to our database.

S RESULTS





The screenshots above show the code and results

Ofthe vanous phases ofthe Data Analysts done by

us on our Covid-19 dataset. The implementation

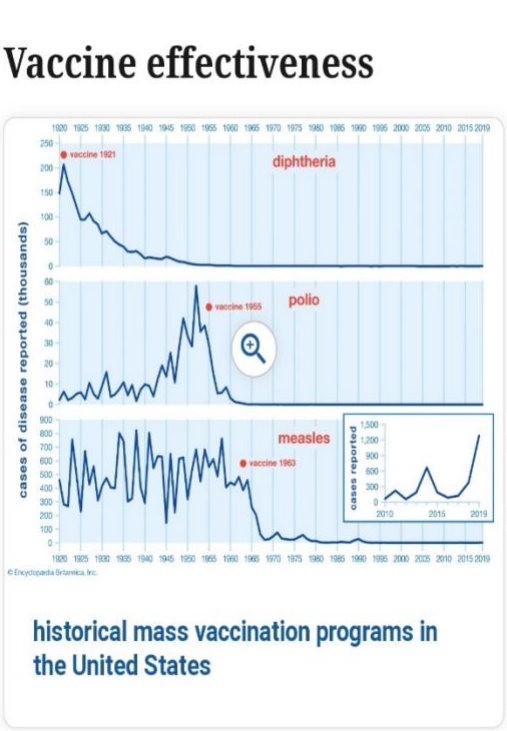
or data analysis has been carried out by varunls algorithms based on their When analysis done by various aV1thms the most accurate results were yielded by the randorn forest classifier algorithm We.wh:ile carrying out the analysls, into consideration the major characteristic features like cough.fever.etc. which the result Of whether the person

is on these symptomsIn phases we were also able to &termine Whether the Frson was eovid negative or positive basedon his data which is taken by a small tkinter interfxe

6. CONCLUSION

- 19 is a huge struggle f'E all Of us. The we are makmg will seek to find the answers to the most questions as to What it that makes 19 such a tragedy and what an are the ones Who arc affected by it it seek to find the approrate can mounted by the and can reach to a place of' the problem and soive it in the best manner there It will also lead to a solution to any medical comhtion might encounter later on •n our lives Where we apply sciences for diagnostics project on the already limned that India have and rvevenis the spread as rxople use it to get an Idea they should go get tested unhealthy and to

USing this SyStCiii effectively and



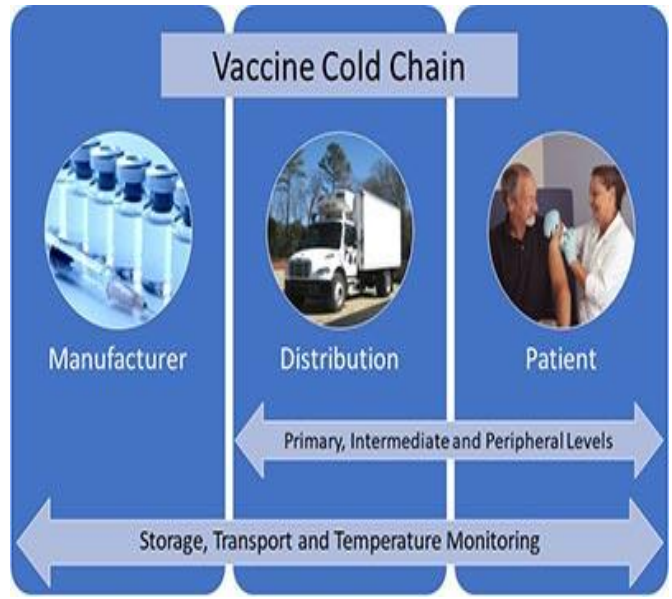
After Pasteur's time, a widespread and intensive search for new vaccines was conducted, and vaccines against both bacteria and viruses were produced, as well as vaccines against venoms and other toxins. Through vaccination, smallpox was eradicated worldwide by 1980, and polio cases declined by 99

efficiently the on sySieiii is stressed out The ability to unbundle those first four functions affected how the pharmaceutical industry was organised heading into the pandemic. Splitting apart the third and fourth steps in particular – the heart of the vaccine manufacturing supply chain – ultimately affected how many doses were produced, where and how quickly.

COVID 19 vaccine data systems

Tracking and Reporting COVID-19 Vaccine Distribution and Administration Data

Tracking COVID-19 vaccine distribution and administration activities requires collaboration between public and private information technology (IT) systems and integration of existing and newly developed IT systems.



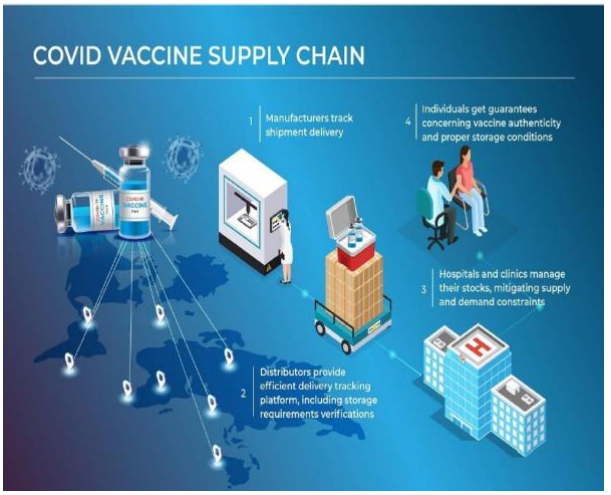
The safe transport of pharmaceuticals, biologics, lab specimens, and temperature-sensitive reagents is mission critical. Our end-to-end

portfolio of custom cold chain solutions helps protect your shipments whether they are going across the country or across the world.

Get it there at the right time and at the right temperature

Cold Chain Storage
Cold Chain Packaging
Transport Management
Visibility and Monitoring
Global Quality Assurance

COVID vaccine supply chain



It is organised as follows. Section 2 provides a simple analytical framework through which to view the vaccine value chain. It identifies the five main steps critical to getting a new vaccine from start to finish: research and development; clinical trials; production of the drug substance and its formulation into drug product; 'fill and finish', or the assembly-line process of putting a vaccine into millions of tiny vials; and then distribution.