CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

This paper represents that prediction of COVID-19 Outbreak is beneficial for the healthcare centres as it helps to minimize the burden on the healthcare centres by providing the best means for diagnosis and medication for COVID-19 affected patients .This project focuses on epidemiological dataset of COVID-19 patients and by analysing that data we will be developing data mining models for the prediction of outbreak of COVID-19 in the upcoming days and also for the prediction of death rate of COVID-19 infected patients. This project would be very helpful for healthcare to fight against the COVID-19.

This project is useful for the government to determine the amount of medical needs (such as equipment, medicines and accommodation)that are to be provided in the upcoming days for the COVID-19 affected patients. The data about the number of cases and deaths will be downloaded dynamically from the website till date. Later the system examines the previous data and will generate the prediction about cases and deaths in upcoming days. This System predicts the outbreak of COVID-19 in the future. This System uses Neural Networks and Polynomial Regression methodologies. This system is helpful for organizations that are trying to help the patients who are suffering with lack of medical facilities. It helps in decreasing the death rate by providing the required number of medical needs to the people.

INTRODUCTION

1.1 INTRODUCTION

COVID-19 is a disease caused by a new virus, which emerged in Wuhan. People can get affected with COVID-19 from others who have the virus .This virus spreads through small droplets from the nose or mouth which are spread by coughing or exhaling. We should note the point that, This virus will stop spreading when it does not find any new bodies to infect.

It is important to investigate the growth of transmission and predict the occurrences of the transmission in the future. Mathematical models which are based on machine learning are chosen to predict the outbreak of the virus. Machine learning and Neural Network technologies are implemented using the python library to predict the total number of confirmed and death cases .This prediction allows us to undertake specific determinations based on transmission growth, such as increasing the lockdown phase, performing the sanitation, and providing hospitality needs and daily support.

Today, a clear solution has not been developed on COVID-19. The vast majority of measures taken on a country basis and individually are to prevent the transmission of this virus to more people. Because of the uncertainty in the transmission dynamics of SARS-CoV-2 and high certainty in its virulence, it is understandable that early responses have relied on blunt interventions, such as movement bans and closures, to save lives.

Given the increasing caseload, there is an urgent need to augment medical and economical skills to face this critical illness. Hence, the scientific challenge now is to identify, through inference and simulation, measures that could provide as-good or better protection with less social cost The growing emphasis on machine learning techniques in medical fields can provide the right environment for change and improvement.

To address this global novel pandemic, WHO, scientists and clinicians in medical industries are searching for new technology to screen infected patients in various stages, to find best clinical trials, control the spread of this virus, develop a vaccine for curing infected patients, and trace contacts. The role of the data science in this scenario consists in helping to speed up the process.

Machine learning has proven to be invaluable in predicting risks in many spheres and since the spread of the virus started, its application is helping us fight against the viral pandemic. Like never before, people all around the world are collecting and sharing what they learn about the virus.
Starting from this, the main goal of this work is to shine a light on their work, high- lighting the importance of the role of machine learning to tackle SARS-CoV-2 (Figure 1.1).

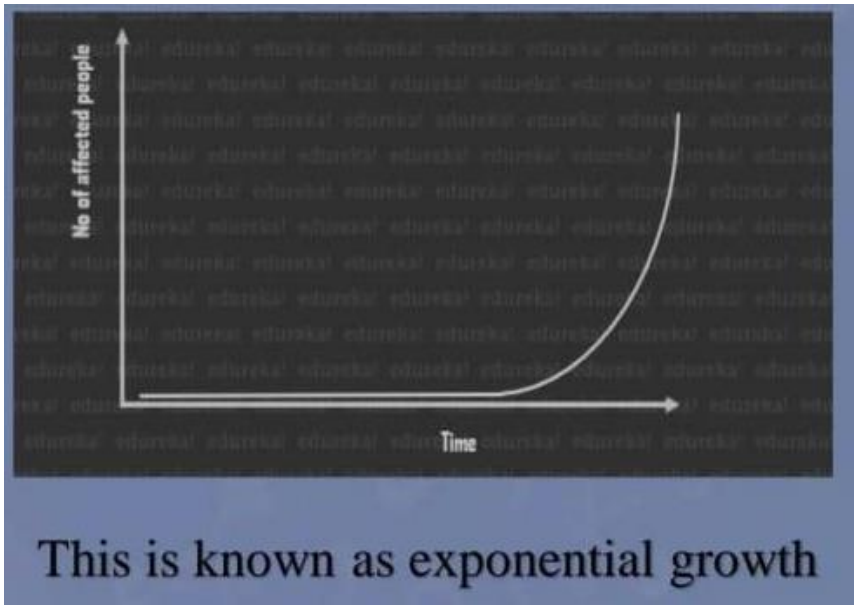

Fig.1.1.1(a): Exponential Growth.

But it cannot go on forever. The virus will gradually stop finding new bodies to infect and ultimately will slow down the count. This is called logistic growth and the curve is known as a sigmoid.

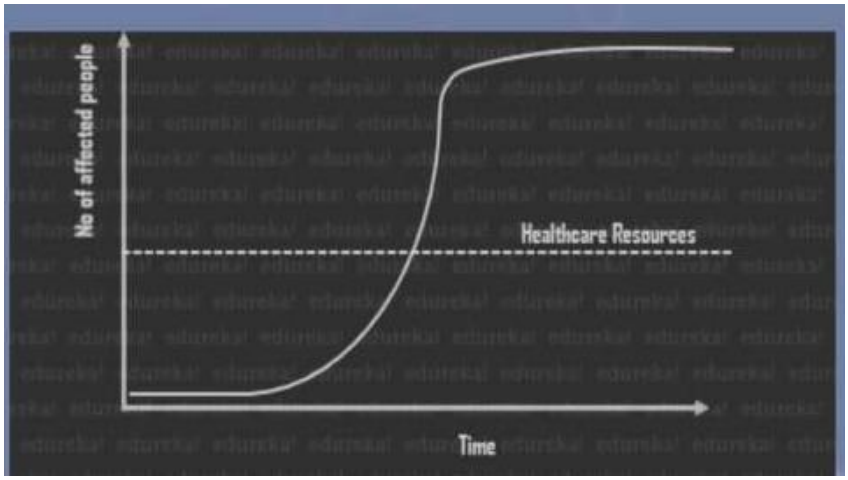

Fig.1.1.1(b): Sigmoid Curve.

Here, every point in the curve will give you the total number of cases on the current day. By plotting the Slope of each day, we can get the new cases on that specific day. There are only a few new cases at the beginning and at the end, with a sharp rise in between.

As we can observe, the peak of the curve may greatly overwhelm our healthcare systems, which is the amount of resources available to us for the care of affected individuals at a given point of time.

The main objective of this paper is to overview the data mining algorithms which are used to predict the outbreak of COVID-19.DataMining techniques are therefore used to study and analyze data and then identify any hidden patterns or information from a data set. These techniques are used to mine the characteristics of the curve and then train the model accordingly to predict the outbreak of corona cases in the future. Using these techniques, we can analyze the death rate and outbreak rate of COVID-19 infected people. Thus, it guides the healthcare centre managers to better develop the organisation by providing sufficient hospitality needs to the infected people.

Regression- is a technique that is made use of to predict a variety of numeric values, including sales, price of a stock, temperatures, that are based on a precise dataset.

1.1.4 WORKING OF DATA MINING

Exploring and analyzing large quantities of information to derive relevant patterns and trends is involved in data mining. Data mining has many uses such as credit risk management, database marketing, spam email filtering, fraud detection, and also to fathom the opinion and sentiment of users. The data mining process is further divided into five steps. First data is collected and loaded into the data warehouse.

Then the data is managed and stored either in the cloud or in the in-house servers. The data is assessed by management teams, business analysts, and information technology professionals and they determine how to organize the data. Then based on the results of the users, it is sorted by the application software. Finally, the data is presented by the end-user in a format like a graph or a table that is easy to share.

The first step in data mining is almost always data collection. Today's organizations can collect records, logs, website visitors' data, application data, sales data, and more every day. Collecting and mapping data is a good first step in understanding the limits of what can be done with and asked of the data in question.

The Cross-Industry Standard Process for Data Mining (CRISP-DM) is an excellent guideline for starting the data mining process. This standard was created decades ago and is still a popular paradigm for organizations that are just starting.

The 6 -DM phases

The CRISP-DM comprises a six-phase workflow. It was designed to be flexible; data teams are allowed and encouraged to move back to a previous stage if needed. The model also provides opportunities for software platforms that help perform or augment some of these tasks.

1. Business understanding

Comprehensive data mining projects start by first identifying project objectives and scope. The business stakeholders will ask a question or state a problem that data mining can answer or solve.

2. Data understanding

Once the business problem is understood, it is time to collect the data relevant to the question and get a feel for the data set. This data often comes from multiple sources, including structured data and unstructured data. This stage may include some exploratory analysis to uncover some preliminary patterns. At the end of this phase, the data mining team has selected the subset of data for analysis and modeling.

3. Data preparation

This phase begins with more intensive work. Data preparation involves preparing the final data set, which includes all the relevant data needed to answer the business question. Stakeholders will identify the dimensions and variables to explore and prepare the final data set for model creation.

- Supervised learning algorithms
- Cross-validation
- Unsupervised learning algorithms
- Various toy datasets
- Feature extraction

MATPLOTLIB

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. Matplotlib is one of the most popular Python packages used for data visualization. It is a cross-platform library for making 2D plots from data in arrays. It provides an object-oriented API that helps in embedding plots in applications using Python GUI toolkits such as PyQt, WxPythonotTkinter. It can be used in Python and IPython shells, Jupyter notebook and web application servers also.

matplotlib. **pyplot** is a collection of functions that make **matplotlib** work like MATLAB. Each **pyplot** function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels.

PROPHET

Prophet ,it is an open-source library designed for forecasting time series data data based on an additive model where nonlinear trends are fit with yearly, weekly, and daily seasonality.It works best with time series that have strong seasonal effects and several seasons of historical data.

Advantages of Prophet:

- Fully Automatic
- Tunable Forecasts
- Fast and Accurate
- Robust

1.1.8 Machine Learning with Python

Machine Learning is simply making a computer perform a task without explicitly programming it. In today's world every system that does well has a machine learning algorithm at its heart. Take for example Google Search engine, Amazon Product recommendations, LinkedIn, Facebook etc, all these systems have machine learning algorithms embedded in their systems in one form or the other. They are efficiently utilising data collected from various channels which helps them get a bigger picture of what they are doing and what they should do.
Python is a widely used high-level programming language for general-purpose programming. Apart from being an open source programming language, python is a great object-oriented, interpreted, and interactive programming language. Python combines remarkable power with very clear syntax. It has modules, classes, exceptions, very high level dynamic data types, and dynamic typing. There are interfaces to many system calls and libraries, as well as to various windowing systems. New built-in modules are easily written in C or C++ (or other languages, depending on the chosen implementation). Python is also usable as an extension language for applications written in other languages that need easy-to-use scripting or automation interfaces.

Machine Learning (ML) is basically that field of computer science with the help of which computer systems can provide sense to data in much the same way as human beings do. In simple words, ML is a type of artificial intelligence that extract patterns out of raw data by using an algorithm or method. The key focus of ML is to allow computer systems to learn from experience without being explicitly programmed or human intervention.

We are living in the 'age of data' that is enriched with better computational power and more storage resources,. This data or information is increasing day by day, but the real challenge is to make sense of all the data. Businesses & organizations are trying to deal with it by building intelligent systems using the concepts and methodologies from Data science, Data Mining and Machine learning. Among them, machine learning is the most exciting field of computer science. It would not be wrong if we call machine learning the application and science of algorithms that provides sense to the data.

What is Machine Learning?

Machine Learning (ML) is that field of computer science with the help of which computer systems can provide sense to data in much the same way as human beings do.
In simple words, ML is a type of artificial intelligence that extract patterns out of raw data by using an algorithm or method. The main focus of ML is to allow computer systems learn from experience without being explicitly programmed or human intervention.

Need for Machine Learning

Human beings, at this moment, are the most intelligent and advanced species on earth because they can

think, evaluate and solve complex problems. On the other side, AI is still in its initial stage and haven't surpassed human intelligence in many aspects. Then the question is that what is the need to make machine learn? The most suitable reason for doing this is, "to make decisions, based on data, with efficiency and scale".
Lately, organizations are investing heavily in newer technologies like Artificial Intelligence, Machine Learning and Deep Learning to get the key information from data to perform several real-world tasks and solve problems. We can call it data-driven decisions taken by machines, particularly to automate the process. These data-driven decisions can be used, instead of using programing logic, in the problems that cannot be programmed inherently. The fact is that we can't do without human intelligence, but other aspect is that we all need to solve real-world problems with efficiency at a huge scale. That is why the need for machine learning arises.

1.1.9Why & When to Make Machines Learn?

We have already discussed the need for machine learning, but another question arises that in what scenarios we must make the machine learn? There can be several circumstances where we need machines to take data-driven decisions with efficiency and at a huge scale. The followings are some of such circumstances where making machines learn would be more effective.

Lack of human expertise

The very first scenario in which we want a machine to learn and take data-driven decisions, can be the domain where there is a lack of human expertise. The examples can be navigations in unknown territories or spatial planets.

Dynamic scenarios

There are some scenarios which are dynamic in nature i.e. they keep changing over time. In case of these scenarios and behaviors, we want a machine to learn and take data-driven decisions. Some of the examples can be network connectivity and availability of infrastructure in an organization.

Difficulty in translating expertise into computational tasks

There can be various domains in which humans have their expertise,; however, they are unable to translate this expertise into computational tasks. In such circumstances we want machine learning. The examples can be the domains of speech recognition, cognitive tasks etc.

1.1.10Machine Learning Model

Before discussing the machine learning model, we must need to understand the following formal definition of ML given by professor Mitchell –

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

The above definition is basically focusing on three parameters, also the main components of any learning algorithm, namely Task(T), Performance(P) and experience (E). In this context, we can simplify this definition as –
ML is a field of AI consisting of learning algorithms that –

- Improve their performance (P)
- At executing some task (T)
- Over time with experience (E)

Based on the above, the following diagram represents a Machine Learning Model –