COVID 19: Prediction & Forecasting Analysis

1,2,3 R Usha, Pramod K, D Surbhi
1,2,3 Department of Computer Science Engineering, Priyadharshini Engineering College, Vellore, Tamilnadu
Email: revatiusha1@gmail.com

Abstract
COVID-19 (‘Coronavirus disease 2019’) is a virus infection caused by the SARS-COV-2 Coronavirus. As the number of cases of COVID-19 spread throughout the world at an alarming rate destroying millions of lives and livelihood, there seems to be no stop or any signs of slowing down, especially in India. The world’s specialized agency responsible for health management issues WHO on March 11, 2020 characterized COVID 19 as pandemic. Eventually started in China, it has now spread to the rest of the world due to its highly contagious nature. There are many arguments regarding the origin of COVID-19 and its source. Whether the virus has been engineered in a lab or created naturally by the ingestion of bats and then to humans? There are several studies which show that the nature of genome sequence of the virus does not seem to be man-made, but of a natural selection.

Keywords
COVID 19, Support Vector Machine, Regression, Estimation, Forecasts

Introduction
Coronaviruses originate from the subfamily called Orthocoronavirinae, a cluster of polymeric molecule Ribonucleic Acid (RNA) virus, predominantly a source of harm in vertebrates. They are majorly known to cause respiratory tract infections, in least cases is the occurrence of common cold. The enfolded micro-organism is a positive-sense single-stranded RNA genome and a nucleic acid core encircled by capsid of spiral uniformity. The genome size of micro-organism ranges from approximately 26 to 32 kb. They have attribute of clover leaf shaped spikes that reckon from their facet, thus an electron micrographs holds an image similar to the solar corona, thus named. COVID-19 is a new strain of virus first identified in the region of Wuhan, China in December 2019. The effects and the dangers of the virus was initially downplayed but later, the spread of the virus across the world has forced the countries to acknowledge and act with a decisive force. This virus is mainly spread in the form of droplets emanating from an infected person. These droplets fall to the ground fast but stay active ranging from a few hours to few days. A recent study [1], has proved that a combination of smooth surfaces and low temperatures help the virus to survive for extended periods of time. The virus is believed to be covered by a bi-lipid layer which encloses the virus and this layer is believed to deteriorate as the temperature increases thus decreasing the potency of the virus [2, 3]. The common symptoms of the virus are fever, cough, fatigue, shortness of breath and loss of smell and taste. Most of the cases result in mild symptoms and the more severe cases will lead to viral pneumonia which will lead to fatalities. The older population is more susceptible to the disease since the immune systems are relatively weaker compared to the younger age spectrum. We know that older people, and people with pre-existing medical conditions (such as Asthma, Diabetes and Heart-Disease) are more vulnerable to becoming severely ill with the virus, as also confirmed by the WHO. This does not mean the younger age group is safer. People are safe only when they follow good hand hygiene and good respiratory hygiene.

Problem Statement
Existing System
COVID 19, it’s almost 8 months passed bearing the rising number of cases in the virus. The Government has taken various measures to break the pandemic chain spread. Despite the continuous efforts of Government in bringing the lockdown effect, the situation is still increasing to its peak. This is majorly caused by the people who lack maintaining the ‘social distancing’ policy, wearing ‘face mask’, & constantly having ‘sanitized’ (washing hands regularly).

Proposed System
With the above rising issues stated in the existing system, the work focuses on building a platform where in the people are educated by showing them with stuffs like visualizations, forecasting models, increase of cases, deaths & recoveries on daily basis. All these concepts involve the idea of Machine Learning algorithms, the time series forecasting models, & Web Scraping.
System Design

- **Web Scraping**
  Web Scraping, a process of uprooting of broad outstretch amounts of data from the websites & having them stored in the local file of your computer, or into other formats like database, json files, or .csv files. Web scraping ingress the contents from the universal accepted Web incorporated with HTTP protocols. The basic idea of the process is just fetching & extracting the data. Fetching is computerization of a page. Web crawling is a chief integrant of this process, to schlepp pages for belatedly processing. In extraction the respective content of a page is traversed, explored, restructured, the data is reproduced to the desired locations.

- **Infographics**
  Infographics is the amalgamation of “information” & “graphics”, used inbuilding higher levels of visualization & cognitive abilities of living beings to understand well. [4] One can see the changes in the patterns & trends of the occurring data. Infographics is most used in Microsoft Excel Sheets.

- **Machine Learning**
  The analysis of thinking machine algorithms that boost spontaneously through experience. It is seen as a sub-element of artificial intelligence. In here, machine learning algorithms are used in making forecasting models using the time series data. The algorithms employed are FB Prophet Model, Arima Model, Linear Quadratic Estimation, & so on.

- **Choropleth**
  Choropleth is a type of colorful map which shows the areas (locations) shaded for the time series statistical data or the geographical characteristics of various countries. The project uses the choropleth to plot the worldwide Covid19 confirmed cases, recoveries & deaths.

System Requirements

- **Hardware Requirements**
  The hardware requirements for having this proposed system are the most which deals with the efficiency & proficiency of the equipment’s. The processor used is Intel core i5, speed is 1.1 GHz, and RAM is 8GB.

- **Software Requirements**
  Software requirements give a brief description of the software amenities that are required for the successful execution of the software with minimal errors. The software used for implementation are Jupyter notebook, Visual Studio Code for UI front end design, Microsoft Excel, Flourish Studio Visualization, Plotly Chart Studio.

System Implementation

- **FB Prophet Model**
  FB Prophet Model is one the Machine Learning Model used in forecasting with the statistical data. This model is an additive model which basically fits the data on yearly, weekly, daily seasonality, & holiday effects. These are sinewy to missing data and capable of handling the outliers.
  The working of prophet model depends on the following functions –
  \[ y(t) = g(t) + s(t) + h(t) + \epsilon \]
  where,
  \( g(t) \) stands for model trend, which means it handles the incessant increase or decrease in the data.
  \( s(t) \) stands for seasonality models which uses the *Fourier Series* concept to predict the change in the seasonal factors occurring during the year.
  \( h(t) \) stands for holidays models, which determine the effects of launch of products or event organizing which play a major role in the business development.
  \( \epsilon \) represents the discordant fallacy term.

- **Arima Model**
  Arima, abbreviated as Auto Regressive Integrated Moving Average. It is a machine learning algorithm extensively used in forecasting with the available time series statistical data. It captures the distinct level of profane structures.

- **Support Vector Machine**
  Support Vector Machine algorithm is the one which can be employed in both regression as well as categorical (binary) circumstances. Support Vectors are the co-ordinates of unique observations.[5] It is a delineation exemplar as extremity in arena, correlated to the disembodied grouping which are split by a comprehensible space that are spread far extent.

- **Polynomial Regression**
  Polynomial Regression analysis is effective in finding out the correlation between an independent variable \( x \) and a dependent variable \( y \) which is collectively compounded into an \( n^{th} \) degree polynomial in variable \( x \).[6] It is meant to
best fit in a nonlinear model

Fig 1: Sigmoid Modelling Prediction

Fig 2: Prophet Model Recoveries Forecasting

Fig 3: Prophet Model Deaths Forecasting
Results
The proposed system contains the live tracking of covid19 cases through web scraping, the choropleth graphs. The home page contains the live tracking of worldwide cases of active, cured & deaths from the web browser. The proposed system also shows the preventive measures and symptoms which indicates precautionary measures to be taken by the users.

Conclusion
The proposed system aims to provide the users with knowledge & insights of the rise in the number of confirmed cases & the deaths. The people have to take necessary precautionary measures like strictly follow the social distancing activity, avoiding unnecessary roaming, wearing face mask, keeping ourselves sanitized. The main aim of the proposed system is to provide the forecasting graphs and its corresponding values through which necessary actions can be taken. The graphs that help in predicting the forecast values indicate a sign of danger in India, the actual cases cross the line of mark from the predicted value.

References