



“CAN MACHINE LEARNING AID IN THE PREDICTION OF COVID-19”

AMAN VERMA

Department of Computer Science Engineering, Galgotias University

ABSTRACT :- This 21st Century is remarkable for witnessing so many upheaval at social, cultural and political level in the world. These changes have affected the lives of every one drastically - Nowadays. La new challenge is in front. The name of this threat is COVID-19, a highly contagious virus whose ruthless grip has left the humanity to breathe in fear and danger. The whole world is putting incredible effort to fight against the spread of this deadly disease. More than 7 percentage of the total population of world in under risk. The need for auxiliary diagnostic tool has there increased as there are no accurate automated toolkit available. In this research paper using one of type of machine learning is supervised learning popular algorithm Linear regression and support Vector machine. It is observed that performance of model I developed using Linear Regression can predict the outbreak up to 15 days in advance. However accuracy of prediction can be increased using more reliable data set this model can be I sealed up at world level.

Keywords : -

Novel Corona virus, SARS-COV-2, COVID-19
2019,-N-COV ACUTE RESPIRATORY DISEASES
PAN DEMICS .

Introduction :-

The catastrophic outbreak of SEVERE ACUTE RESPIRATORY SYNDROME - CORONA VIRUS (SAR-COV-2) COVID-19 is respiratory syndrome. The origin of COVID-19 is Undefined but it was firstly in WUHAN the capital of China HUBEI provinces and has spread globally resulting in the ongoing 2019 -2020 and became pandemic. Corona virus are group of related RNA viruses that cause disease in mammals and birds, the Corona virus is zoonotic which means the transmission of disease from animals to human, There are many existing solution to COVID -19 but there is no perfect solution to this pandemic[1] yet. In biological term the Convalescent plasma therapy is an experimental that some doctors are using for people with severe corona virus disease 2019 (COVID-19) No drug has been proved to be safe and effective for treating COVID-19, But people who've recovered proteins The body uses to fight off infectious to the disease in their blood. The blood from people convalescent plasma. Now I want to tell about the something about the machine Learning algorithm alpha-fold. Algorithm to I predict a variety, of Protein Structure associated with

COVID-19. Alpha fold which recently won the CASP B competition, (Critical Assessment of Technique for Protein structure Prediction)

for protein structure Prediction) | by, passes these Protein structure prediction that prevent distance and angle between amino acids, Scored with gradient, descent. In this research paper give early prediction of a COVID - 19 to identify potential outbreak of a 2019[5] In this paper using supervised learning popular algorithm Linear Regression can predict the Outbreak of upto 15 days in advance. However accuracy of prediction can be increased using more reliable dataset. This model can be scaled upto at world level.

RELATED WORK :-

With the help of the Python language and the Online editor Google Colab. In this Research paper use to predict the COVID- 19 cases from data given by WHO (World Health Organisation). on kaggle website -

<https://drive.google.com/drive/folders/195LWL0va1KI86qlmK5HuyskXkxmx6F9d>

In making this prediction there are six library in one package the first command first package is import pandas as pd it is used for data analysis and data analysis and data cleansing Pandas provides us with some powerful objects like Data frame. The second command is import matplotlib. plot as plt. It is used for data virtualization which means to create a virtual version of a device or resources, or even an operating System. The Third command is import seaborn as sns seaborn is an advance tool analysis data without default parameter It is used to create more attractive and informative statistical graphics • The fourth command is import Numpy as np which means. Numpy is a python package which stand for NUMERICAL PACKAGE . which contains a powerful, n-dimensional. The fifth command is from sklearn.sum import svr learning is a free software machine learning for python. It features various algorithms like Support Vector Machine

(SVM), Random forest and K- neighbours, and it also support python numerical and scientific libraries like Numpy Now, I want to share out for designing paper Importing,

- 1.1 necessary package, libraries and then Reading the data from the packages
- 1.2 "Covid = pd. Read_csv ("Covid-19data") Grouping
- 1.3 different types of cases I confirmed, Recovered, and death).
- 1.4 and then segregating country wise Analysis of COVID - 19 cases.

then Covid-19 predict on Data Analysis to find total number of cases in India and then comparing data where India has taken to get to present number of cases and creating a model Training using the current Dates it from (2/01/2020) to (24/04/2020)

Methodology :-

Dataset provided by W. H. O(World Health Organisation) on Kaggle website has been used for the empirical result analysis. The time period of date is from 2/01/2020 to 24th/04/2020) The data includes confirmed cases, death cases and recovered cases of all countries. However this paper focuses only on India's date for analysis and prediction of COVID - 19 confirmed cases patients, and will date the effect of COVID-19 less per million is less than 1, is the motivation behind this research for analysis and prediction of number of covid-19 patients in India,

```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import datetime as dt
from datetime import timedelta
from sklearn.linear_model import LinearRegression
from sklearn.svm import SVR
from statsmodels.tsa.api import Holt

covid = pd.read_csv("covid_19_data.csv")
covid.head()

```

SNo	ObservationDate	Province/State	Country/Region	Last Update	Confirmed	Deaths	Recovered
0	01/22/2020	Anhui	Mainland China	1/22/2020 17:00	1.0	0.0	0.0
1	01/22/2020	Beijing	Mainland China	1/22/2020 17:00	14.0	0.0	0.0
2	01/22/2020	Chongqing	Mainland China	1/22/2020 17:00	6.0	0.0	0.0
3	01/22/2020	Fujian	Mainland China	1/22/2020 17:00	1.0	0.0	0.0
4	01/22/2020	Gansu	Mainland China	1/22/2020 17:00	0.0	0.0	0.0

RESULTS

In India, the first case of COVID-19 was reported on 30th January 2020. During the month of there many, February, the number of cases reported was 3 and remained constant during the entire month. The major rise in the spread of disease started in the month in confirmed cases and death cases from **02/01/2020 to 24/04/2020**.

```

#Data Anlysis for India
india_data= covid[covid["Country/Region"]=="India"]
datewise_india=india_data.groupby(["ObservationDate"]).agg({"Confirmed":"sum","Recovered":"sum","Deaths":"sum"})
print(datewise_india.iloc[-1])
print("Total Active Cases",datewise_india["Confirmed"].iloc[-1]-datewise_india["Recovered"].iloc[-1])
print("Total Closed Cases",datewise_india["Recovered"].iloc[-1]+datewise_india["Deaths"].iloc[-1])

```

```

Confirmed    24530.0
Recovered    5498.0
Deaths        780.0
Name: 2020-04-24 00:00:00, dtype: float64
Total Active Cases 19032.0
Total Closed Cases 6278.0

```

Data Analysis of India from 02nd Jan 2020 to 24th April 2020

```

#Data Anlysis for Us
us_data=covid[covid["Country/Region"]=="US"]
datewise_us=us_data.groupby(["ObservationDate"]).agg({"Confirmed":"sum","Recovered":"sum","Deaths":"sum"})
print("Total Active Cases",datewise_us["Confirmed"].iloc[-1]-datewise_us["Recovered"].iloc[-1])
print("Total Closed Cases",datewise_us["Recovered"].iloc[-1]+datewise_us["Deaths"].iloc[-1])

```

```

Total Active Cases 886254.0
Total Closed Cases 151028.0

```

Prediction of covid - 19 from 02nd Jan 2020 to 24th April 2020

```

print("Basic Information")
print("Total number of Confirmed cases around the world",datewise["Confirmed"].iloc[-1])
print("Total number of Recovered cases around the world",datewise["Recovered"].iloc[-1])
print("Total number of Deaths cases around the world",datewise["Deaths"].iloc[-1])

```

```

Basic Information
Total number of Confirmed cases around the world 2811193.0
Total number of Recovered cases around the world 793601.0
Total number of Deaths cases around the world 197159.0

```

Fore cast of covid - 19 from 02nd Jan 2020 to 24th April 2020



DISCUSSION

The current trend shows there will be linear trend continued in the next few days as the control mechanism taken by Government of India are fairly strict and working well for the time being. Also with linear trends the patients getting recovered can be managed easily.. there were few challenges associated with the date The date was not stationary that showed an exponential growth after 112 days from (02/01/2020) to (24/04/2020)

Also rapid changes in number of infected cases occurred in mid-march In this study, we have only predicted the number of confirmed cases to I predict the number of health death cases. we food many problem of data stationary Using the other data related to weather, geographic layout of the country state-level population and governance parameter.

Conclusion

In this study, two machine learning model Linear regression and Support Vector Machine were used to analyse and predict the change in spread of COVID-19 disease. We have analyse in data and found out that the number of cases per million in India is less them 0.5 till 24/04/2020. (29th April, 2020)

Also we predicted the number of confirm cases of COVID-19. for the next 15days in advances

The result obtained from this study is taken from training data set up-to 24th April, 2020. Doctors, health workers and people involved in providing essential

services have to be protected in accordance with prescribed medical norms.

In Future, an automated algorithm can be developed to fetch data in regular intervals Government and hospital facilities can also maintain a check to the supply and medical assistance, isolation required for new patients



REFERENCE

1.

World Health Organization 2020. Coronavirus (COVID - 19): situation report 67.

2.

Wu, Z, and Mc Googan J.M.. 2020. Characteristic of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China summary of a report of 72,314 cases from the Chinese Center for Disease Control and Prevention, Jama.

3.

Mai, M.V. and Krauthammer, m, 2016. Controlling testing volume for respiratory viruses using machine learning and text mining. In AMEA (Annual Symposium Proceedings).

4.

Mandal, S., Bhatnagar T, Anu Arinaminpathy to Agarwal, A., Chowdhury, A., Murhikar M gangakhedhor, R.R. and Sarkar, sey 2020,

5.

By online video from the youtube video.

<https://codegna.com/python-training-...>

