**COVID-19 DATA ANALYSIS**

**Abstract**

COVID-19, an infectious disease caused by the SARS-CoV-2 virus, was declared a pandemic by the World Health Organisation (WHO) in March 2020. At the time of writing, more than 2.8 million people have tested positive. Infections have been growing exponentially and tremendous efforts are being made to fight the disease. We are going to aware people by giving them graphical representation and pie representation for the cases for the better understanding of ongoing Corona pandemic we are trying to give you the better visualization of the pandemic so everyone can easily compare the situations. For this project we are going to use different python libraries such as PANDAS, NUMPY, MATHPLOTLIB, CUFFIN etc... Expected outcomes of my project are side by side graphical representation of the pandemic of top most affected countries pictorial representation of corona cases in India with India map and so many other graphical and pictorial representation of the corona data

**INTRODUCTION**

On 31 December 2019, the first reported case in the COVID-19 outbreak was reported in Wuhan, China. The first case outside of China was reported in Thailand on 13 January 2020. Since then, this ongoing outbreak has now spread to more than 50 other countries. WHO declares COVID-19 outbreak as a Public Health Emergency of International Concern (PHEIC) by WHO on 30 January 2020 . There are over 76,000 cases of confirmed COVID-19 worldwide as of 20 February An infectious disease outbreak is the occurrence of a disease that is not usually expected in a particular community, geographical region, or time period. Typically, a rising infectious disease involves fast spreading, endangering the health of large numbers of people, and thus requires immediate action to prevent the disease at the community level. COVID-19 is caused by a new type of coronavirus which was previously named 2019-nCoV by the World Health Organization (WHO). It is the seventh member of the coronavirus family, together with MERSnCoV and SARS-nCoV, that can spread to humans. The symptoms of the infection include fever, cough, shortness of breath, and diarrhea. In more severe cases, COVID-19 can cause pneumonia and even death. The incubation period of COVID-19 can last for 2 weeks or longer. During the period of latent infection, the disease may still be infectious. The virus can spread from person to person through respiratory droplets and close contact. An ‘infodemic’ has accompanied the COVID-19 outbreak which is essentially an overabundance of information regarding the outbreak. As some of the information available to the public may not be accurate, it becomes hard for people to find reliable sources and trustworthy guidance when they need it. Because of the high demand for appropriate and trustworthy information about 2019-nCoV, WHO technical risk communication and social media teams have been working closely to track and respond to myths and rumors via its headquarters in Geneva, its six regional offices and its partners. The organization is working continuously to identify the most widespread rumors that can possibly harm the public’s health, such as inaccurate prevention measures or claims of cures. These myths are then rebutted with evidence-based information. WHO is making public health information and advice on the COVID-19. Therefore, we make Covid – 19 time and series forecasting project which will help us showing us the quick updates of the corona virus spread all among us, which will compare the spread with other countries and gives us a real and quick information of the Covid - 19 spread

**PROBLEM STATEMENT**

The pandemic has already taken grip over people’s life. Since the start of the pandemic, some countries are facing problem of ever-increasing cases. Through the data analysis of cases one can analyse how countries all over the world are doing in terms of controlling the pandemic. Analysing data leads to adapt the prevention model of the countries that are doing great in terms of lowering the graph. Predictions are made with the dataset available to the individual/country/organisations, thus helping them to decide how far they are able to control the pandemic or up to how much extent they should guide preventive measures.

Therefore, this project will help us by easily understanding the spread of Corona Virus in our country and comparing other countries with the help of graphical and pictorial representation of Time series forecasting Data Analysis Covid – 19

**PREREQUISITES**

* Python Knowledge
* Jupiter Notebook installed
* Knowledge of Python Library PANDAS
* Knowledge of Python Library NUMPY
* Knowledge of Python Library MATPLOTLIB
* Knowledge of GGPLOT
* Knowledge of Python Library FOLIUM
* Knowledge of Python Library PLOTLY
* Knowledge of Python Library CUFFLINKS

**TECHNIQUES AND CONCEPT**

**Machine Learning**

Machine learning is a field of study or process of teaching a computer to learn the fed data without being explicitly programmed. It makes computer make decisions similar to humans. Now a days, it is actively being used in various field. E.g. Medical, Industries, Astronomy etc. The major types of Machine learning are Supervised Learning, Unsupervised Learning and Reinforcement Learning.

**Supervised Learning**

The machine learning task of learning a function that can map an input data to output data and performs analysis based on that input-output pair.

**Unsupervised Learning**

A type of machine learning that draw an inference from dataset consisting of input data without labelled responses. One of the common unsupervised learning methods called cluster analysis, is used find the hidden pattern or grouping of data.

**Reinforcement Learning**

A type of machine learning that is bound to learn from experiences. There is no training dataset provided \*(such methods work in the absence of datasets). An agent in Reinforcement learning that rewards or penalise for actions done by the algorithm. The task is to find the best possible path to reach the goal.

**Some important terms**

**Data frame**

Pandas Data frame is 2D, mutable and heterogeneous tabular data structure with labelled axes. Data frame can be made of more than one series (series can only contain single list with index).

**Hypothesis**

In Machine learning, Hypothesis is a model that is used to approximate the target function and performs mapping of input with output.

**Regression**

In Machine Learning is about predicting the continuous value-based learning gained by dataset. The correctness of the output can depend on the size of dataset, features, hypothesis used etc.

**Classification**

The problem of identifying that in which sub-population a new example/observation belongs to, on the basis of learning obtained through training set containing observations along with the category they belong to.

**REFRENCES**

1. GitHub
2. YouTube
3. Coursera
4. Udemy
5. Kaggle
6. WHO (World Health Organization)
7. Datacamp
8. Edureka
9. Databricks
10. ResearchGate

Thank You