

**Artificial intelligence Project Deliverable 1**

**Members:**

**Iqra Amin (221441560)**

**Sumera Shafi (231452028)**

**Idwah Sajjad (221440716)**

**Course: Intro to Artificial Intelligence (COMP 360)**

**Section: B**

**Instructor: Dr. Aasia Khanum**

1.Project Title

**Predicting spread of covid-19 (cases or positivity ratio) using machine learning algorithms by integrating socio-economic and demographic factors**

2. Project Introduction

In this project we will be using Artificial Intelligence, mainly Machine Learning in order to check the effect of some factors like demographic and socio-economic. To do this task, we will be needing authentic data sets from around the world. We will be using data from various countries about the covid cases in the respective countries. We are going to predict the ratio of covid and the various effects on it with much more accuracy using Machine learning. This prediction as a result will help people to analyze the various aspect to be considered. For example, the prediction can be used by any organization to update their rules and regulations for the number of people to be in office at a time as the prediction will include the contribution of such factors. Our project result can help people to know at what places the implementation of measures to control spread of COVID-19 are to be taken. Using these techniques, the world can experience a good reduction in the ratio of corona positive cases.

3. Background

* How COVID-19 has devastated the world

The emergence of the coronavirus disease 2019 (COVID-19), which is caused by infection from the previously unknown severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has devastated economies and caused unprecedented challenges to healthcare and food systems around the world. Globally, billions of people have been ordered to stay at home as a result of lockdowns, while almost three million people have died (as of the end of March 2021).

Each new wave of COVID-19 has been more devastating than the one before. The COVID-19 pandemic has overwhelmed healthcare systems around the world, having a knock-on effect on the diagnosis and treatment of other diseases. Social distancing and lockdowns have reduced diagnosis rates of infectious diseases such as seasonal influenza, as would be expected with reduced social contact. But this pandemic continues to effect billions around the world with little idea of when and how a new wave of the virus strikes.

* How AI and machine learning are helping to fight COVID-19

As the world grapples with COVID-19, every ounce of technological innovation and ingenuity harnessed to fight this pandemic brings us one step closer to overcoming it. Artificial intelligence (AI) and machine learning are playing a key role in better understanding and addressing the COVID-19 crisis. Machine learning technology enables computers to mimic human intelligence and ingest large volumes of data to quickly identify patterns and insights.

In the fight against COVID-19, organizations have been quick to apply their machine learning expertise in several areas: scaling customer communications, understanding how COVID-19 spreads, and speeding up research and treatment. Machine learning is also helping researchers and practitioners analyze large volumes of data to forecast the spread of COVID-19, in order to act as an early warning system for future pandemics and to identify vulnerable populations. Machine learning is helping leaders make more informed decisions in the face of COVID-19.

* Prediction and tracking

Artificial Intelligence is also used in developing mathematical models to study the transmission rate of COVID-19. The different mathematical models that are used include: (1) SIR (Susceptible, Infectious and Recovered) model (ii) GLEaM (Global Epidemic and Mobility) model (iii) TRANSIMS (Transportation Analysis and Simulation) system and (iv) IBM (Individual Based model).

These models can theoretically predict the number of positive cases and the rate of transmission for the COVID-19 pandemic. Artificial Intelligence can easily identify the most vulnerable regions by tracking the number of confirmed cases and take necessary actions to curb the spread.

* **Spreading of Misinformation**

AI is also used to response to this crisis by controlling misinformation. Social media is now actively using personalized AI technology to mitigate the spread of false information across their platform. This becomes particularly important because a series of conspiracy theory and false medical information are circulating across the media. AI can make the process faster for screening of false information regarding COVID-19.

4. Objectives

* Collecting Datasets

The first and foremost that we need for our project are datasets with sufficient data for the number of covid cases, demographics and socio-economic indicators. We will then analyze this data to find out the how covid cases differ in different areas according to various factors.

* Applying Machine Learning Algorithms

Since our task is mainly concerned with prediction number of cases, it falls into the prediction Machine Learning Algorithm, for that we can use a number of algorithms that are mainly used for numerical prediction such as Linear Regression, Decision Trees, Neural Networks, and K-Nearest Neighbors.

5. Datasets

**Cases Data:**

The data we are going to use through this link displays the total number of cases, deaths due to COVID, tests for corona diagnosis, and vaccinations for corona being done all over the world till to date. The data is collected from various authentic organizations including “World Bank World Development Indicators, sourced from Food and Agriculture Organization and World Bank estimates”, “National government reports”, “European CDC for European countries / UK Government / HHS for the United States”, “COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University”, “Our World in Data”, “UN Population Division, World Population Prospects, 2017 Revision” and many more. These data will help us analyzing different numbers of cases and how they are a component of spread of COVID. We will be using these to make our ML algorithm more precise and accurate.

**Demographic data:**

It is one of the most important facts and figures to our project. It has the specifics of socio-economic Indicators. The data contains the economic statistics of different countries around the world from 2014 to 2019. It will assist us in our project when we will be comparing each countries’ economic particulars and COVID cases. There are different factors for which the data is available for different countries such as total population of each country for every year (population growth annual). The more helpful is the fact that it has explicit data for different age groups as it can help in finding out which age group is being affected the most and many more contributing factors will be used through this dataset. The input of data is genuine and taken from “United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects: 2019 Revision”, for some special cases, population data are derived from “Eurostat (Demographic Statistics) or National Statistical Offices”, “The World Bank. World Development Indicators: December 2020”,” International Monetary Fund”, “World Economic Outlook database: October 2020, and United Nations Statistics Division, National Accounts Main Aggregates Database (December 2020).

**Socioeconomic indicators data:**

The World Bank organization is a great source for getting our task done. It helps addressing the crisis. The datasets show a lot of helpful information such as dept in the world, particularly in emerging market and developing economies. We will be relating these to the COVID records and make our system learn how they are affecting it. The data has many factors that can be included for our project like research and development expenditure, etc. How private and public sector are being affected can also be inspected using the data. Poverty rate is also proportional to the corona case and the figures for that will be taken from here It has different statistics that are going to aid us in achieving our output.

6. Output of project

The output will be a script that will predict covid spread based on the input socio-economic and demographic factors of an area.