INT 246 : Soft Computing Techniques

Academic Task : Continuous Assessment 1

Project : Covid-19 Case and Death Prediction System

Devyansh Gupta
11806355

Puneet Karamchandani
11807623

What we had done?

To implement the Soft Computing Techniques practically we had considered a real life problem, the outbreak of the novel Covid-19 pandemic. It took birth in China and spread all over the world causing so many death casualties.

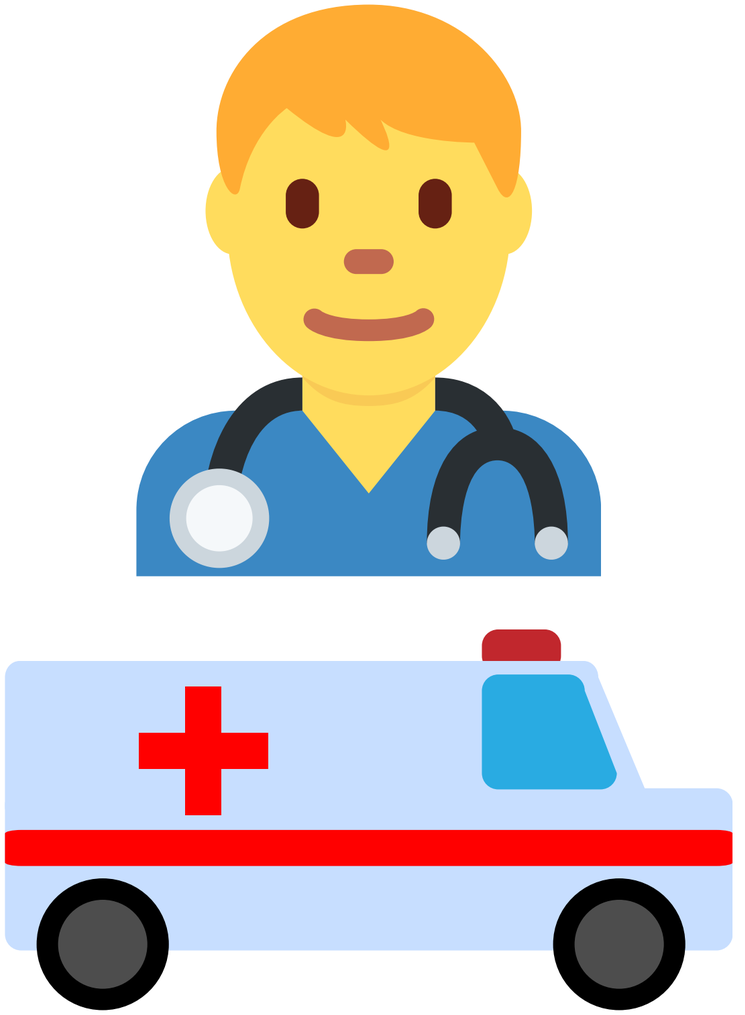
Considering the same we had taken two different data-set.

- **Infected.csv --------- Number of infected People on a particular day**
- **Deaths.csv ---------- Number of Deaths on a particular day.**

We had prepared a Machine Learning Model that can predict the number of infected people and death casualties in the future.

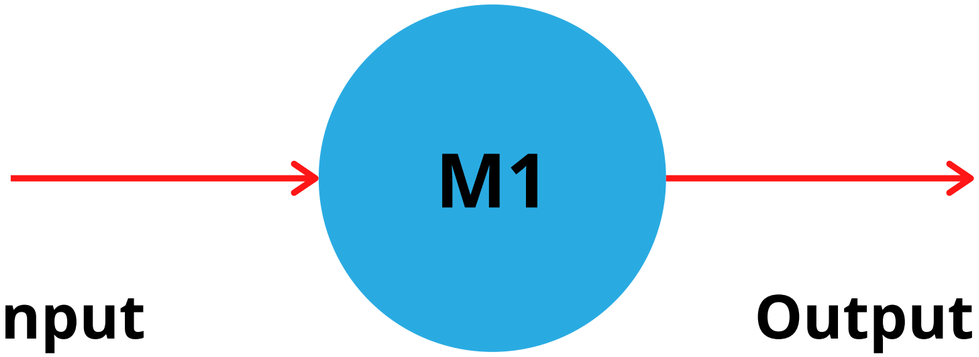
APPLICATION

This ML model will help us to predict the number of infected people and deaths in the upcoming days.
It will help us to do the required healthcare arrangements in advance so that we will be better prepared to deal with it.
Moreover the improved healthcare infrastructure and services will reduce the death casualties.



MACHINE LEARNING MODELS

Number of Infected People



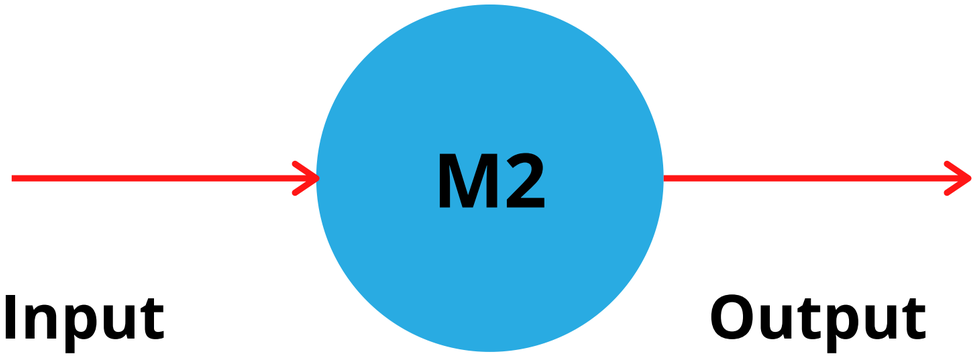
Input

Day Count

Output

Number of Infected People

Number of Deaths



Input

Day Count

Output

Number of Deaths

Defining the Problem Statement and Data Set used

Infected.csv

Date	num_of_date	num_of_patients	Date_test	num_of_date_test	num_of_patients_test
2020-01-21	1	1	2020-02-05	16.0	11.0
2020-01-22	2	1	2020-02-16	27.0	13.0
2020-01-23	3	1	2020-02-22	33.0	15.0
2020-01-24	4	2	2020-02-27	38.0	58.0
2020-01-25	5	2	2020-03-03	43.0	118.0
2020-01-26	6	5	2020-03-05	45.0	217.0
2020-01-27	7	5	2020-03-12	52.0	1663.0
2020-01-28	8	5	2020-03-23	63.0	43847.0
2020-01-29	9	5	2020-03-30	70.0	161807.0
2020-01-30	10	5	2020-04-05	76.0	312237.0
2020-01-31	11	7	2020-04-10	80.0	460252.0
2020-02-01	12	8	2020-04-19	89.0	657996.0
2020-02-02	13	8	NaN	NaN	NaN

Training Data | **Testing Data**

Deaths.csv

Date	num_of_date	Deaths	Date_test	num_of_date_test	Deaths_test
2020-03-02	1	6	2020-03-18	17.0	171.0
2020-03-03	2	9	2020-03-24	23.0	957.0
2020-03-04	3	11	2020-04-01	31.0	6394.0
2020-03-05	4	12	2020-04-11	41.0	23843.0
2020-03-06	5	14	2020-04-22	52.0	47894.0
2020-03-07	6	17	2020-04-28	58.0	59266.0
2020-03-08	7	21	2020-04-05	35.0	11793.0
2020-03-09	8	22	2020-04-16	46.0	34619.0
2020-03-10	9	30	NaN	NaN	NaN
2020-03-11	10	36	NaN	NaN	NaN

Training Data | **Testing Data**

Feature Extraction

In the above data sets the only feature variable is the Day Count.

- no_of_date
- no_of_date_test

INDEPENDENT VARIABLES

The Target variable or the Output variable are:

- no_of_patients
- no_of_patients_test
- no_of_deaths
- no_of_deaths_test

DEPENDENT VARIABLES

OBJECTIVE

Depending on the Day Count predict the number of infected people and the number of deaths.

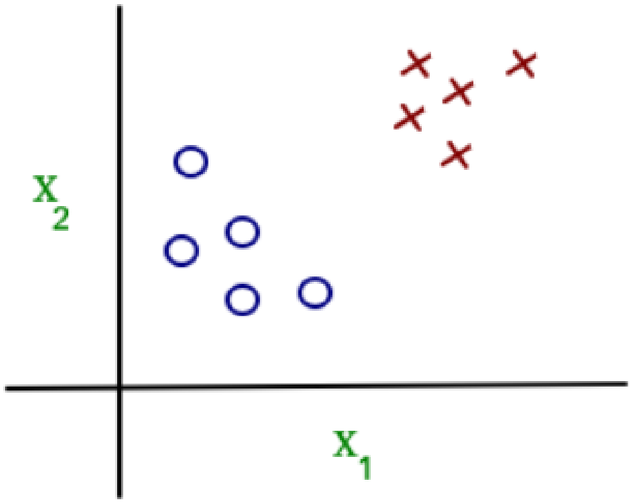
Type of Learning Used

Here we are using Supervised Learning as we are providing the Day Count along with the Number of Infected People and the Number of Deaths.

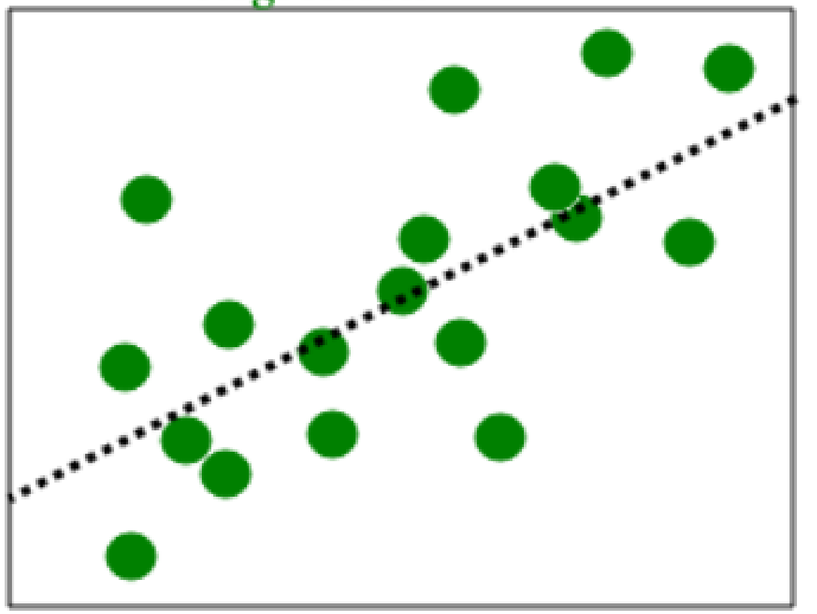
Type of Problem

The outputs are continuous rather than discrete. Thus we are having a Regression Problem.

Supervised Learning



Regression



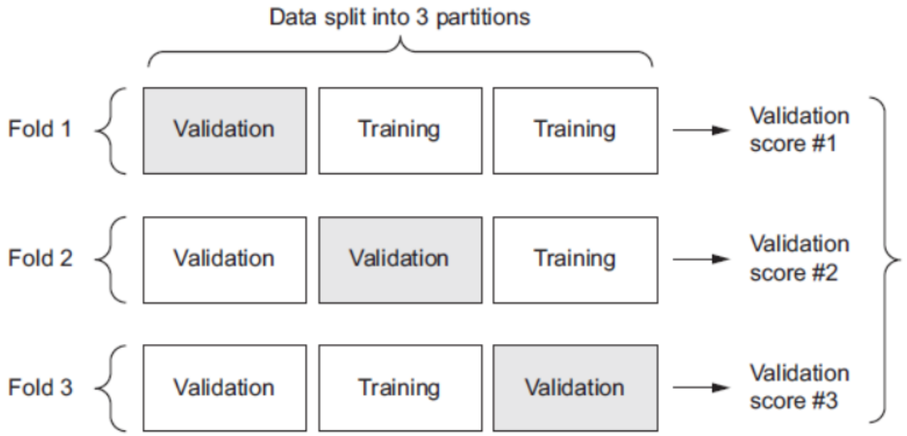
Measure of Success

Regression problems use certain evaluation metrics such as mean squared error (MSE).

Evaluation Protocol K-Fold Validation

K-Fold consists in splitting the data into K partitions of equal size. For each partition i, the model is trained with the remaining K-1 partitions and it is evaluated on partition i.

The final score is the average of the K scored obtained. This technique is specially helpful when the performance of the model is significantly different from the train-test split.



We have used 3 different ML techniques for implementing each ML model

- Linear Regression with log values
- Support Vector Regression (SVR)
- Artificial Neural Network (MLP Regressor)

The MLP Regressor model is found to give the best results