# Project Report

**On**

# Corona Virus (COVID-19) Prediction System

## BACHELOR OF TECHNOLOGY

**in**

## COMPUTER SCIENCE AND ENGINEERING

By

**Devyansh Gupta (11806355)**

**Puneet Karamchandani (11807623)**

**Section: KM110**


**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab(India)

# INTRODUCTION

*Today, the world is facing a widespread spread of the corona virus and an increase in the number of infections and deaths due to this disease, wishing safety and health for all.*

## What is Corona Virus (COVID-19)?

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus.

Most people infected with the COVID-19 virus will experience mild to moderate respiratory illness and recover without requiring special treatment. Older people, and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness.

The best way to prevent and slow down transmission is to be well informed about the COVID-19 virus, the disease it causes and how it spreads. Protect yourself and others from infection by washing your hands or using an alcohol based rub frequently and not touching your face.

The COVID-19 virus spreads primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes, so it’s important that you also practice respiratory etiquette.

# PROJECT

## Aim

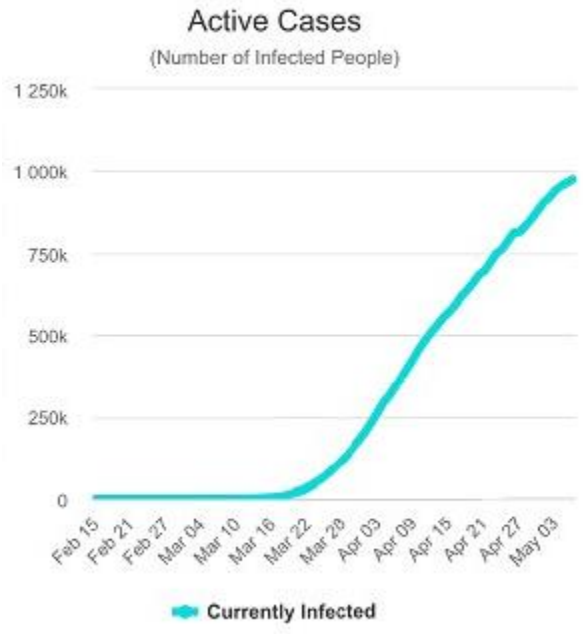
Predict using ML the number of infected people and the number of deaths of coronavirus on the upcoming days.

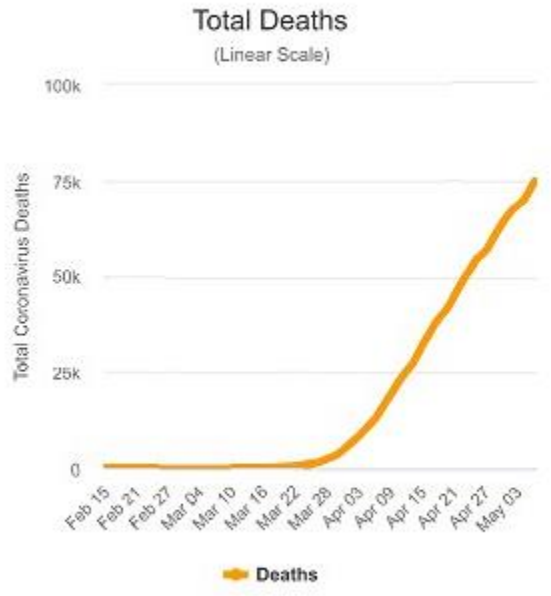
## Technologies

- Python 3.8.0
- Jupyter Notebook

## Data Set

We used data for the **India** to train our model and we got data for **95 days** that includes the date of the day and the number of people with the disease and for the data of the dead we got data for **60 days** and also includes the date of the day and the number of deaths due to disease.



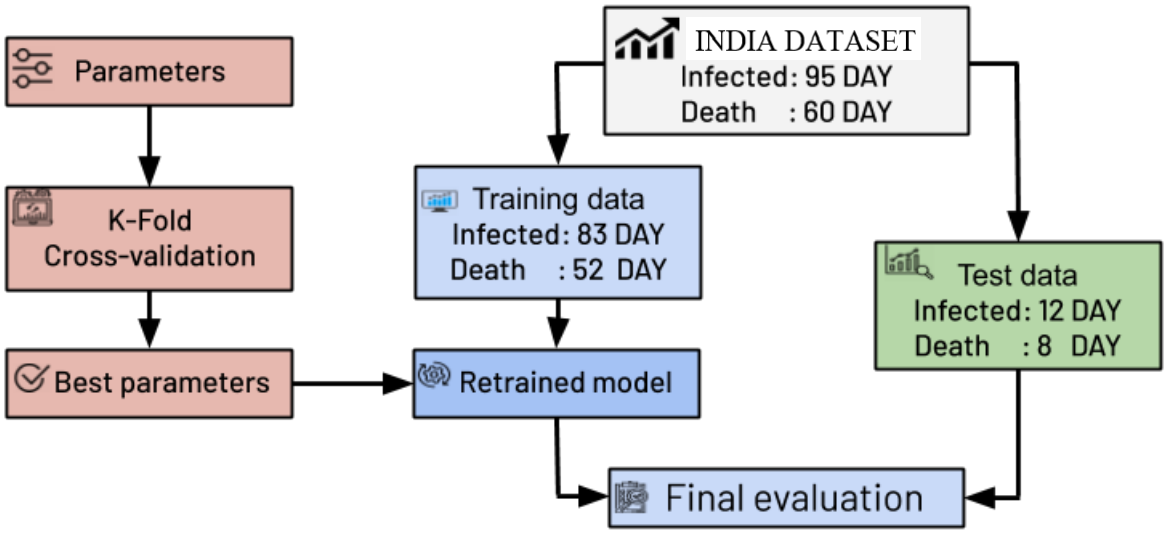



## We have 3 types of ML models

- *Linear regression with log values*
- *Support Vector Regression (SVR)*
- *Artificial Neutral Network*

## Work methodology :

- The data was divided into two sections, for **training** and **testing** .
- We adopted the **K-Fold Cross-validation** methodology when selecting the best parameter for a single model. The following figure shows the work methodology:



# INFECTED

When looking for data for the **India - Active Cases (Total)**, we obtained data for 95 days, starting **21/1/2020** , and the number of injured was **one** and we assumed on the first day. The data ended on **25/4/2020**, today is 95, and the number of injured is **788,233**.

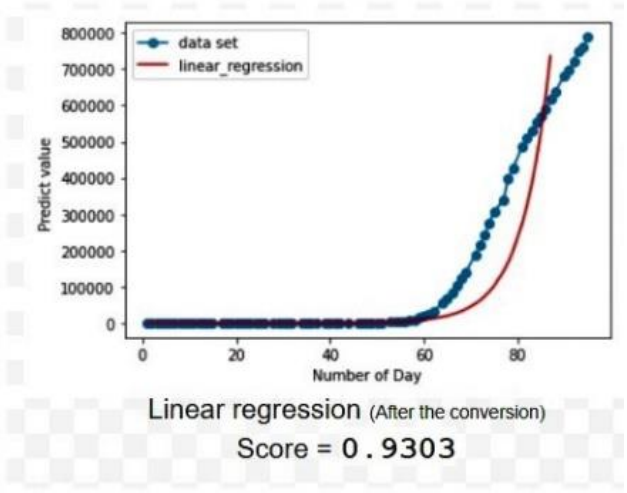
We took data for **12 days** randomly for the final **test and evaluation stage**. After that we trained the different models, got a **final evaluation** and represented them with **score for each model**.

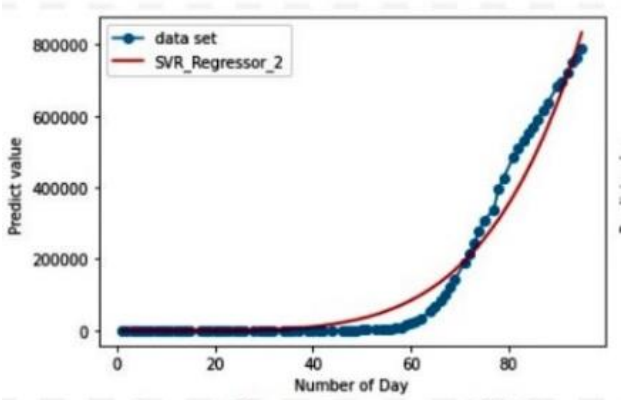
*The following figure represents the **curve for each model** in addition to the **score** obtained:*

**Note:** The data were represented using two methods, the first is a **logarithmic representation** and the second is a **Linear representation (Exponential)**. The models were trained in both methods, and we chose the best among them for each model individually.

- When looking at the Linear Regression curve, we used logarithmic representation of data when training it, and the resulting curve was converted to the exponential representation.

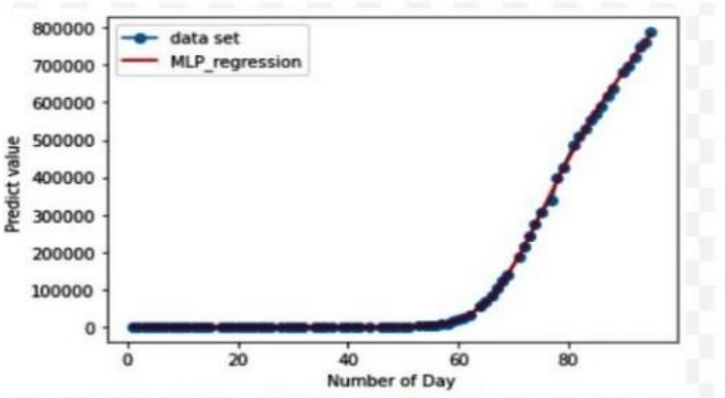
## Linear representation (Exponential)






Support Vector Regression (SVR)

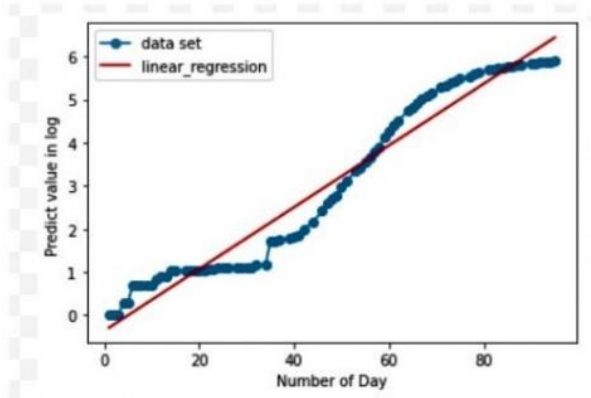
Score = 0.9587




Artificial Neural Network (MLPRegressor)

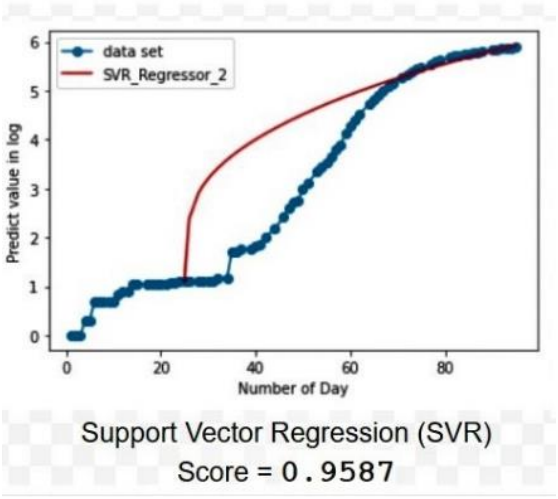
Score = 0.9984

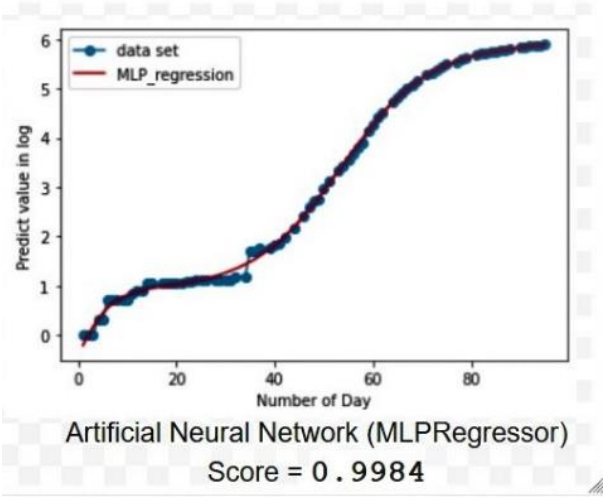
## Logarithmic Representation




Linear regression (After the conversion)

Score = 0.9303




After looking at the results, we realize that **the best model is the Artificial Neutral Network ,** and it has the best score !

# DEATHS

When searching for data for the **India**, we obtained data for 60 days, starting on **2/3/2020,** and the number of dead was **6** and we assumed on the first day. The data ended on **30/4/2020**, today 60, and the number of dead **63856**.

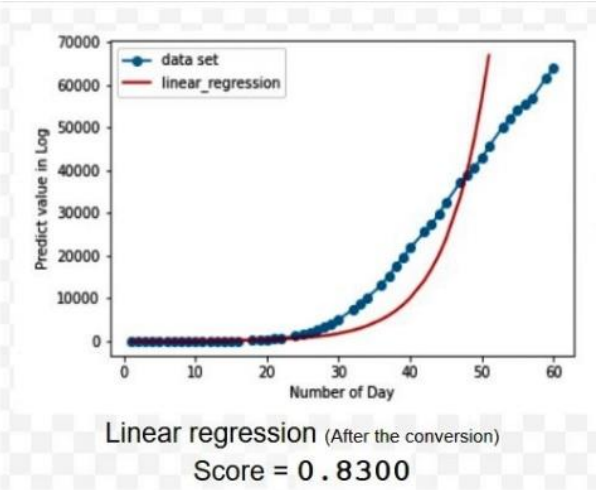
We took data for **8 days** randomly for the final **test and evaluation stage**. After that we trained the different models, got a **final evaluation** and represented them with **score for each model**.

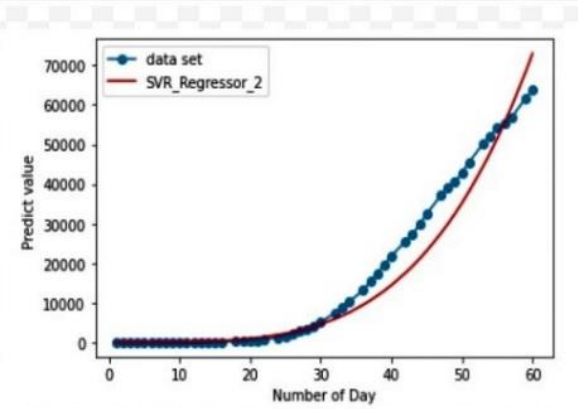
*The following figure represents the **curve for each model** in addition to the **score** obtained:*

**Note:** The data were represented using two methods, the first is a **logarithmic representation** and the second is a **Linear representation (Exponential)**. The models were trained in both methods, and we chose the best among them for each model individually.

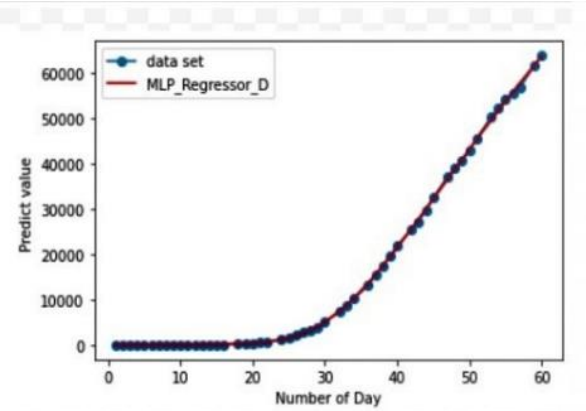
- When looking at the Linear Regression curve, we used logarithmic representation of data when training it, and the resulting curve was converted to the exponential representation.

## Linear representation (Exponential)



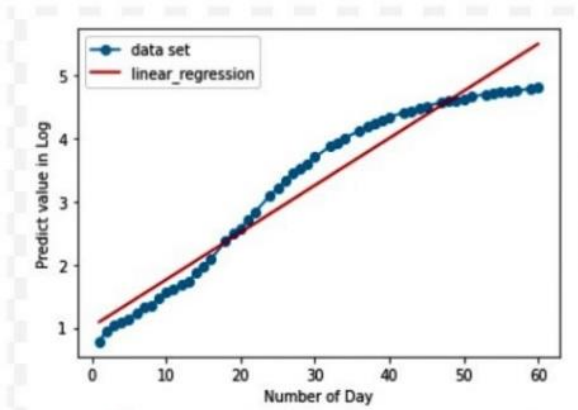



Support Vector Regression (SVR)
Score = 0.9342

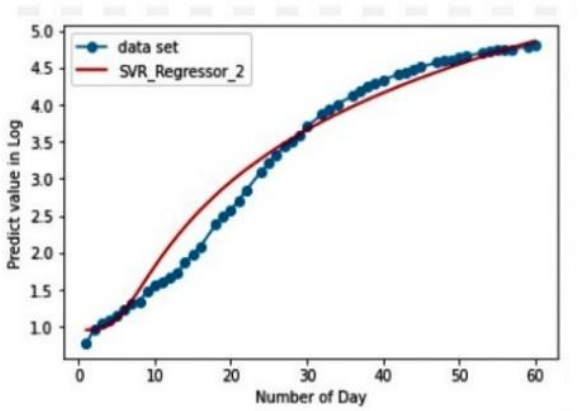


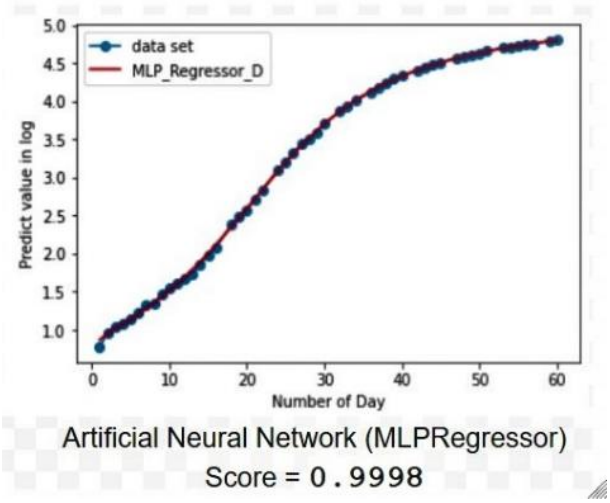

Artificial Neural Network (MLPRegressor)
Score = 0.9998

## Logarithmic Representation




Linear regression (After the conversion)
Score = 0.8300




After looking at the results, we realize that **the best model is the Artificial Neutral Network ,** and it has the best score !