**COVID-19 CASES PREDICTOR**

Submitted in Partial Fulfillment of the Recruitment for the Degree of

Bachelor of Technology

In

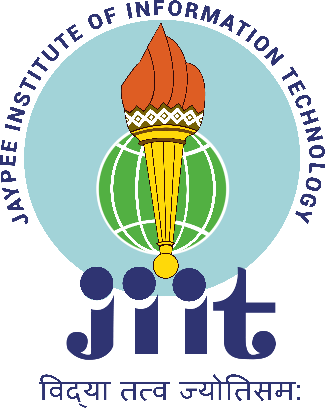
Computer Science & Engineering

2021

Submitted to

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY

NOIDA (U.P)



Submitted by

ARSH THAKUR (19103132)

CHINMAYA PARAKH (19103140)

Under the guidance of

PARUL AGARWAL

DEPARTMENT OF COMPUTER SCIENCE

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY

YEAR – 2021

**Problem statement**

Making fast and accurate decisions are vital these days and especially now when the world is facing such a phenomenon as COVID-19, therefore, counting on current as well as projected information is decisive for this process.

This project aims to forecast and compare the total number of COVID-19 cases in multiple countries in the next n days given current statistical information, hoping it can be used as foundation support to take action in this difficult scenario. To accomplish this objective, polynomial regression has been applied to the model.

**Motivation**

Since the World Health Organization declared a pandemic for COVID-19 virus on March 11, 2020, the Americas, Europe, South-East Asia and Eastern Mediterranean regions have the most documented cases. Globally, nationally, and at every sub-governmental level, there is a need to monitor the current caseload and project the rate and nature of the spread to guide public health awareness, preparedness, and response.

Hence the idea of this project is to gather current statistical information about the total number of cases and predict the dynamics of COVID-19 infections during the current pandemic.

**Literature**

Machine learning plays a major role in disease prediction and analysis. Though COVID-19 is a new disease still works have done in COVID-19 prediction using machine learning. Continuous contribution is going on in this field of research.

Ahmed Hamad et.al has proposed an accurate classification model for COVID-19 using KNN variant. They have used the dataset collected from Italian society of medical and intervention radiology society. A clear comparative analysis has given among Modified KNN (MKNN), KNN for imperfect data (KNNimp) and cost sensitive KNN (csKNN) and the KNN variant has achieved most efficient performance.

In another study authors have collected blood samples from 404 patients in Wuhan China for identification of predictive biomarkers of the disease. An artificial intelligence based COVID-19 mortality prediction model has been proposed by authors. They have used Artificial Neural Networks, Random Forest, Decision Tree, Logistic Regression, and K-Nearest Neighbor (KNN) for the mortality prediction.

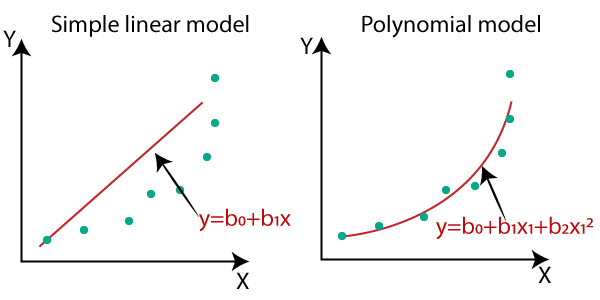
Another very effective model has been proposed to predict the need for ventilation in COVID-19 patients using machine learning framework.

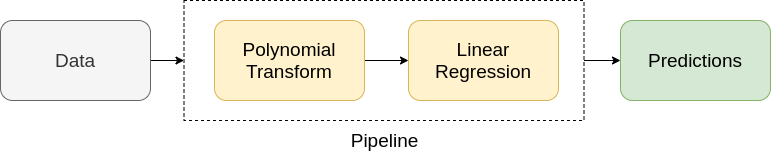
Mohammadreza Nematiet al. has proposed a model which predicts discharge time of COVID-19 patient with survival analysis.

**Model Used**

**Polynomial Regression:**

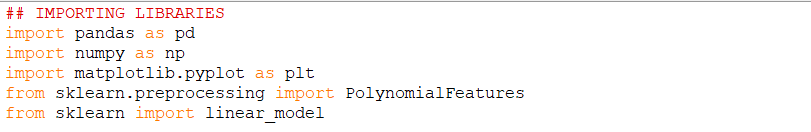
In statistics, polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modeled as an nth degree polynomial in x. Polynomial regression fits a nonlinear relationship between the value of x and the corresponding conditional mean of y. Although polynomial regression fits a nonlinear model to the data, as a statistical estimation problem it is linear, in the sense that the regression function is linear in the unknown parameters that are estimated from the data. For this reason, polynomial regression is considered to be a special case of multiple linear regression.



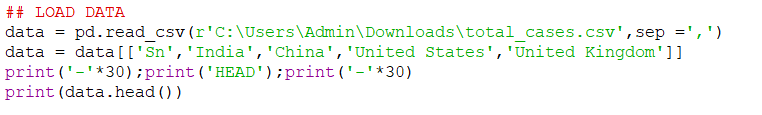


**Implementation**

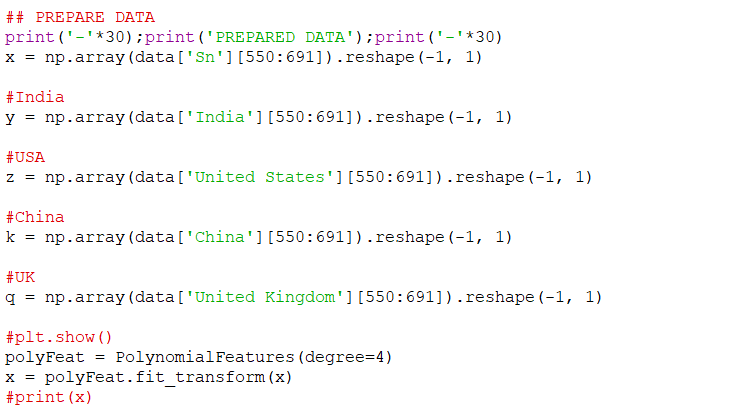
**Libraries Used**



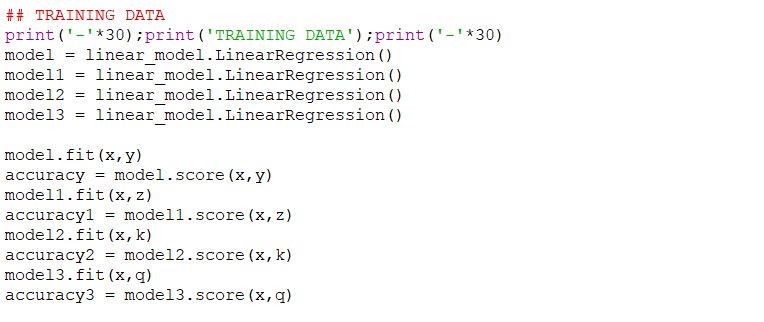
**Loading Data**

****

**Preparing Data**

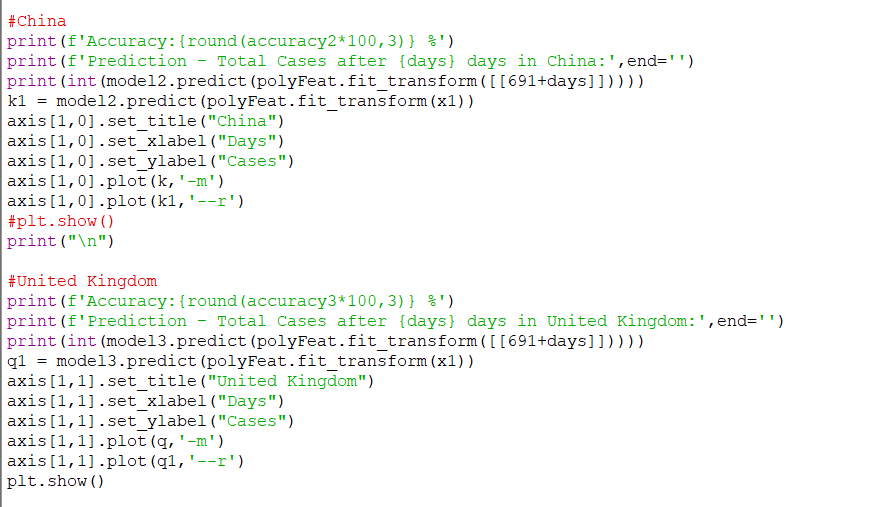
****

**Training Data**

****

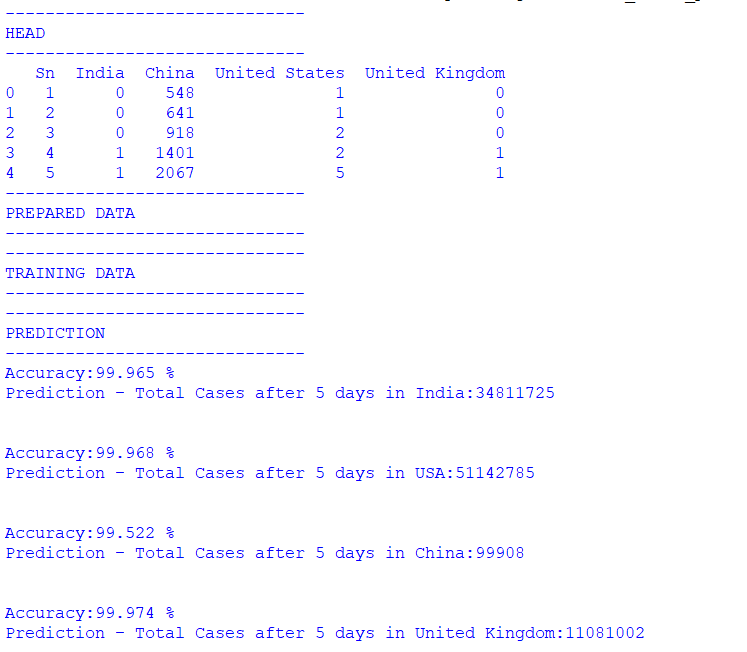
**Prediction**

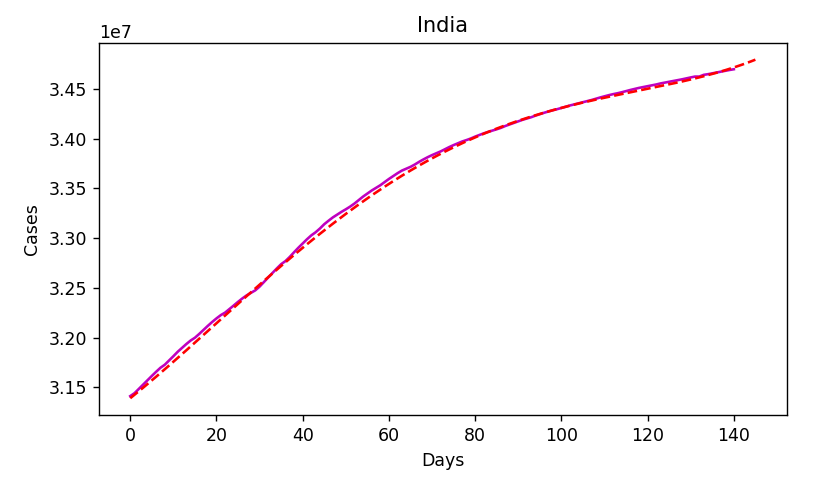
****

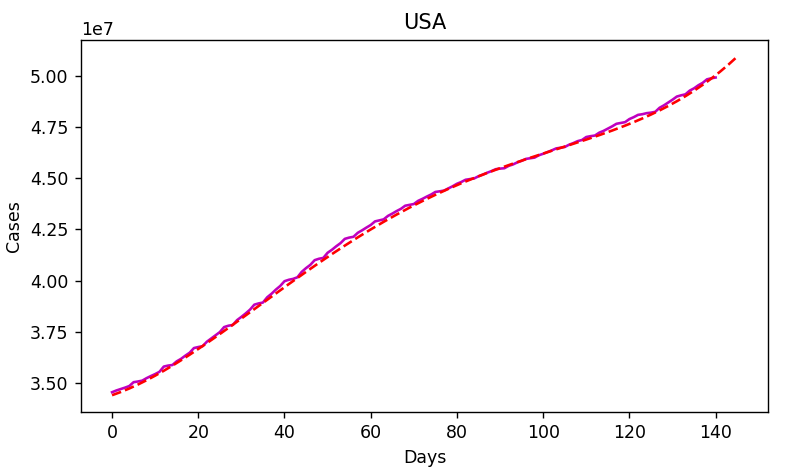
****

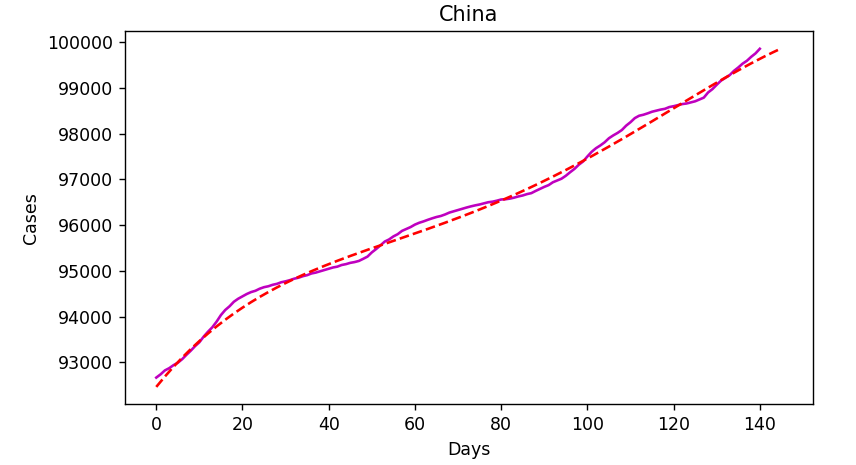
**Results**

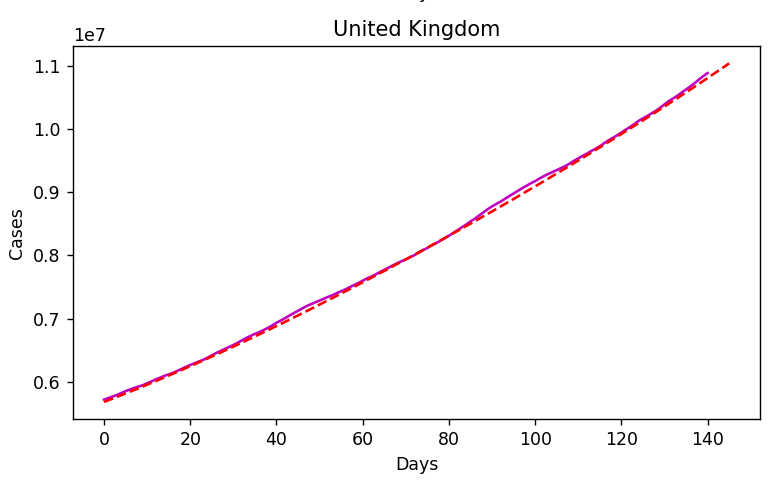
Forecasting of total number of COVID-19 cases using Polynomial Regression is shown below:

****

****

****

****

****

**Conclusion**

Our Covid-19 case predictor has been successful in predicting total number of cases of multiple countries. We have studied and learnt forecasting data using Polynomial Regression model.

Accuracy of prediction:

* India: 99.96%
* USA: 99.97%
* China: 99.52%
* United Kingdom: 99.97%

**References**

[Polynomial regression - Wikipedia](https://en.wikipedia.org/wiki/Polynomial_regression)

[Polynomial Regression. This is my third blog in the Machine… | by Animesh Agarwal | Towards Data Science](https://towardsdatascience.com/polynomial-regression-bbe8b9d97491)

[Plot multiple plots in Matplotlib - GeeksforGeeks](https://www.geeksforgeeks.org/plot-multiple-plots-in-matplotlib/)

[JPCSJ17971011-.pdf (iop.org)](https://iopscience.iop.org/article/10.1088/1742-6596/1797/1/012011/pdf)

[(PDF) Accurate Classification of COVID-19 Based on Incomplete Heterogeneous Data using a KNN Variant Algorithm (researchgate.net)](https://www.researchgate.net/publication/341170227_Accurate_Classification_of_COVID-19_Based_on_Incomplete_Heterogeneous_Data_using_a_KNN_Variant_Algorithm)