BYTE PANACHE COMPETITION

FOUNDATIONS OF MODERN MACHINE LEARNING

OBJECTIVES

To produce a reasonable estimate of future COVID-19 cases using Regression techniques. More specifically: **Linear Regression**, **Polynomial Regression**, and **Ridge Regression**.
Key topics:

1. Data Analysis
2. Linear Regression
3. Polynomial Regression
4. Ridge Regression

MATERIALS & METHODS

The following materials were required to complete the research:

- Jupyter Notebook
- COVID-19 dataset
- Python libraries for data manipulation and visualization
  1. Pandas
  2. SciPy
  3. Matplotlib
  4. Scikit Learn

Regression approaches and models are employed to understand the effect of COVID-19 and eventually predict the future cases.
Linear Regression analysis, Polynomial Regression analysis and Ridge Regression analysis have been modelled and effect of COVID-19 has been analysed. The variables considered are: confirmed cases and number of days.

INTRODUCTION

Machine Learning enables computers to learn without being programmed. Machine learning is divided into two types: supervised learning and unsupervised learning. In supervised learning, new instances are mapped by analysing the training data's input-output correlation. To forecast COVID-19 cases, regression approaches and models are employed in this project.

RESULTS 2

Linear Regression:
Mean Absolute Error and Mean Squared Error

	MAE	MSE
1	181791.4230	33128805402.9171

Table 1: Linear Regression

Polynomial Regression:
Mean Absolute Error and Mean Squared Error

	MAE	MSE
1	31989.7526	1314905592.3565

Table 2: Polynomial Regression

Ridge Regression:
Mean Absolute Error and Mean Squared Error

	MAE	MSE
1	181588.7612	133054982293.9402

Table 3: Ridge Regression

RESULTS 1

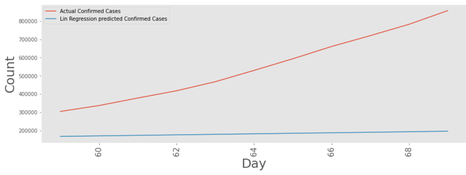


Figure 1: Linear Regression

The results of Linear Regression are shown in the figure (1). It is observed that the actual confirmed cases which are depicted with the red line, are constantly provided the number of days increased. While the blue line, which depicts the confirmed cases predicted by Linear Regression is steady with no deviations.

Polynomial Regression results are depicted in the figure (2). It is seen that the actual confirmed cases, denoted by the red line, are consistently increased as the number of days increases. The blue line, which depicts the confirmed cases predicted by Polynomial Regression, predicts almost the same results as the actual cases with slight differences.

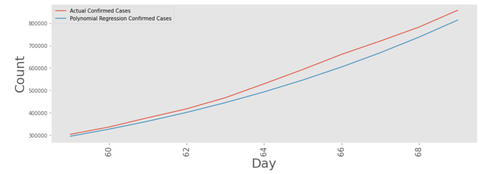
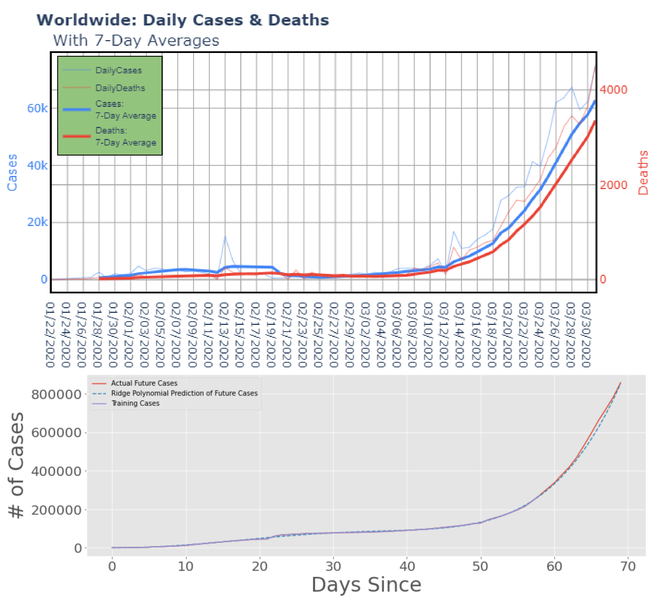


Figure 2: Polynomial Regression

CONCLUSION



- The findings may differ based on the data set used. This approach may be extended to larger datasets. The same approach of the analysis can also be performed on confirmed and recovered instances.
- According to the analysis made, the number of confirmed cases would rise in relation to the number of days.
- Furthermore, more attributes can be included in the study in order to add more accuracy during the process.
- From Mean Absolute Error and Mean Squared Error values, Ridge Polynomial Regression Model seems to be the best model.
- Further investigation may involve the analysis on a different dataset by predicting the number of cases in future and how the mortality rate varies with the rise in the number of cases.

REFERENCES

[1] Raji P. Deeba Lakshmi. *Covid-19 Pandemic Analysis using Regression*. medRxiv, 7th edition, 2020.

[2] Mohammed Ayoub K. Rijwan Khan. Performance evaluation of regression models for covid-19. August 2021.

FUTURE RESEARCH

This experiment on predicting the future cases of COVID-19 which is taken with respect to the parameters: Count of cases and Number of days has given depicted the considerable results with adjoin very close to the actual data set. The same approach can be further applied in various aspects of finding relativity of any different parameters and also various large datasets. The idea of regression technique is considerably the most intuitive way of predicting, given any type of data.

CONTACT INFORMATION

Group members (Group 9):
Thota Varshith, Shaantanu Agarwal, Mungani Sreeja, Pradeep Kumar Pal, Shubham Sharma, Gokulakannan S