

OBJECTIVES

The project aims to make predictions for the covid-19 virus which can be very helpful in the early detection of virus and containing the spread of virus. The predictions will be helpful for making public policies and estimating the economic variations over time. The project aims to:

1. Understand the current spread of covid19
2. Model the virus trend using various methods
3. Predict the number of people the virus may affect in the future

MATERIALS & METHODS

To complete the research a data set containing information about covid-19 outbreaks was required along with the information regarding the date and region of outbreak. The data also contains separate information about confirmed cases, deaths and recovered cases.

The following models were applied on the data:

- Linear Regression
- Polynomial Regression
- Ridge Regression
- Random Forest Regressor
- SVM Regressor

The first 85% of the data sorted according the date was used for testing while the latest 15% samples were used to test the model. Various algorithms were trained and were used for predictions on future dates contained in the test samples. The models were evaluated using mean squared error and mean absolute error metrics.

REFERENCES

- [1] Mishra D. Bhardwaj S. Aggarwal M. Painuli, D. Forecast and prediction of covid-19 using machine learning. April 2021.
- [2] McNulty A Radcliff TA Cote MJ et al. Zhao H, Merchant NN. Covid-19: Short term prediction model using daily incidence data. *PLoS ONE*, April 2021.

INTRODUCTION

COVID-19 is a deadly widespread virus that has affected people all over the world. With its origin in Wuhan city, China it started to spread and reached almost every corner of the world. This had a huge impact on people and one of the worst blows to the economies over the world. As prevention is the only cure for the disease, various analytics and predictions have come up. This approach is a part of the effort to deal with covid-19 by using machine learning to predict its future trajectory.

RESULTS 2

The predictions made by the model are quite accurate and can be used as an indicator for deciding the future course of actions that may be affected by the rise of number of cases of the virus. The predictions on the test set show steep increase.

Day	Predicted Cases	Actual Cases
0	298348	304528
1	331417	337020
2	368468	378287
3	409833	417966
4	455863	467594
5	506922	529591
6	563394	593291
7	625680	660706
8	694196	720117
9	769380	782365
10	851687	857487

Table 1: Predictions

FUTURE RESEARCH

The project can be further used to predict outbreak according to region. The active cases can be segregated region-wise to predict the future spread within the region. Also, by monitoring the

RESULTS 1

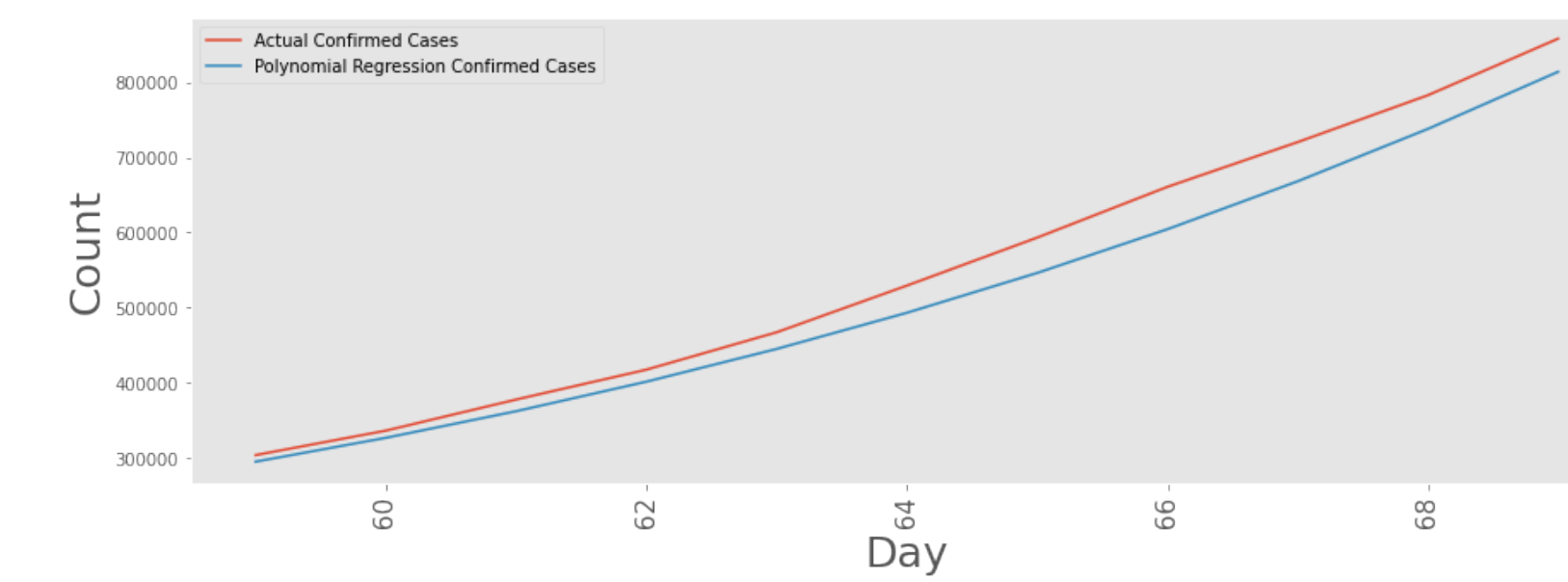


Figure 1: Polynomial Regression

Ridge regression is an effective method of generalizing the model. Applying this on polynomial regression model brings up a model with excellent accuracy and minimal error. It is more efficient and accurate than random forest and support vector machines with polynomial kernel.

Polynomial regression is found to be a good fit for the given data. This technique fits a polynomial curve to the data and the optimal value of the polynomial is found to be 5. In comparison, Linear Regression when used on the same data performs very poorly by under predicting the values. Also, random forest regressor fails to achieve a similar accuracy.

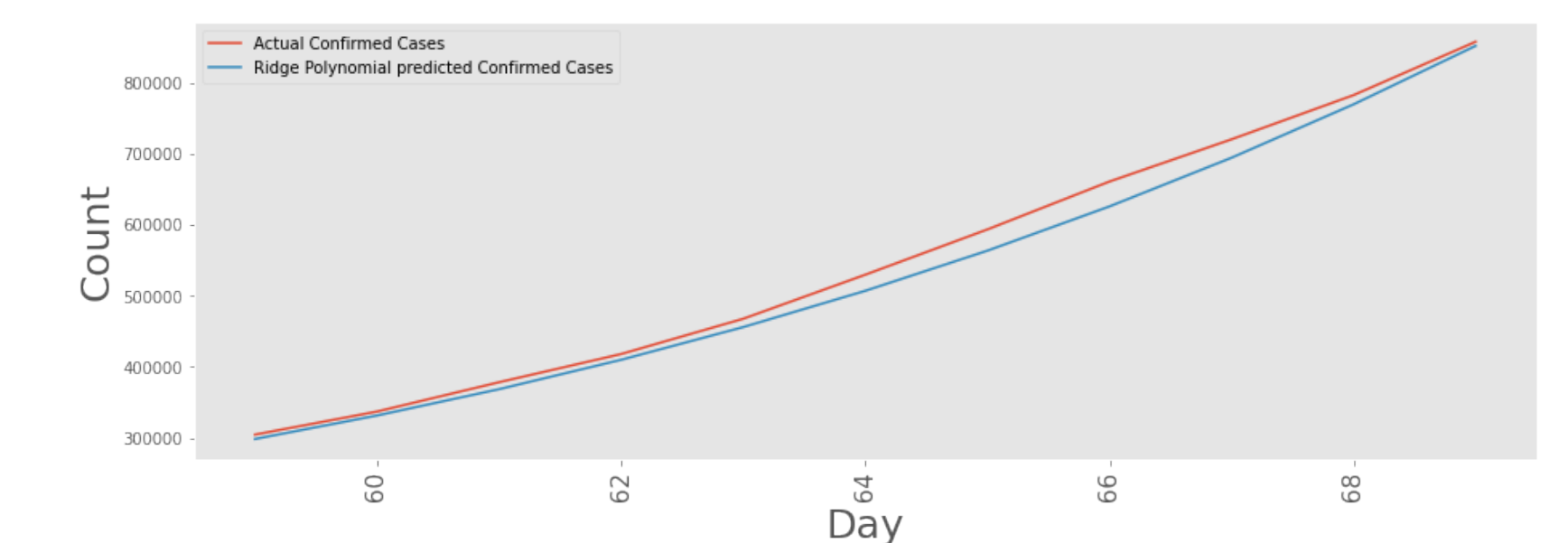


Figure 2: Polynomial Ridge Regression

CONCLUSION

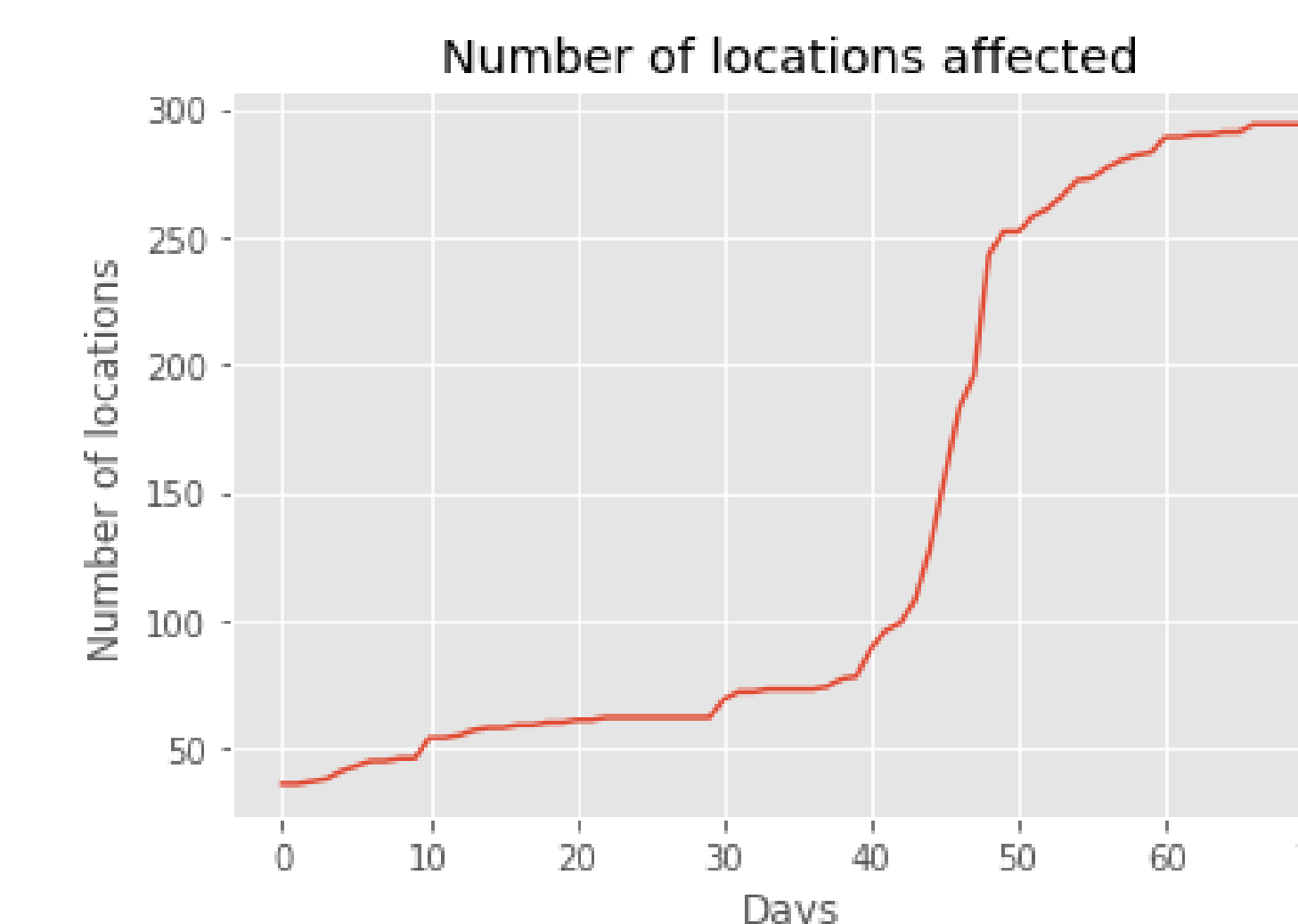


Figure 3: Spread among locations

For the set of locations given in the data, the number of covid cases have been rising rapidly and the virus has already spread to most of the considered regions (as evident in the graph). Thus, it can be estimated that the number of active cases is set to grow in the short term especially in the regions where the first case has been detected recently. The model presented here can easily fit to suit the needs of the authorities to give more specific predictions. Data containing demographics can be used to get more precise estimates about a certain class of people or those living in a certain region.

CONTACT INFORMATION

Email harshitg2306@outlook.com
Phone +91 9810906860