BYTE PANACHE COMPETITION

FOUNDATIONS OF MODERN MACHINE LEARNING

OBJECTIVES

In this project, we used the **concepts of regression and regularization to predict early Covid-19 cases**. We used **linear regression**, **polynomial regression** and **ridge regression** to obtain a reasonably good estimate of the future cases. We also experimented with **hyperparameters** to obtain better results.

1. Compared all the above mentioned regressions technique to get better results.
2. Used Visualization techniques to make data more understandable.
3. We used metrics like growth factor to understand the task in hand .

INTRODUCTION

Covid-19 is a pandemic which has affected all parts of the world which can be controlled only by maintaining social distancing, proper hygiene, wearing mask, hand sanitation and to a extend by wearing face shield. Awareness among citizens and behaving as responsible citizens is very important in controlling this disease. Contact tracing plays an important role in controlling this pandemic. This project deals with the effect of Covid-19 in various regions of China and also forecasts its effect using machine learning techniques . This helps in understanding the much-affected regions and helps to understand the infrastructure requirements to handle this pandemic efficiently.

MATERIALS & METHODS

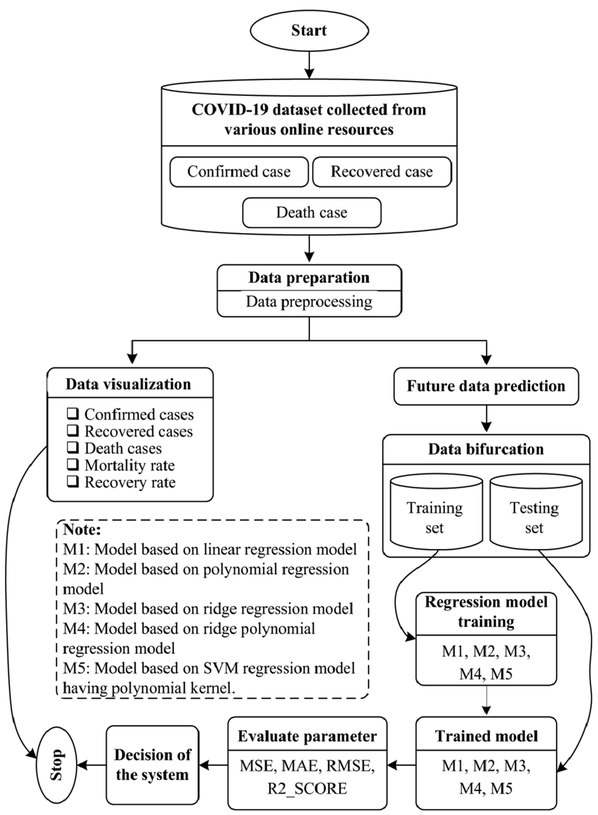

Figure 3: Experimental Work Flowof the Model.

In this model we have utilized the overall details of coronavirus from January 22, 2020, to July12, 2021, was assembled from the online resources like Kaggle, Weka 3.8.4 and Orange . The datasets give us the quantity of affirmed cases, recouped cases, and passing cases everywhere throughout . The datasets are accessible in time arrangement position with death, month and year so fleeting parts are not neglected. We isolated our datasets into a preparing set (85 perc.) on which our model readied and a testing set (15 perc.) to test the exhibition of our framework.

The experimental work flow diagram of the work is shown in Fig. 3. From the Fig. 3, it has been found that the complete work divided into two sections.

1. Section one is used for data visualization and section two is used for future data prediction. For data visualization, progress bar, recovery rate and mortality rate are shown.
2. In Section 2, linear, polynomial, ridge, ridge polynomial and support vector machine (SVM) having polynomial kernel function is used for future case prediction. Each model is evaluated using MSE, MAE, RMSE and R2-SCORE parameters.

RESULTS

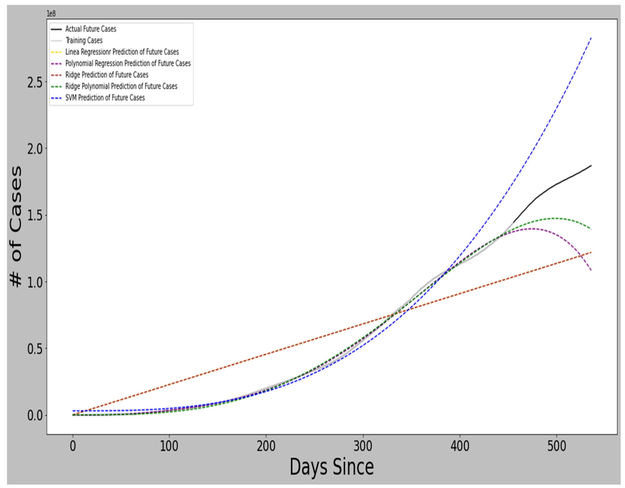

Figure 1: Days Since : Future Cases Prediction

- Figure 1(Days Since) shows the performance of all the models on the basis of trained data.

We also concluded that only polynomial ridge model comes out with better result in case of actual death pattern .

- Here, Red line : actual death cases
- Black line : the training cases.
- Yellow line : the LR
- Brown line: the ridge regression
- Purple line : the polynomial prediction of death cases.
- Green line : the polynomial ridge prediction of death cases.
- Blue line: the SVM prediction of death cases

From this we have examine that polynomial ridge, SVM and polynomial predicts good result and following the pattern which are near to actual confirmed cases. LR and ridge predict similar result and does not follow the pattern of actual cases.

- Here, black line represents actual future cases and grey line represents the training cases.
- Yellow line represents the LR and Brown line represents the ridge regression and the future prediction of both the models are very much similar and overlap each other.
- Purple line represents the prediction of confirmed cases using polynomial regression.
- Green line represents the prediction of confirmed cases using polynomial ridge regression.
- Blue line represents the prediction of confirmed cases using SVM.

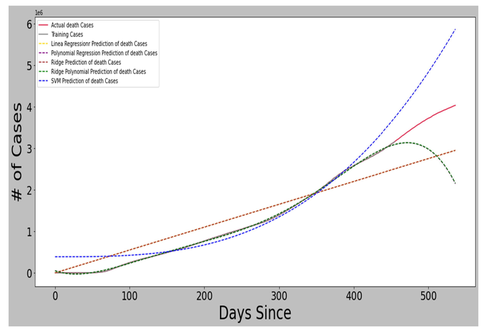

Figure 2: Days Since :Performance of Different Regression Models on Prediction of death cases.

CONCLUSION

From our investigation we concluded that polynomial ridge is the best model for predicting confirmed cases, linear regression model for recovered cases and SVM regression model for death cases.

FUTURE RESEARCH

There is no remarkable difference in the prediction accuracy of algorithms with and without feature selection algorithms, so there we need to find out the optimal features for reproduction rate.

CONTACT INFORMATION

LinkedIn linkedin.com/in/prince-dwivedi/
Email fmml20210418@ihub-data.iiit.ac.in
Phone +91 7440 324 634