

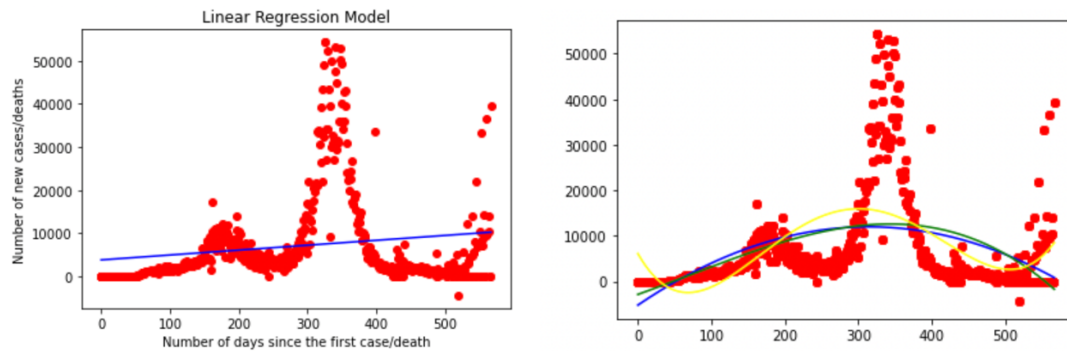
MEMBER TASK

Lahari Chilakuri, Graduate Student

Linear and Non-Linear regression models:

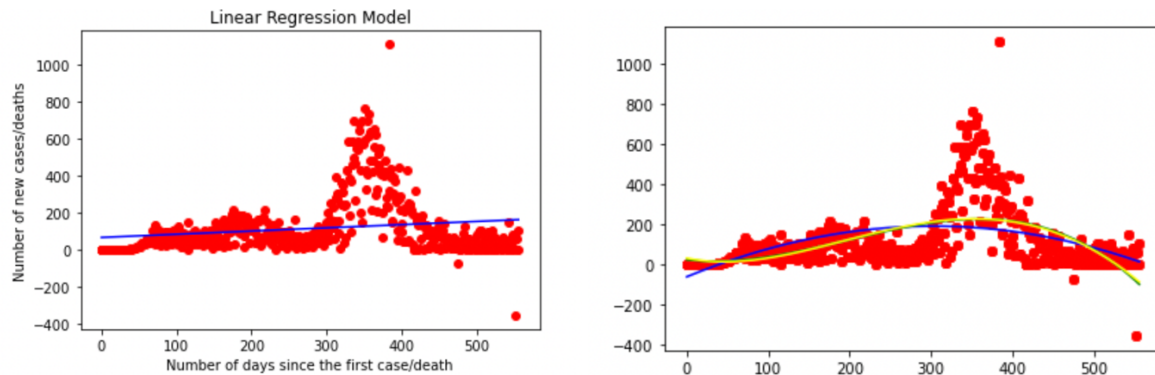
Here I have utilized the linear and non-linear regression models to compare trends for the California state.

The linear and non-linear regression models for the California state cases are as below:



Here, the number of days since the first case is taken as X-axis and number of new cases is taken as Y-axis.

The linear and non-linear regression models for the California state cases are as below:



Here, the number of days since the first case is taken as X-axis and number of new deaths is taken as Y-axis.

The RMSE values for Linear Regression of the cases and deaths of the California state are:

RMSE for Linear Regression(cases): 10476.23830557541

RMSE for Linear Regression(deaths): 148.44545350797327

The RMSE values for Non-Linear Regression of the cases and deaths of the California state are:

RMSE for Polynomial Regression with degree 2(cases): 9575.05052634204

RMSE for Polynomial Regression with degree 3 (cases): 9525.803271903882
 RMSE for Polynomial Regression with degree 4 (cases): 8875.273767113391
 RMSE for Polynomial Regression with degree 2 (deaths): 132.56486847759808
 RMSE for Polynomial Regression with degree 3 (deaths): 125.46921883076848
 RMSE for Polynomial Regression with degree 4 (deaths): 125.40479047246164

Identifying counties which are at risk:

From stage 2, I have listed out the five counties which are at high risk. The five counties are:

1. Alpine County
2. Amador County
3. Lassen County
4. Mono County
5. Yuba County

Just like for the California state cases and deaths, the linear and non-linear regression models are plotted for these five counties and the RMSE values are as follows:

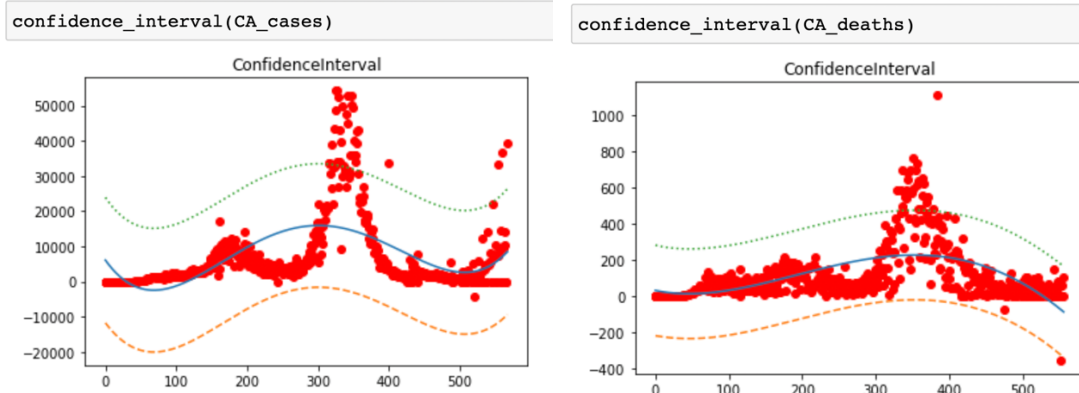
	Linear Regression	Non-Linear Regression (degree 2)	Non-Linear Regression (degree 3)	Non-Linear Regression (degree 4)
Alpine County	3.2624696	3.2593192	3.2587870	3.256191
Amador County	3.262659	3.2601515	3.2590737	3.2577669
Lassen County	39.929284	39.6801802	39.656286	39.1872564
Mono County	0.708065	0.70597434	0.70573497	0.70494954
Yuba County	58.9448301	58.21545	58.125787	58.125787

Performing Hypothesis Test:

I have taken the Hospital Beds dataset as my enrichment dataset. The hypothesis question is “Whether the number of people who got vaccinated affects the number of Covid cases?” By performing Chi-square test for categorical hypothesis testing, its concluded that the number of people who got vaccinated effects the decrease in the number of covid cases. The increase in the number of vaccinated people decreases the confirmed covid cases.

Confidence Interval:

The confidence interval for the number of cases and deaths of the California state are plotted:



Similarly, the confidence interval of both cases and deaths of the five counties which are at risk are also plotted.

The trend lines of both cases and deaths are as follows:

