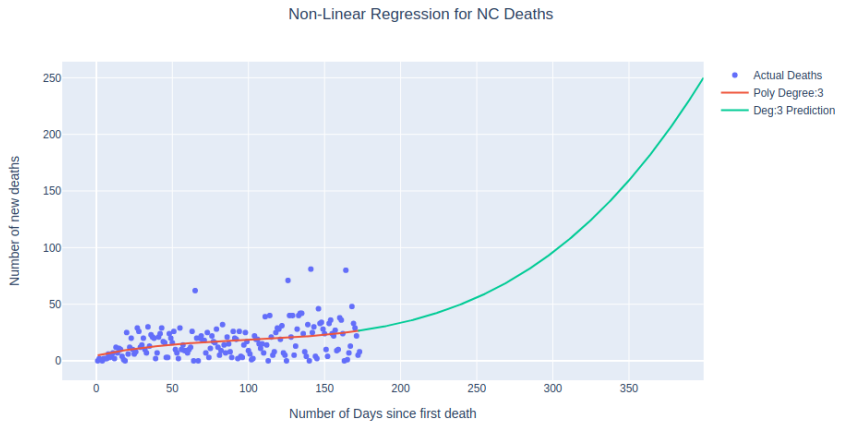
Project Stage -3 Report
Submitted By : Rashmi Hassan Udaya Kumar

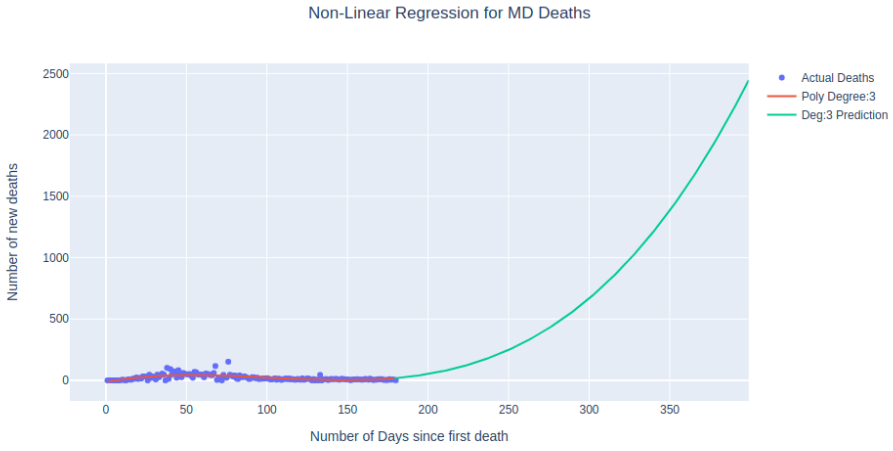
Member task 2 Conclusion

- After analyzing the trend lines and prediction paths of the 5 counties, we can observe that the Montgomery county is at higher risk and the number of cases are predicted to increase at a higher rate when compared to other counties.
- The second county which is at most risk is the Robeson County

Member task 3 conclusion

- When we use non-linear model with degree 3 for modeling the deaths in NC, we observe from the prediction of the number of deaths is increasing slowly, it will take 792 days from the day, the first case was detected the number of deaths are crossing 2648 which is the number of ICU beds in the state of NC. Thus after 792 days NC will go into point of no return.
- when we use non-linear model with degree 3 for modeling the deaths in MD, we observe from the prediction of the number of deaths is increasing rapidly, it will take 346 days from the day, the first case was detected the number of deaths are crossing 1311 which is the number of ICU beds in the state of NC. Thus after 346 days MD will go into point of no return.
- All the other states have a decrees in the predicted number of deaths.
- Thus MD might reach the point of no return faster when compared to other states

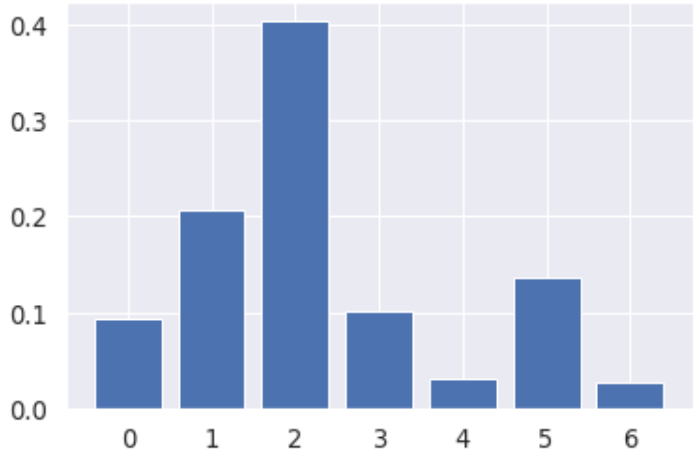




Member Task 4 Conclusion
- Among the 3 models , Random Forest has less RMSE value and better fits the number of cases data.

Conclusion on Relative importance of variables with respect to cases

```
Feature: 0, Score: 0.09406
Feature: 1, Score: 0.20765
Feature: 2, Score: 0.40282
Feature: 3, Score: 0.10117
Feature: 4, Score: 0.03095
Feature: 5, Score: 0.13578
Feature: 6, Score: 0.02757
```




```
Root Mean Squared Error for Random Forest Regressor: 12979.40691052977
--------------------------------------------------------------------------------
```

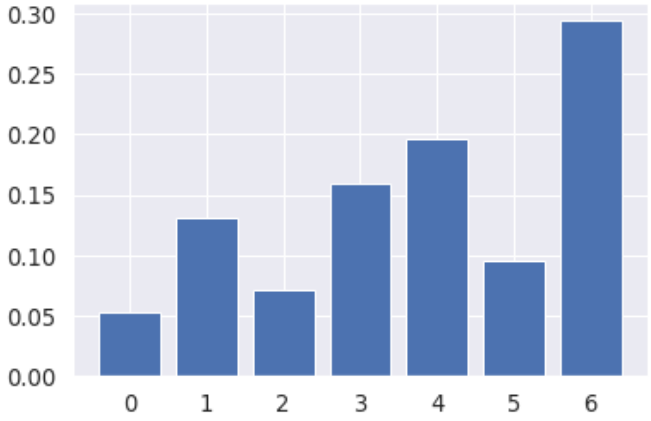
The Chosen variables are :
- 0: 'TotalMales'
- 1: 'TotalFemales'
- 2: 'age<5'
- 3:'ageFrom45to54'
- 4: 'age75to84'
- 5: 'whitePopulation'
- 6: 'blackPopulation'

We observe by the RMSE value of Decision tree and Random Foreset that Random Forest model better fits the new cases data.
From the above random forest plot we observe that
- 'TotalFemales' has more feature importance than 'TotalMales'. it means that, more number of females are affected by covid-19 than males
- Among all the age groups, the age group that has more importance is age<5 which means that kids with the age<5 are more affected by Covid-19

- 'whitePopulation' ha more feature importance score when compared to 'blackPopulation' . This means that white population is more affected by covid-19 than black

Conclusion on relative importance of variables with respect to deaths

```
Feature: 0, Score: 0.05295
Feature: 1, Score: 0.13106
Feature: 2, Score: 0.07126
Feature: 3, Score: 0.15977
Feature: 4, Score: 0.19630
Feature: 5, Score: 0.09504
Feature: 6, Score: 0.29362
```




```
Root Mean Squared Error for Random Forest Regressor: 391.14589483899425
```

The Chosen variables are :
- 0: 'TotalMales'
- 1: 'TotalFemales'
- 2: 'age<5'
- 3:'ageFrom45to54'
- 4: 'age75to84'
- 5: 'whitePopulation'
- 6: 'blackPopulation'

We observe by the RMSE value of Decision tree and Random Forest that Random Forest model better fits the new deaths data.
From The above random forest plot we observe that the
From the above random forest plot we observe that
- 'TotalFemales' has more feature importance than 'TotalMales'. This means that death rate in females is more than male population
- Among all the age groups, the age group that has more importance is 'age75to84'which means that death rate in old people with the age>75 are more.
- 'blackPopulation' has more feature importance score when compared to 'whitePopulation' . This means that black population has more death rate than white population.