

Name: **Muley, Tushar**

Assignment: **DSC 680 - Week 4 Milestone 3 Questions & Answers**

Date: **January 9, 2022**

Questions:

1. Where was the data obtained from?

The data was obtained from 'Our World in Data' the 'Natural Disaster' subsite. The Our World in Data' site did do some transformation and combining of data to get their data. It is all documented in the reference section of the White Paper.

2. Why were two different datasets used to perform the analysis?

The reason for the two different dataset was the information on the Natural Disaster subsite needed to be combined with additional hurricane data like ACE information. Everything was split out and I need a more cohesive data set.

3. Why did the models have such low R-Squares?

I believe this can be attributed to two factors. The data was robust. But more detail information is needed. As I started going into this further I learned having storm strength data, population and common storm tracks might have help determine the number deaths a little better. I tried binning the deaths so we could put a range on the expectation. This help the R^2 value went from -2.41 to -1.07. An improvement but still falling short of anything meaning full. The binning of data would be too broad.

On the economic impact I saw a good starting value of 0.410. That was after using OLS model. I think I can improve this by selecting a better model and identifying common storm tracks. That information was a little hard to find.

4. What other different predictive models could you have run?

I ran linear regression, decision tree and OLS. OLS seems to produce the best results for the limited amount of time I have to work on this. I would try using larger amounts of data and running classification model for deaths and maybe for economic damage as well. If you knew which areas these storms might head in you would be able to determine the amount of damage they might cause. Deaths would be hard because human are becoming desensitized to warnings and requests to leave when big storms like this approach. Having sentiment analysis along with running ensemble model might help.

5. What different data points could you have used to get better correlation?

The correlation is not bad. It could be better. I believe having some additional data like population, common storm tracks and maybe even some type of storm sentiment. Storm sentiment would be how people are perceiving the threat alerts. How they feel about the first storm of the season compared to the second, third all the way to average. This might get better death rates. Or even areas where deaths have occurred.

6. What understanding is needed to perform this analysis?

When I started the analysis I thought it was straight forward. There is a lot to understand about hurricanes, their tracks, what cause changes and other details. This was something I did not expect. The volume of data needed is large and I believe a review of how various data effect the correlation would need to be done. This was not simple and unexpected. With the amount of data that would be need an the different types of analysis that would need to be done before coming to a predictive model is a lot. This would have been better served as an EDA project.

7. Can additional detail data at a lower-level help make the analysis better?

Yes, more detail is always better but too much detail will probably slow the model down and cause other issues. For example, when you watch Metrologies show different storm tracks they are taking into account so many different variables like water temp, pressure and water currents in real time it is difficult to gauge where a storm will go. Detail would help but deaths might be harder to predict. You could come up with a range or even category but it would be wide and saying we can expect 0 to 100 deaths and be off by more than 50%.

8. Would you be able to obtain better data from NOAA?

NOAA has a lot of the data. But they don't just make it available as it is very large. They use different source to gather the data like airplanes, boats, and sensors in the water. It can be done but it will be a large amount of data and my PC would not be able to handle it.

9. Can hurricanes deaths be predicted?

Yes. Even if categorized into say small, medium or large can be done. It will be very inaccurate on the level that it would help the public make the connection that a storm is bad enough to warrant leaving the area. That every storm is like finger print. They are all a little different and different enough that you can't get a large population to understand the deadliness of it.

10. Can anything else be predicted from the data gathered?

Yes, I had some success but limited. I was able to gauge economic damage from the storm. But it is low at 41% accuracy. I believe that can be increased if I knew what additional data I could get and changing the model to classification or running a ensemble of models.