Storm Events of 2010 and 2011

CIS 3200-02
Group 1
Arne Mueller, Erik Pacheco, Michael Do, Shadman Sayef, & Steven Vi

Abstract: With this paper, we are using storm data events from the National Centers for Environmental Information (NOAA) to research trends and insights about storms that occurred in the U.S. We use both Elasticsearch and Kibana cloud servers to analyze and make insights from various storm data event CSV files. The data that we researched were details, locations, and fatalities from the events of 2010 and 2011. The reason we chose this is because we want to analyze how we could use storm events data to predict future storm events as well as other reasons. Overall, we found a lot of insights from the CSV files.

Introduction

In the years 2010 and 2011, there was a large recording of storm events that took place in the United States, which was conducted by the NOAA. Seeing the amount of information on the storm events in the years 2010 to 2011, we decided we wanted to analyze these datasets. It must be noted that due to page constraints, we use a mixture of 2010 and 2011 data for this paper. In our analysis, we will be exploring a predictive analysis that focuses on the magnitude variable, a geo-map of the US region with the number of storm events that have taken place, and look into the deaths and injuries that these storms have done.

Specifications

- dataset size: 153.85300 mb.
- dataset URL: https://www.ncei.noaa.gov/pub/data/swdi/stormevents/csvfiles/
- GitHub URL: https://github.com/MichaelCIS/3200_Stormevent

Body

Storm Locations 2010

On the storm locations map for 2010 we can see that the majority of storm events happened in the states that are located in the middle east and east coast of the US. Only a very small amount of storm events took place on the west coast. The highest number of storm events is located in the area of and around Iowa, with the highest value being 1955. In second place comes the area of and around North Carolina with a value of 1573 storm events. The lowest number of storm events is located in the area of northern California, with a minimal number of one storm event.

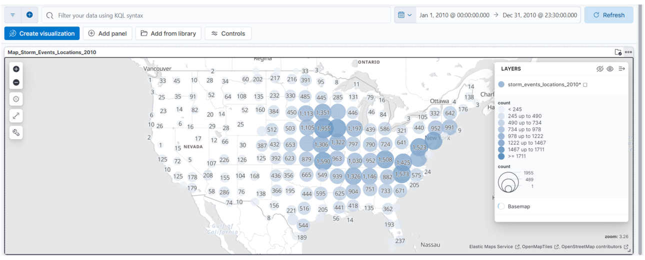

Figure 1: Cluster map of storm event locations in the US for the whole year of 2010, including storm event numbers.

Storm Locations 2011

Next we look at the storm locations map for 2011. Here we can recognize a similar pattern of storm events that took place in the US. In this year, the majority of storm events slightly shifted to the lower Middle East, and Upper East Coast states. Now, the highest number of storms that occurred is located in the area of Pennsylvania and New Jersey with a value of 2849. The second highest number is visible in the area of Missouri and Arkansas, with 2746 storm events. The lowest number of storms occurred in the area of northern California again. The value stays the same with a total of one storm event in that specific area.

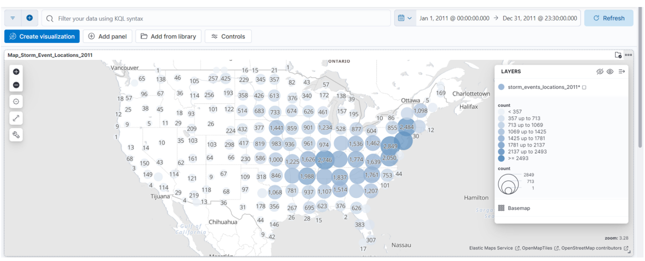

Figure 2: Cluster map of storm event locations in the US for the whole year of 2011, including storm event numbers.

Storm Events Fatalities 2010

In this section, we will have a look at the fatalities that were caused by storm events in the whole year of 2010. With the total number of deaths being 674, it is visible that the most deaths for this year occurred around the summertime. The highest peak can be seen around the end of April and the beginning of May, with a value of almost 50 deaths. The rest of the higher numbers of deaths took place from June to the middle of August, with up to almost 40 deaths in the middle of July for the year 2010.

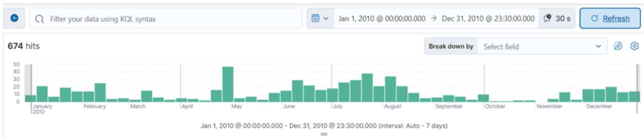

Figure 3: Bar chart visualizing storm event fatalities of the whole year of 2010 in the US. Time period listed as months.

When it comes to the locations of how and where these deaths occurred, we can see that vehicles and towed trailers are ranked first, with a value of 207 deaths in total. Open areas and permanent homes are ranked as second and third place with values consisting of 133 and 107 deaths.

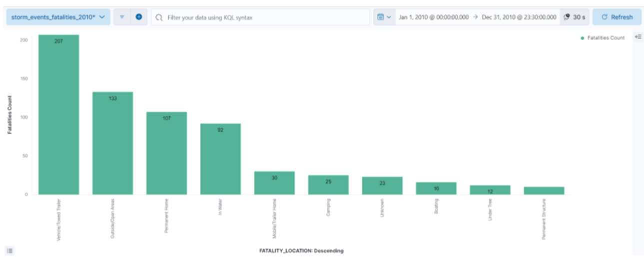

Figure 4: Bar chart visualizing locations and causes of storm event fatalities for the whole year of 2010 in the US.

Storm Events Fatalities 2011

When we now look at the storm event fatalities of 2011 in comparison, it is recognizable that the total number of deaths has a value of 1335. This is almost twice the amount of deaths than in the year of 2010. This time the most deaths occurred specifically around the time of late spring and early summer. The highest number of fatalities can be seen at the end of April reaching almost 350 deaths. The second highest value is located in the middle of May with over 150 deaths.

Figure 5: Bar chart visualizing storm event fatalities of the whole year of 2011 in the US. Time period is listed as months.

In terms of locations of how and where these deaths for 2011 occurred, we now can see that permanent homes got ranked in the first place with a value of 404 deaths. Vehicles and towed trailers now dropped to the second rank with a total value of 283 deaths. Also, open areas dropped one rank, now being in third place of fatality locations with a number of 149.

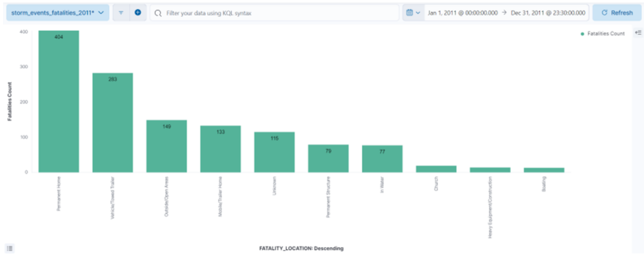

Figure 6: Bar chart visualizing locations and causes of storm event fatalities for the whole year of 2011 in the US.

Top 10 Causes of Storm Events [2010]

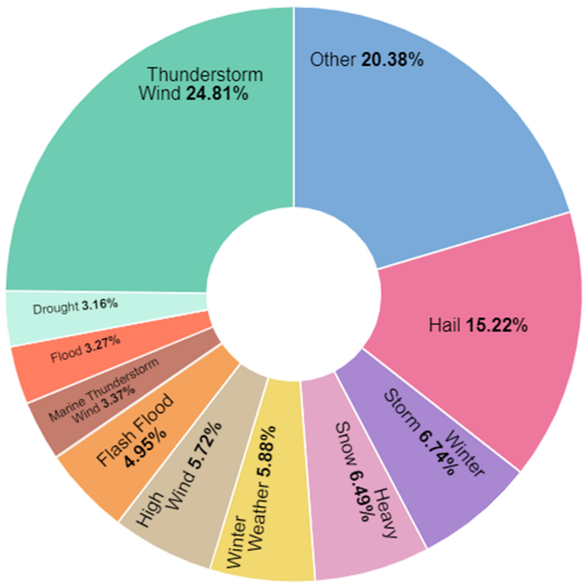

Figure 7: The pie chart presents the top 10 storm events in the year 2010.

This visualization describes the top 10 causes of storm events recorded in the dataset for the year 2010. In order, from highest to least, thunderstorms accounted for the most at a rather substantial ~25%, followed by hail, winter storms, heavy snowfall, winter weather, high wind, flash floods, marine thunderstorm wind, floods, and lastly, droughts. The proportion of thunderstorm winds to every other storm type is quite high, showing that the real focus should be on that factor.

Top 10 States with Recorded Weather Events [2010]

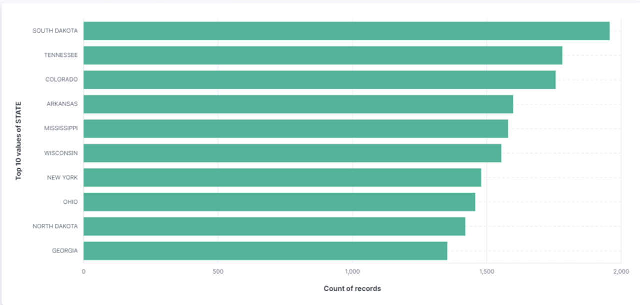

Figure 8: This vertical bar chart shows the 10 states with the most weather events that have taken place.

This bar chart visualization presents the top 10 states with weather events described in the dataset. Northern US State South Dakota tops the list, followed closely by Tennessee, Colorado, Arkansas, Mississippi, Wisconsin, New York, Ohio, North Dakota, and Georgia. Surprisingly, states all over the country are included in this top 10 list, highlighting the importance of storm preparedness and well-engineered infrastructure that can withstand the force of these various weather events.

Number of 2011 Storm Events in 20 States

In this analysis, we explored the number of storm events that took place in the 20 states in 2011. From the chart, we found that Texas was the number one state to have the most occurrences of storm events. The number of storms that have taken place in Texas is 28 storm events, which means in 2011, Texas had the most storms to occur in the state. If we wanted to look at the second state to have the most storms, it would be California, with 26 storm events to occur in their state.

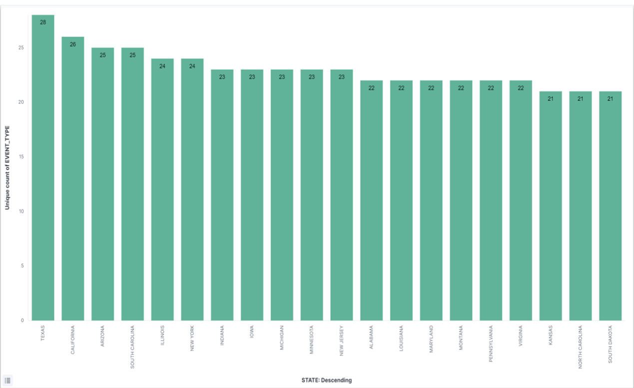

Figure 9: This bar chart presents the 20 states that had the most storm events to occur in their states. Unique count of Event_type and Term of States were used to create this bar graph.

Percentage of Direct Injury from Different Types of Storms in 2011

For this pie chart, we explored the percentage of direct injury caused by storms in 2011. Looking at our pie chart, we discovered that the Tornado had done the most direct injury in the year 2011 with a percentage of 33.33%, and as for the second storm to cause the most direct injury, it would be excessive heat with a percentage of 17.89%. Overall, the tornado dealt more direct injuries compared to the other storms.

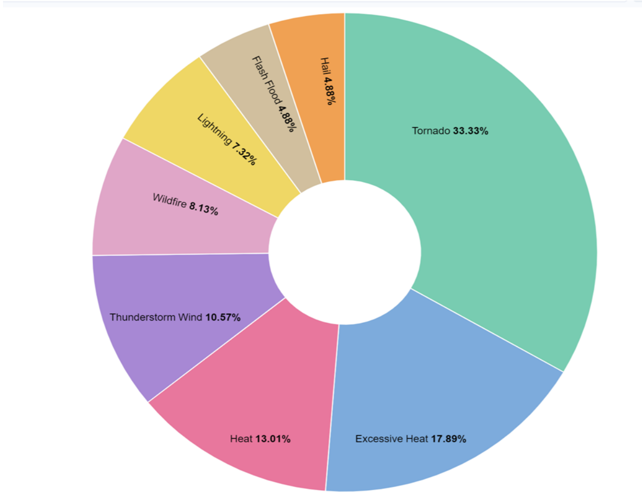

Figure 10: This pie chart presents the percentage of direct injury caused by 8 different storm types in 2011.

Regression Data with Feature Importance

Figure 11: This figure represents the feature importance from 2010

For the making of this chart, I used machine learning and used magnitude as our dependent variable to analyze the impact of storm events. In this context, magnitude means the size of the storms. It is important for us to analyze this because we want to understand how magnitude could affect the impact of the storm event and could use it to predict future storm events. From here, this shows the labels that display the most influential feature importance. The most relevant feature importance is EVENT_TYPE, with the magnitude of 24.4.

Machine Learning Feature Importance

ml.is_training	ml.MAGNITUDE_predi...	MAGNITUDE	ml.feature_importance	BEGIN_AZIMUTH	BEGIN_DATE_TIME	BEGIN_DAY
true	51.733	50	[{"feature_name":["BEGIN_L...	SE	28-FEB-11 18:00:00	28
true	51.452	50	[{"feature_name":["BEGIN_L...	NE	28-FEB-11 18:14:00	28
true	51.508	55	[{"feature_name":["BEGIN_L...	S	28-FEB-11 18:02:00	28
true	50.248	50	[{"feature_name":["DAMAGE...	NW	24-FEB-11 23:20:00	24
true	53.783	52	[{"feature_name":["DAMAGE...	E	24-FEB-11 00:00:00	24
true	53.917	56	[{"feature_name":["CZ_NAM...	WNW	24-FEB-11 00:00:00	24
true	42.835	41	[{"feature_name":["EPISODE...		28-DEC-11 10:00:00	28
true	50.995	50	[{"feature_name":["BEGIN_L...	N	22-DEC-11 04:30:00	22
true	53.498	52	[{"feature_name":["END_LO...	NNW	22-MAY-11 18:20:00	22
true	50.876	48	[{"feature_name":["BEGIN_L...	W	25-FEB-11 01:50:00	25
true	50.901	50	[{"feature_name":["BEGIN_L...	NW	07-DEC-11 17:51:00	7
true	55.454	56	[{"feature_name":["BEGIN_L...	W	07-DEC-11 17:51:00	7
true	59.74	60	[{"feature_name":["DAMAGE...	E	07-DEC-11 17:23:00	7
true	1.146	1.75	[{"feature_name":["END_LO...	SSE	30-JAN-11 17:25:00	30
true	33.627	37	[{"feature_name":["EVENT_T...		01-FEB-11 12:00:00	1
true	30.875	30	[{"feature_name":["EVENT_T...		01-FEB-11 11:30:00	1
true	62.387	61	[{"feature_name":["DAMAGE...	SSE	25-JAN-11 17:20:00	25
true	33.2	33	[{"feature_name":["EVENT_T...		03-APR-11 07:00:00	3
true	33.663	35	[{"feature_name":["DAMAGE...		03-APR-11 07:00:00	3
true	36.35	42	[{"feature_name":["EVENT_T...		03-APR-11 07:00:00	3
true	51.646	50	[{"feature_name":["BEGIN_L...	S	28-FEB-11 16:08:00	28

Figure 12: This figure presents the outcome of the machine learning feature importance for storm events of 2010

For this graphic, this delves into a closer look at feature importances and how much of an impact various labels have towards the dataset.

Influencers - 2010

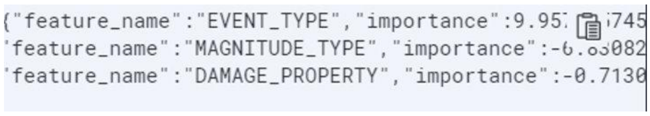

Figure 13: This figure shows the influencers using the json view for 2010 storm events

Continuing on from the previous graphic, we utilize the json aspect to see how much of an impact each label has towards our dataset. EVENT_TYPE has the highest impact with 9.95 magnitude. This is based on the 2010 storm details dataset, but it has similarities with the 2011 storm details dataset with varying magnitudes

Training Data - 2010

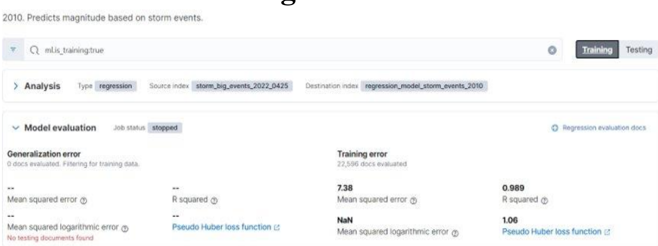

Figure 14: This figure shows the training data of the 2010 storm event dataset

Using the 2010 dataset, we analyze the training data. MSE is 7.38 and R-squared is 0.989. This means that this data has good accuracy as to predicting future storm events based on the magnitude.

Testing Data - 2010

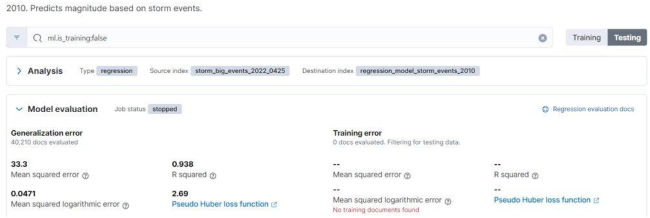

Figure 15: This figure presents the testing data of the 2010 storm event dataset

From here, using the 2010 dataset, we analyze the testing data. MSE is 33.3, and R-squared is 0.938. Just like training data, this has good accuracy in predicting future storm events based on magnitude.

Influencers - 2011

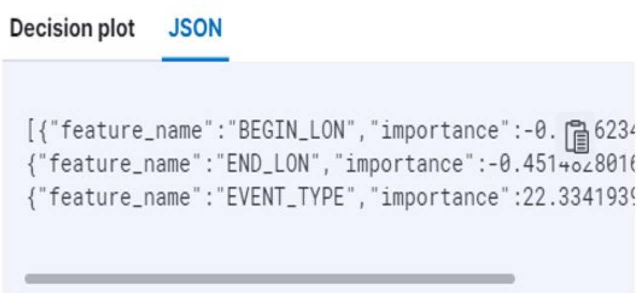

Figure 16: This figure shows the json view of influencers of the 2011 storm event dataset

The key feature is EVENT_TYPE, which has a maximum value of 22.33. This implies that events classified as EVENT_TYPE have a significant impact or influence and may even play a critical role in the dataset.

Training Data - 2011

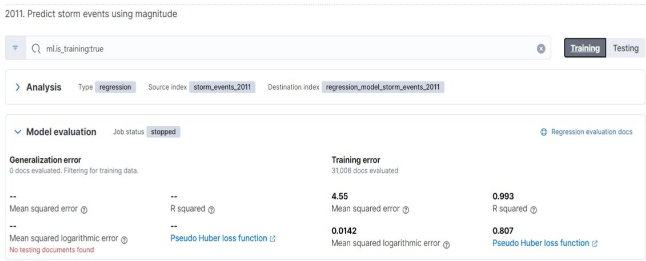

Figure 17: This figure shows our training data result of the 2011 storm event dataset.

The model shows a good fit to the training data, as indicated by a low Mean Squared Error (MSE) of 4.55..The high R-squared value of 0.993 indicates that the model explains about 99.3% of the variance in the training data, reflecting a strong ability to capture patterns.

Testing Data - 2011

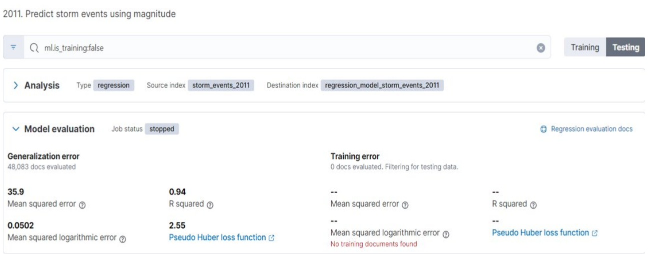

Figure 18: This figure presents our testing data result of the 2011 storm events dataset.

When applied to new data, the model encounters difficulties, with a higher MSE (35.9) and a lower R-squared (0.94) compared to the training data. This suggests that generalizing to previously unseen data may be difficult, possibly due to overfitting.

Understanding Model Performance:

Understanding the model's behavior on both training and testing data is critical for evaluating its generalization capabilities. While a model may excel at fitting training data, the ultimate goal is to ensure that it performs well on new, previously unseen data.

Conclusion

In conclusion, we have learned a lot when using Elasticsearch and Kibana to analyze storm event files. In addition, we have gained a lot of insights because there are quite a few significant similarities and differences. In terms of similarities, we have learned that the majority of storm events happened around the mid-east and eastern states of the U.S. for both 2010 and 2011. We also learned that the majority of deaths occurred in vehicles and towed trailers. In addition, the majority of the storm events happened around late spring to summer. In terms of differences, most storm events happened to be thunderstorms in 2010, and that South Dakota and Tennessee had the most storm events during that year. However, in 2011, Texas had the most storm events. In addition, during that same year, most storm injuries were tornadoes, and most of them occurred at home. Finally, for machine learning, we learned that the magnitude of the storms depends on the event types, which makes sense. Depending on the event type of the storm, the higher or lower of the magnitude of the storm. From all of these insights, we learned a lot from using Elasticsearch and Kibana to help us make these conclusions.

Works Cited

***Index of /pub/data/swdi/stormevents/csvfiles*. (n.d.).**
https://www.ncei.noaa.gov/pub/data/swdi/stormevents/csvfiles/

https://www.ncei.noaa.gov/pub/data/swdi/stormevents/csvfiles/StormEvents_details-ftp_v1.0_d2010_c20220425.csv.gz

https://www.ncei.noaa.gov/pub/data/swdi/stormevents/csvfiles/StormEvents_details-ftp_v1.0_d2011_c20230417.csv.gz

https://www.ncei.noaa.gov/pub/data/swdi/stormevents/csvfiles/StormEvents_locations-ftp_v1.0_d2010_c2022042

https://www.ncei.noaa.gov/pub/data/swdi/stormevents/csvfiles/StormEvents_locations-ftp_v1.0_d2011_c20230417.csv.gz