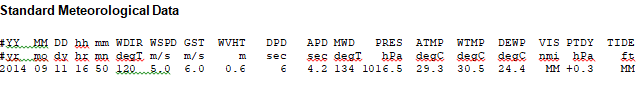
|  |
| --- |
| **Buoys: Riders on the Storm** |
| Buoy Data and their Impact on Hurricane and Tropical Storm Predicting in the Gulf of Mexico -- Ruthy Amkraut-Megan Moroney-Rick Gentile |
| This document details the fifteen years of buoy data utilized in a Machine Learning predictive model for when Hurricane and Tropical Storms will occur in the Gulf of Mexico. |
|  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Data Data from the National Data Buoy Center was compiled from the time period of January 2005 through December 2019 from buoys in the Gulf of Mexico. The data was further delineated to have datasets consisting of data when storms occurred and when storms did not occur. The delineation was based on a listing of Hurricanes and Tropical Storms for the time period detailed above. References: <https://www.ndbc.noaa.gov>; https://www.aoml.noaa.gov/hrd/hurdat/All\_U.S.\_Hurricanes.html   |  |  | | --- | --- | | Abbreviation | Description | | WDIR | Wind direction (the direction the wind is coming from in degrees clockwise from true N) during the same period used for WSPD. See Wind Averaging Methods | | WSPD | Wind speed (m/s) averaged over an eight-minute period for buoys and a two-minute period for land stations. Reported Hourly. See Wind Averaging Methods. | | GST | Peak 5 or 8 second gust speed (m/s) measured during the eight-minute or two-minute period. The 5 or 8 second period can be determined by payload, See the Sensor Reporting, Sampling, and Accuracy section. | | WVHT | Significant wave height (meters) is calculated as the average of the highest one-third of all of the wave heights during the 20-minute sampling period. See the Wave Measurements section. | | DPD | Dominant wave period (seconds) is the period with the maximum wave energy. See the Wave Measurements section. | | APD | Average wave period (seconds) of all waves during the 20-minute period. See the Wave Measurements section. | | MWD | The direction from which the waves at the dominant period (DPD) are coming. The units are degrees from true North, increasing clockwise, with North as 0 (zero) degrees and East as 90 degrees. See the Wave Measurements section. | | PRES | Sea level pressure (hPa). For C-MAN sites and Great Lakes buoys, the recorded pressure is reduced to sea level using the method described in *NWS Technical Procedures Bulletin 291* (11/14/80). ( labeled BAR in Historical files) | | ATMP | Air temperature (Celsius). For sensor heights on buoys, see Hull Descriptions. For sensor heights at C-MAN stations, see C-MAN Sensor Locations | | WTMP | Sea surface temperature (Celsius). For buoys the depth is referenced to the hull's waterline. For fixed platforms it varies with tide, but is referenced to, or near Mean Lower Low Water (MLLW). | | DEWP | Dewpoint temperature taken at the same height as the air temperature measurement. | | VIS | Station visibility (nautical miles). Note that buoy stations are limited to reports from 0 to 1.6 nmi. | | PTDY | Pressure Tendency is the direction (plus or minus) and the amount of pressure change (hPa)for a three hour period ending at the time of observation. (not in Historical files) | | TIDE | The water level in feet above or below Mean Lower Low Water (MLLW). |  Heading 2 You can easily change the formatting of selected text in the document text by choosing a look for the selected text from the Quick Styles gallery on the Home tab. You can also format text directly by using the other controls on the Home tab. Most controls offer a choice of using the look from the current theme or using a format that you specify directly.  To change the overall look of your document, choose new Theme elements on the Page Layout tab. To change the looks available in the Quick Style gallery, use the Change Current Quick Style Set command. Both the Themes gallery and the Quick Styles gallery provide reset commands so that you can always restore the look of your document to the original contained in your current template. Heading 3 On the Insert tab, the galleries include items that are designed to coordinate with the overall look of your document. You can use these galleries to insert tables, headers, footers, lists, cover pages, and other document building blocks. When you create pictures, charts, or diagrams, they also coordinate with your current document look. |  | • • • |

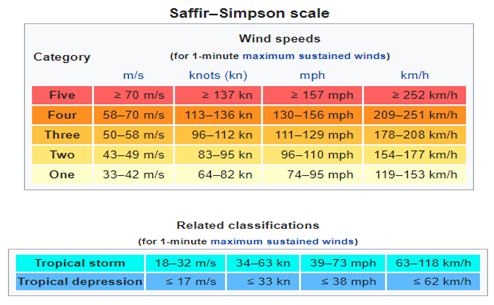
## Data Analysis



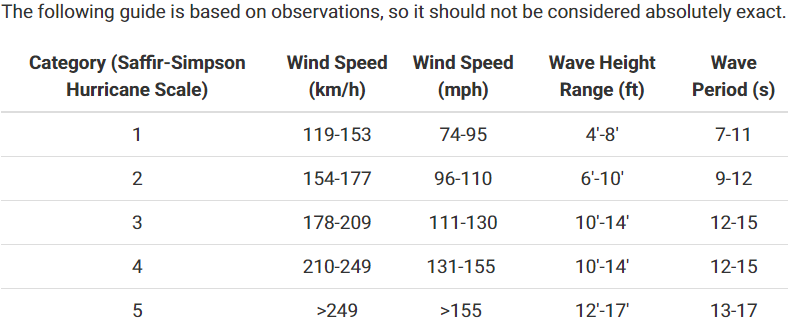
In reviewing the buoy data and important factors of a hurricane, wind speed (WSPD(T m/s), water temperature (WTMP degC)\* and WVHT(m)\*\*\* have been deemed the three (3) data points that will be utilized in the Machine Learning model for storm predicting.

‘\* - Water temperature must be 26 degrees Celsius (79 degrees Fahrenheit). Below this threshold temperature, hurricanes will not form or will weaken rapidly once they move over water below this threshold. (Source: <https://www.weather.gov/.../hurricane_anatomy.html>)

‘\*\* - Windspeed(WSPD m/s) When Converted these are the wind speeds for the Hurricane Categories and Tropical Storm Classifications. (Source: <https://en.wikipedia.org/wiki/Saffir%E2%80%93Simpson_scale>)



‘\*\*\* - Wave Height(WVHT m) When converted to miles per hour below and how they relate to the hurricane categories (Source: <https://www.surfertoday.com/surfing/what-are-hurricane-generated-waves>)



## Heading 2

You can easily change the formatting of selected text in the document text by choosing a look for the selected text from the Quick Styles gallery on the Home tab. You can also format text directly by using the other controls on the Home tab. Most controls offer a choice of using the look from the current theme or using a format that you specify directly.

To change the overall look of your document, choose new Theme elements on the Page Layout tab. To change the looks available in the Quick Style gallery, use the Change Current Quick Style Set command. Both the Themes gallery and the Quick Styles gallery provide reset commands so that you can always restore the look of your document to the original contained in your current template.

## Heading 3

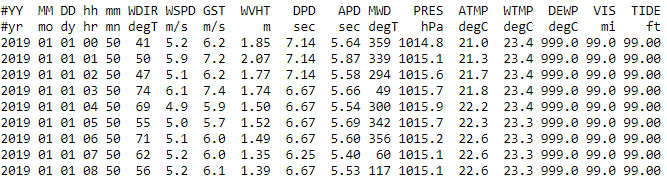
On the Insert tab, the galleries include items that are designed to coordinate with the overall look of your document. You can use these galleries to insert tables, headers, footers, lists, cover pages, and other document building blocks. When you create pictures, charts, or diagrams, they also coordinate with your current document look.

We began with a deep dive for the data and information to build our database and utilized the following sources:



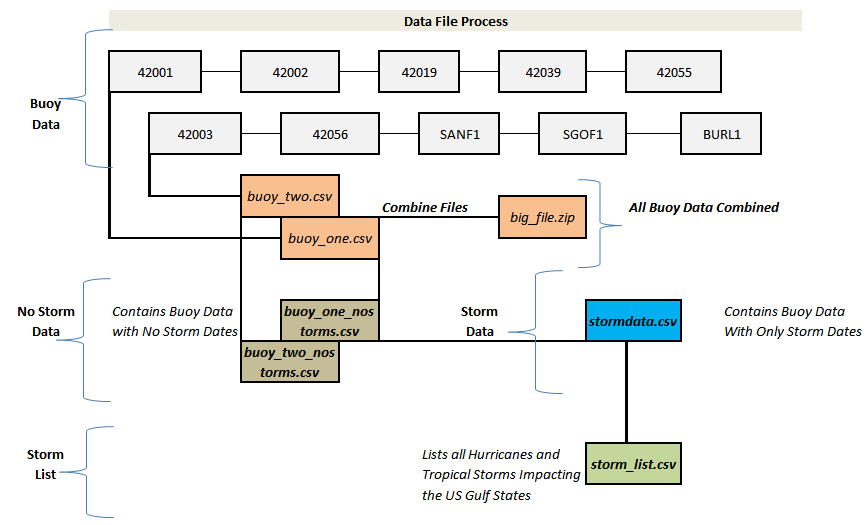
## Data Collection from Buoys

**Sample Data – Standard Meteorological Data**



**Buoy Locations**

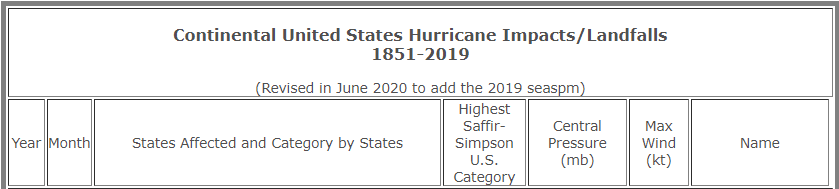
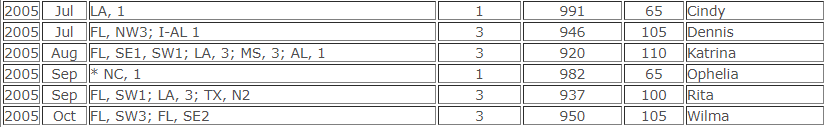


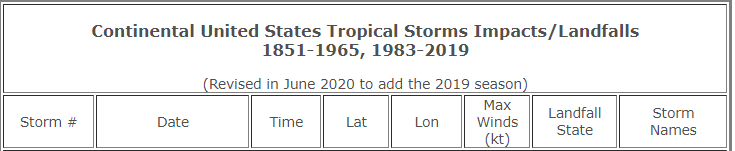
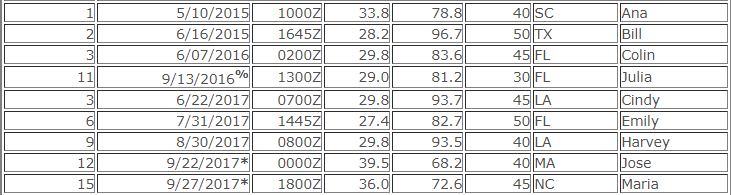


* Buoy data is collected via satellite and sent to the National Data Buoy Center which collects, analyzes and sends out the data. We had to pull data for each buoy separately, which then made it necessary to do a lot of data cleanup and joining in order to get all the data in the same format and same document
* Once this was put together, we were able to determine dates for the data from each buoy when there was a storm versus no storm using the hurricane and tropical storm data explained below.

**Hurricane and Tropical Storm Data**

* The data we found for hurricanes and tropical storms was pulled from the Hurricane Research Division - Atlantic Oceanographic & Meteorological Laboratory. We were able to pull hurricane and tropical storm data from the Atlantic for the last 15 years from this site.
* This data provided storm category, some conditions, affected states, date of the storm, and the name of the storm. We had to combine this data with our buoy data in order to match up the buoy data with each day there was a storm and also to know which days did not have storms.

**Limitations**

* In 2005, two buoys (Buoys 42003 (8/28 5am) and BURL1 (8/29 6am)) were knocked out of commission during hurricane Katrina. Because of this, there is no data for those two buoys from late August to mid-October in 2005.
* There were hurricanes dating back as early as 1851 we could have utilized, but the buoy data did not go back that far.
* We used a static number for each hurricane category, which means we did not encompass the category changes of the hurricane throughout its lifecycle.
* Not all buoys we used may have been close to or in the path of each hurricane. The data from these buoys would most likely be less useful and cause our model to learn inaccurately how to predict storms.
* We understand there are other data sources out there that help predict hurricanes and tropical storms, however we wanted to focus solely on the buoys data for this project.



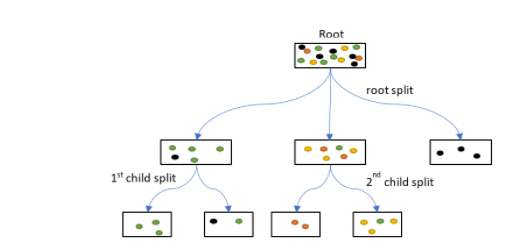
**If We Had More Time……**



* We could have used more buoy data not only from more buoys in the Gulf of Mexico, but a larger area in the Atlantic.
* We would have loved to look into Neural Networks a bit more, how it could have been used with our data, or what we would have needed in order to make it happen.
* We could have analyzed the paths of the hurricanes and storms in order to best select buoys to choose data from
* We could have looked into removing the data that potentially interfered with our machine learning process (as mentioned in Limitations section).

## Machine Learning Predictive Model

**Decision Trees and Random Forests**

* For our model we used classification trees, both decision trees and random forests. Classification trees are essentially a series of questions designed to assign a classification.
* In our case, our classification trees are a series of data points, including our most important variables; water temperature, wind speed and wave height. With this data, our model is trying to answer the question, "Are these the conditions for a storm? If yes, what category storm?"

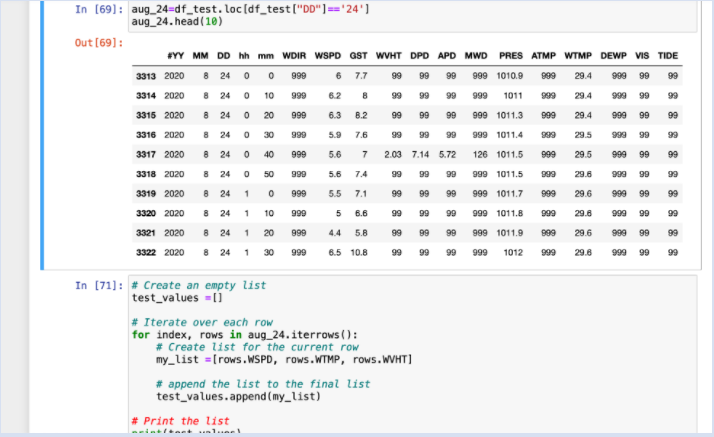
**Methodology**

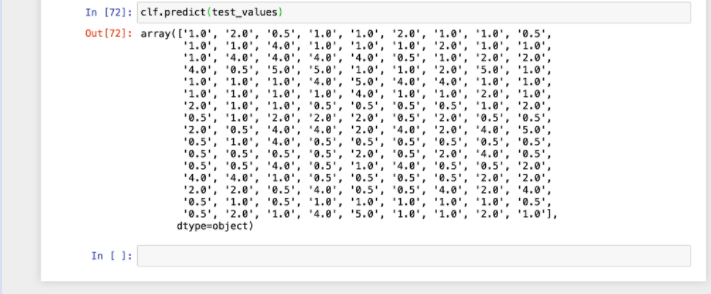
* We **first** looked at a dataset that only had the buoy data from days there were storms, along with the data from that particular storm or hurricane.
  + First looking at a decision tree, we were able to see that the model was only 48% accurate at predicting storms.
  + Next, we ran the random forest and found that was only able to predict with 52% accuracy.
  + We then tested the model to find out if it really could predict storms during a previous month where we already knew all the storms that occurred.
  + **We found that this model was the most accurate at predicting storms when doing this.**
* **Next**, we looked at our full dataset that included the buoy data from days with storms, but also all of the buoy data from days without storms.
  + Looking at the **decision tree** in this model, we were able to see a much better **predictive score from our model, 91%**.
  + The **random forest** had an even better ability to **predict storms at 93%**. However, when we tested this model with a previous month of data, we saw that it was actually less accurate at predicting storms, despite the higher predictive score.

Below is some of our code that we used to build the decision tree and random forest models.

**Code using just the Storm Data**

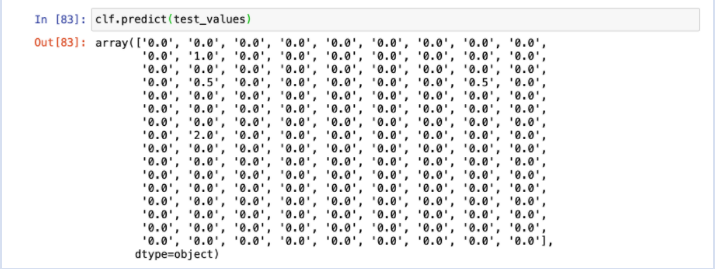






**Code using all data - days with storms and without**





**Limitations**



* Our complete dataset has the better predictive score, however, we found it is actually not predicting storms on days with conditions where we know there are storms.
* We found just the storm data (the smaller dataset) had a lower predictive score, but predicted storms more accurately.
* We had a lot of data from the buoys, but there have not been enough storms in the last 15 years to account for all of the buoy data we had.

**If We Had More Time……**



* We would dig further into why the model worked less accurately with more data.
* We would potentially remove some of the non-storm data points to lessen the difference between non-storm days and storm days in a year.

## Tools Used



* Python/Pandas - used to combine data frames and filter data
* HTML/CSS Bootstrap - used to create and format our site
* SQL Database - used to create our database with the storm and buoy data
* Tableau - used to create buoy visualizations and manipulate our data
* Scikit-Learn - python machine learning library used in order to do our machine learning and random forest
* Random Forest Classifier - within scikit-learn, used to create our predictive model.
* Github - used to host our repository as well as host our site

## References (all are Links)

* [National Data Buoy Center](https://www.ndbc.noaa.gov/)
* [Hurricane Research Division - Atlantic Oceanographic & Meteorological Laboratory (Hurricane Data)](https://www.aoml.noaa.gov/hrd/hurdat/All_U.S._Hurricanes.html)
* [Hurricane Research Division - Atlantic Oceanographic & Meteorological Laboratory (Tropical Storm Data)](https://www.aoml.noaa.gov/hrd/hurdat/uststorms.html)
* [National Weather Service](https://www.weather.gov/source/zhu/ZHU_Training_Page/tropical_stuff/hurricane_anatomy/hurricane_anatomy.html)
* [Wikipedia- Saffir – Simpson scale](https://en.wikipedia.org/wiki/Saffir%E2%80%93Simpson_scale)
* [Surfer Today - "What are Hurricane Generated Waves?"](https://www.surfertoday.com/surfing/what-are-hurricane-generated-waves)
* [Understanding Decision Trees for Classification (Python)](https://towardsdatascience.com/understanding-decision-trees-for-classification-python-9663d683c952)
* [Our Github Repository](https://github.com/ramkraut/final_project)