Atlantic Hurricane Season Forecasting Using NDBC Data

Capstone Project, *Professional Certificate in Machine Learning and Artificial Intelligence*, UC Berkeley Executive Education

Barrett Travis | July 2024

Glossary

NOAA: National Oceanic and Atmospheric Administration

NDBC: National Data Buoy Center; a part of NOAA National Weather Service (NWS)

ACE: Accumulated Cyclone Energy; a metric that accounts for both the strength and duration of storms. Storms that are stronger for longer contribute more to a season’s total ACE.

Oceanic Nino Index (ONI): A measure of the El Nino-Southern Oscillation (ENSO); generally understood to be negatively correlated with Atlantic hurricane activity

Named Storms: Storms are typically named when they reach sustained wind speeds of 38mph

Contents

Introduction and Modeling Objective

Introduction: What is Hurricane Forecasting?

The Atlantic hurricane season is carefully watched by meteorologists and storm enthusiasts, by the property insurance industry, and by communities at risk. Tropical Cyclones have inflicted more than $1.4 trillion in damage to the United States since 1980 (CPI adjusted), making it easily the most costly disaster in the U.S. over that time period.

Every year at the beginning of the hurricane season (which officially kicks off on June 1st), universities, private institutions and government agencies like the National Oceanic and Atmospheric Administration (NOAA) publish “hurricane season outlooks” that look at meteorological and ocean conditions and attempt to predict the number of hurricanes. Many of these forecasts also forecast how many major hurricanes (based on storm strength) or landfalling hurricanes (hurricanes that hit the mainland United States) will occur. For example, NOAA on 23 May 2024 published the article "NOAA predicts above-normal 2024 Atlantic hurricane season" detailing their official 2024 season forecast, forecasting "a range of 17 to 25 total named storms (winds of 39 mph or higher). Of those, 8 to 13 are forecast to become hurricanes (winds of 74 mph or higher), including 4 to 7 major hurricanes (category 3, 4 or 5; with winds of 111 mph or higher). Forecasters have a 70% confidence in these ranges."

Researchers at Colorado State University also publish a well regarded annual hurricane season forecast several times throughout the year; in a July 9th report this year they predicted 25 named storms, 12 hurricanes, and 6 major hurricanes, all well above the 1991-2020 historical averages of 14.4 named storms, 7.2 hurricanes, and 3.2 major hurricanes. This year there was broad consensus that the conditions are ripe for an overly active hurricane season, and unfortunately that seems to have begun to materialize with Major Hurricane Beryl having become the earliest Category 5 hurricane on record in the Atlantic. Most forecasts mention higher than average (in fact, record-breaking) sea surface temperatures and La Nina conditions as factors expected to drive the busier-than-usual hurricane season.

Sources:

- - NOAA National Centers for Environmental Information (NCEI) U.S. Billion-Dollar Weather and Climate Disasters (2024). https://www.ncei.noaa.gov/access/billions/, DOI: 10.25921/stkw-7w73
- - https://www.noaa.gov/news-release/noaa-predicts-above-normal-2024-atlantic-hurricane-season
- - https://tropical.colostate.edu/forecasting.html

Modeling Objective and Scope

Objective: Explore techniques in predictive modeling through the training of regression models to predict hurricane activity (measured by number of storms that occur or by Accumulated Cyclone Energy, or ACE).

Scope: The data used in the training of these models is deliberately limited. From two buoys in the NDBC network we have ocean temperature and some limited local atmospheric data, and from NOAA, the monthly ONI index, to create the features used in the training of the models. The particular NDBC data used spans 1976 to 2022. As we will see, this creates very few data samples if we are creating annual forecasts, and makes modeling perilous. Nonetheless, the goal is to explore techniques and see where modeling with this limited data is useful, and where it falls short.

Data Collection and Feature Engineering

Data Collection: What data matters AND is readily available?

Many factors influence tropical cyclone development. Due to the limited scope of this model, only a couple sources of data were included:

- Sea surface temperatures and other oceanic conditions, from two NDBC buoys (Mid-Gulf and South Hatteras). Meant to be proxies for basin-wide temperatures, but an indexed basin-wide temperature value or, better yet, Ocean Heat Content (OHC) data would be excellent additions.
- Oceanic Nino Index
  - From ONI, calculated the slope of change of ONI (e.g. in May of the given year is ONI increasing or decreasing? ONI plus ONI slope helps better predict ONI deep into hurricane season)

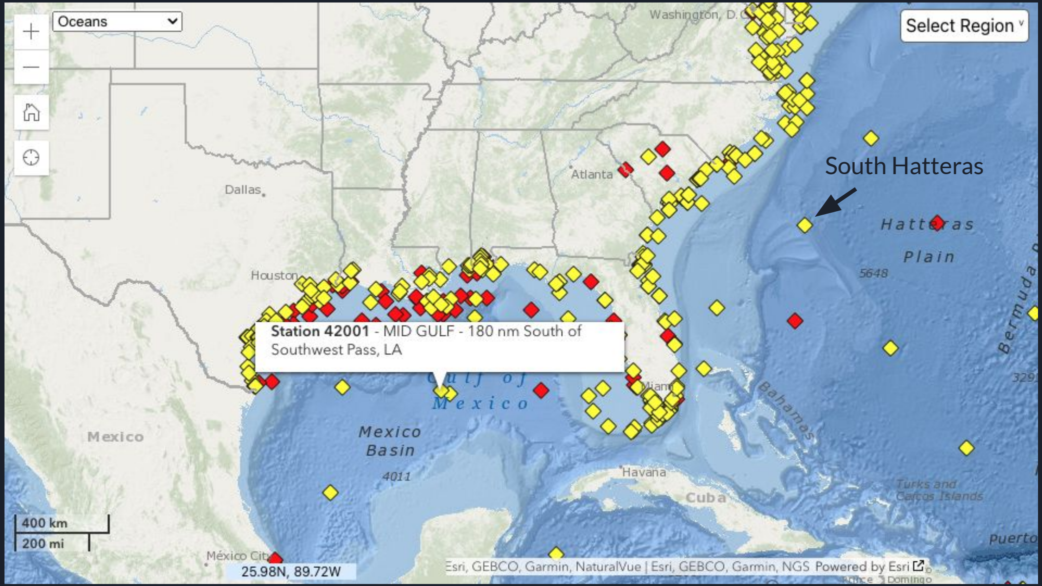
Other factors important to tropical cyclone development that were not used for this model include:

- Oceanic Heat Content or other, better proxies for sea surface temperatures (SSTs)
- Aerosols
- Wind shear

One of the challenges of including data like aerosols (e.g. Saharan dust) or wind shear directly, rather than through longer term meteorological conditions like El Nino, is that direct data on aerosols in May or June may not be predictive of conditions later in the hurricane season. Are there good data sources for these factors?

NDBC Data, 1976 - 2022

NDBC Data was collected from two buoys: MidGulf (Station 42001) and South Hatteras (Station 41002). These stations were selected due to relatively few gaps in data going back to 1976.



Features retained from these buoys:

- WTMP (water temperature)
- WSPD (wind speed)
- WVHT (wave height)
- PRES (air pressure)
- ATMP (air temperature)

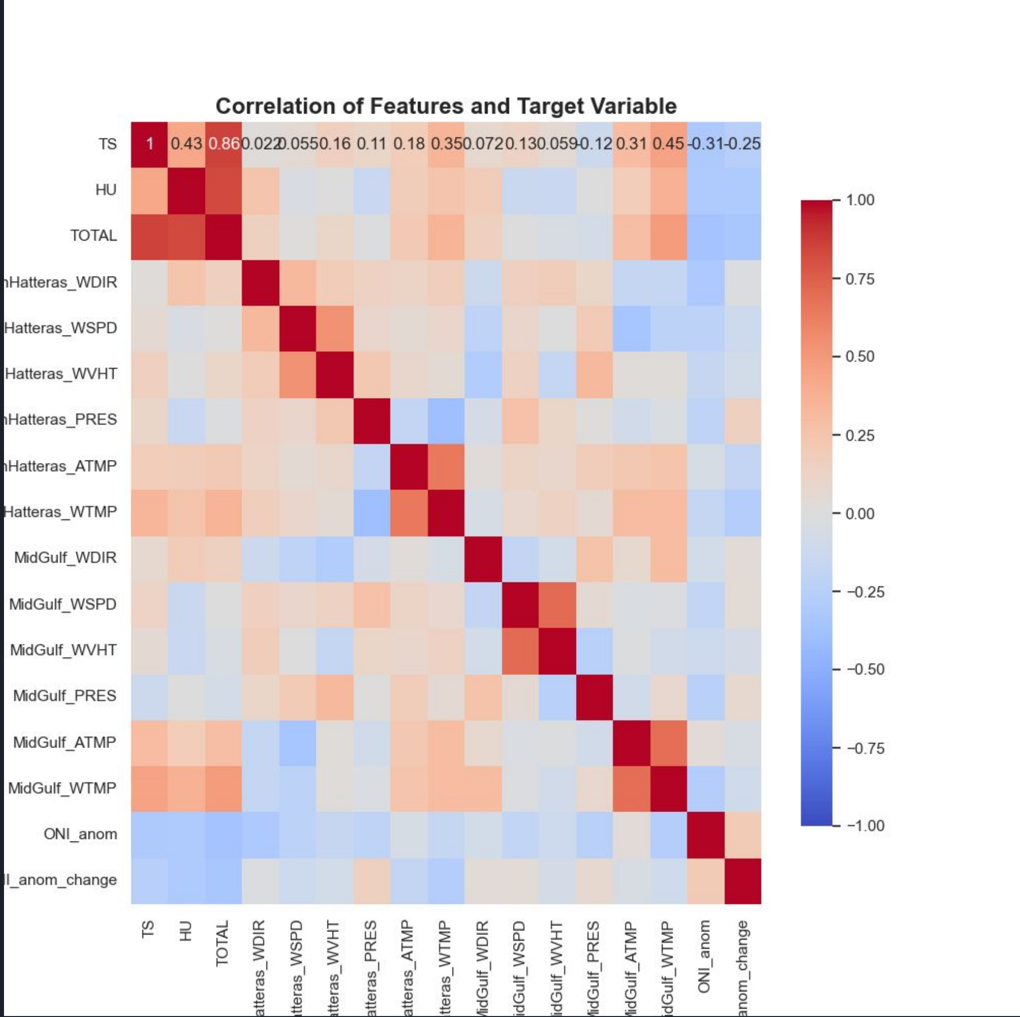
Map and Data source: https://www.ndbc.noaa.gov/

See data/data_NDBC.ipynb to pull NDBC data directly into python.

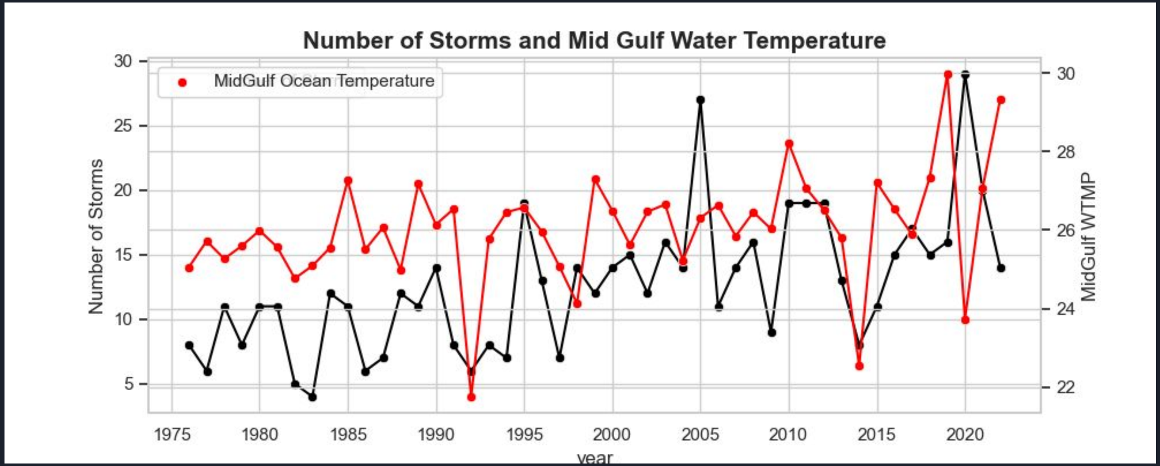
Exploratory Data Analysis: Correlation

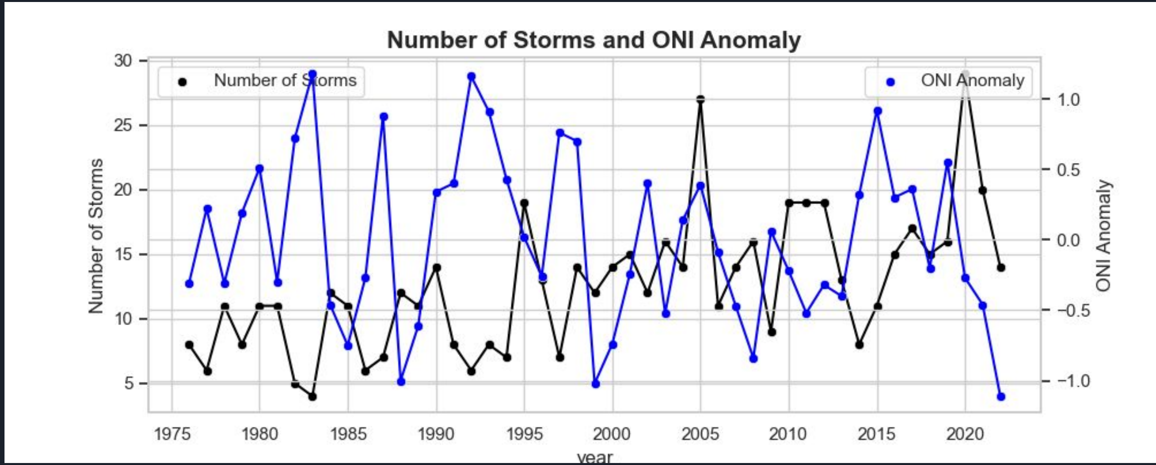
The variable we are using as the initial Target Variable is "TOTAL", which represents total named storms during a given season.

The rest of the variables (excluding TS and HU) represent ocean or meteorological averages for the month of May of that storm season (since May is the common time to issue hurricane season forecasts).



Exploratory Data Analysis





These are the features I expect to be most important in the training of the model. The first is May Mid-Gulf ocean temperature, which is meant to be a proxy for basin-wide sea surface temperature, and should be positively correlated with storm activity. It does seem to be somewhat loosely correlated.

The second feature shown here is May Oceanic Nino Index (ONI) anomaly from long-term historical average, which should be negatively correlated with storm activity.

Modeling & Results

Modeling Methodology

Feature Engineering: As part of Grid Search, implemented Polynomial Features with degree ranging from 1 to 6 on the aforementioned features

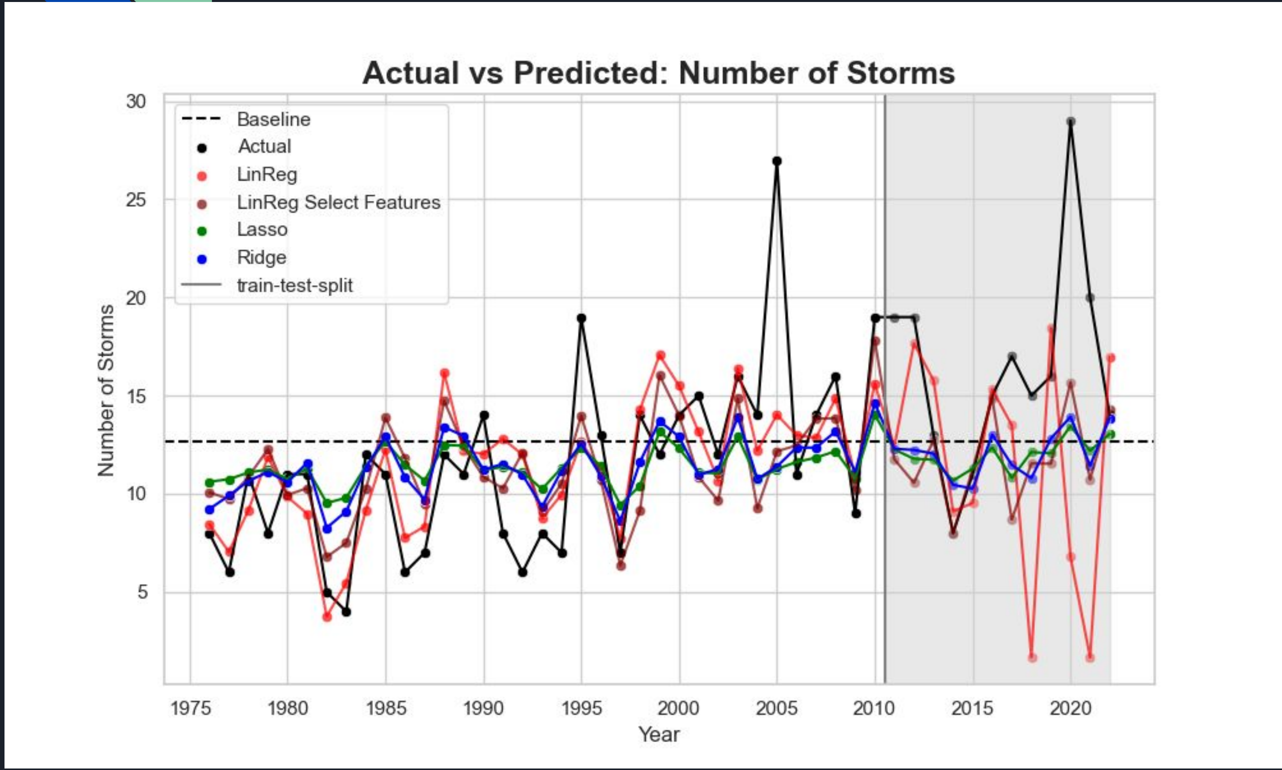
Train-Test Split: Years 1976 to 2010 included in training; 2011 to 2022 set aside for testing

Scaling: Feature scaling via StandardScaler(). Only numeric features are used.

Tested 4 different models using sklearn implementations, <u>with a Grid Search to find optimal hyperparameters</u>:

- DummyRegressor(): strategy = 'mean'; this creates a baseline to compare against
- LinearRegression(): only hyperparameter is polynomial degree for PolynomialFeatures()
- LinearRegression but limited to 4 most important features (WTMPx2 + ONIx2)
- Ridge(): sklearn's implementation of linear regression with L2 regularization
  - Hyperparameter: alpha (controls amount of regularization)
- Lasso(): sklearn's implementation of linear regression with L1 regularization
  - Hyperparameter: C (controls amount of regularization)

Results: Predicted Number of Storms

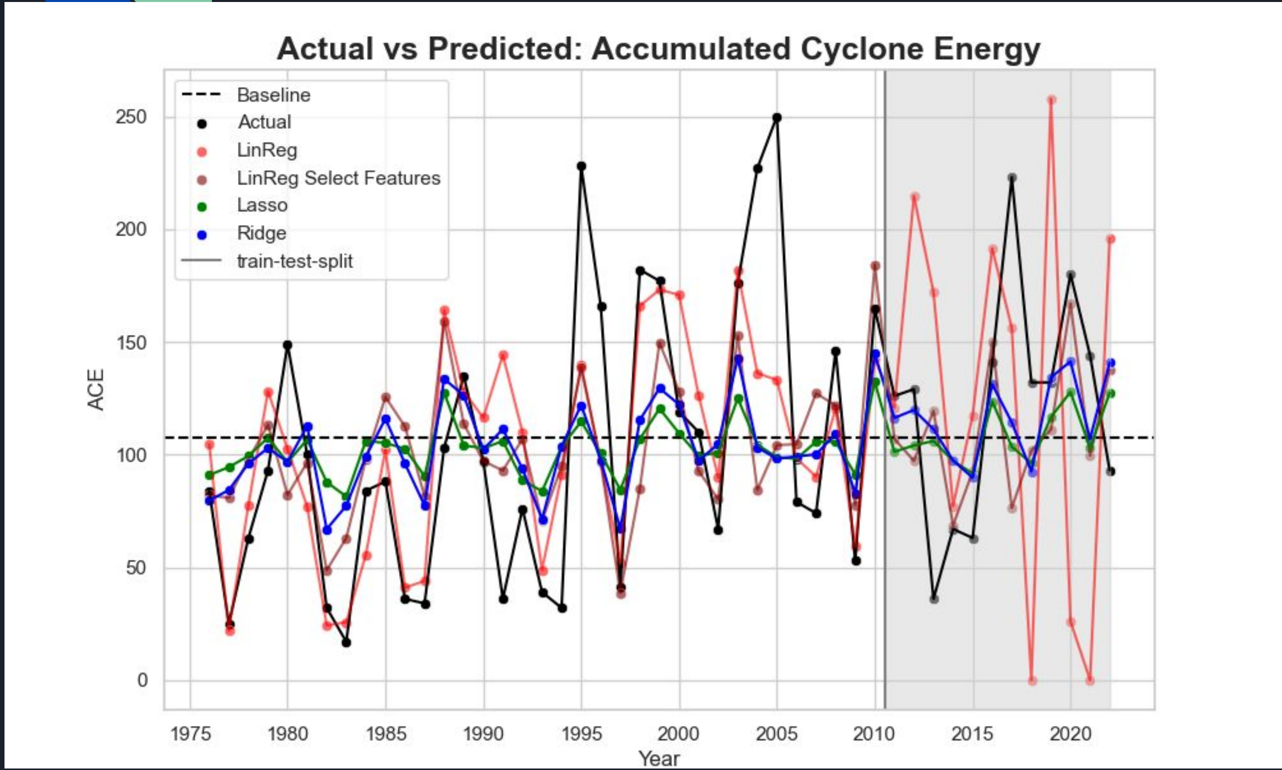


The plot shown depicts the actual and predicted number of storms for the various models. Recall that everything post-2010 was NOT used during training.

Observations:

- Full, unregularized linear regression performed terribly, like due to overfitting and noise from unimportant features
- Other models provide helpful signals (e.g. correctly detect a bump in 2020) but due to regularization and modeling limitations never capture the true range of outcomes in the test data.

Results: Total Accumulated Cyclone Energy



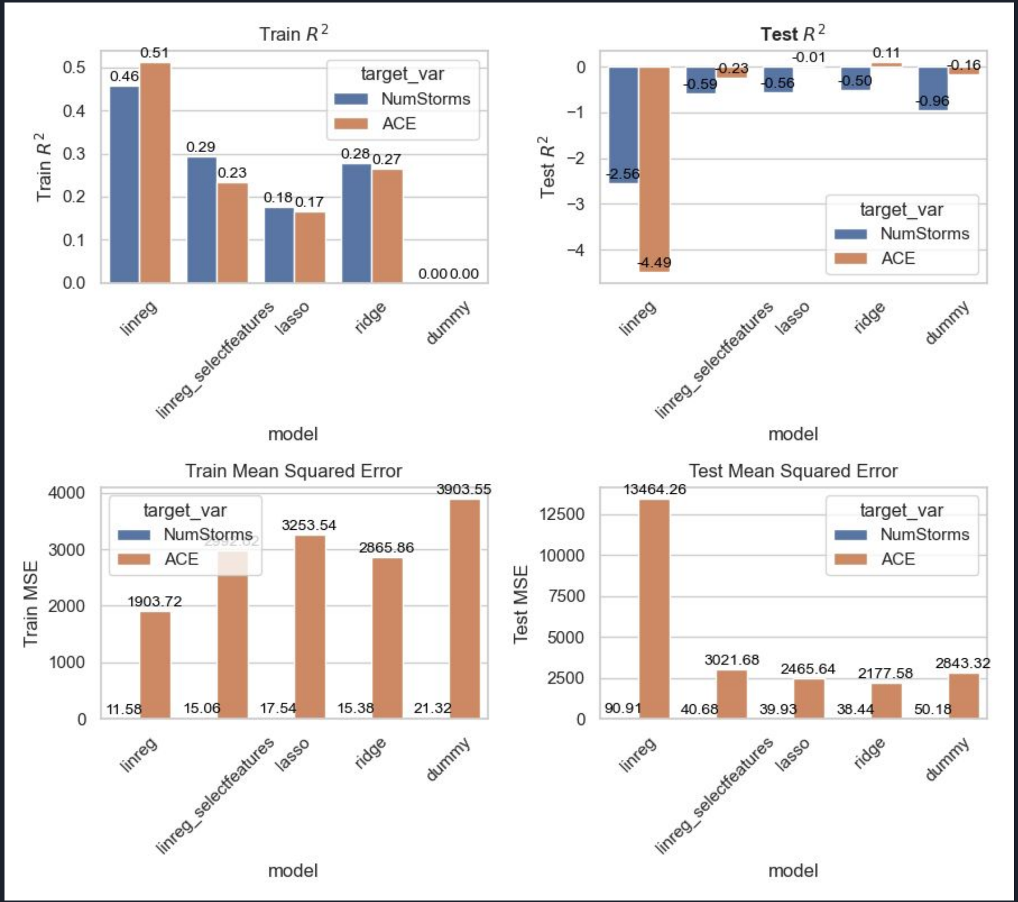
The same model setups were used to train to predict Total season ACE, with results shown here.

Error and R^2

In the models I developed, the Test R^2 scores were very poor, with values mostly below 0, indicating really poor model performance. The model does better predicting ACE than number of storms.

Lasso and Ridge vastly outperform linear regression and also outperform the linear regression with select features model.

Among the tested models, Ridge regularization yielded the best performance in terms of mean squared error and R^2. Despite the low Test R^2 scores, all the models outperformed the DummyRegressor() baseline model, perhaps suggesting some predictive value.



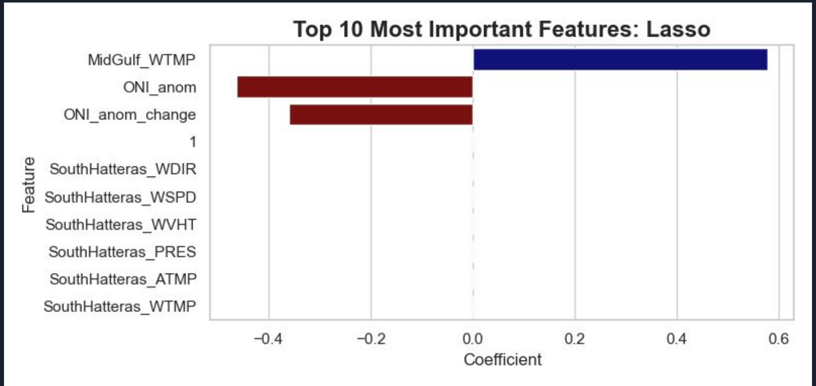
Most Important Features

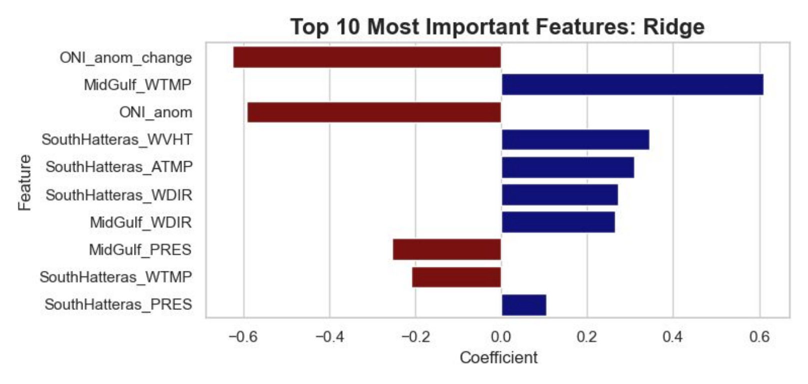
Ridge and Lasso regression are regularization techniques that add penalties to the model's coefficients (Ridge shrinks them uniformly, while Lasso can shrink some to zero), helping to prevent overfitting and aiding in feature selection by reducing the impact or entirely removing less important features.

The magnitude of the coefficients here imply importance, while sign (positive or negative) imply positive or negative correlation with storm activity.

For example, ONI is negatively correlated with storm activity according to the models (as expected), while water temperature (WTMP) is positively correlated.

Interestingly, the Lasso regression did not find South Hatteras water temperature to be a useful predictive feature.





Next Steps

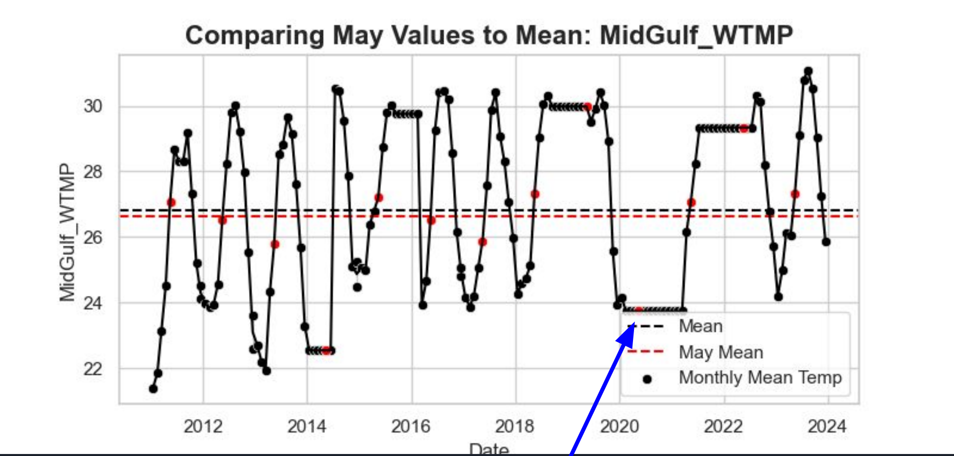
Possible Next Steps

- Monthly, rather than season predictions
- Incorporate more data, more features
- Improve data cleansing – e.g. for filling in missing data
- Find data that extends further back than 1976; see if substituting that data can produce a better model

Appendix

Discussion: Why did Models miss the very-active 2020 Hurricane Season?

- Moderate La Nina and slightly cooler than average sea surface temperatures at Mid-Gulf buoy (in May)
- Data issues in 2020 at both MidGulf and South Hatteras buoys (see right)
- Despite a record number of storms (29) in 2020, the Total ACE that year was only moderately high at 180. It's possible that our modeling doesn't capture storm generation as well as we'd like, since there are many additional meteorological factors (e.g. wind shear) that are not included in the training data here.



Looks like data was missing for Mid-Gulf buoy for most of 2020, and in our data preparation we forward-filled missing data. The last known temperature from January 2020 was 24 degrees C, much cooler than the May average of closer to 27 degrees. This is a likely factor in the model "missing" the very active 2020 season. Perhaps different data preparation could be used to populate the May 2020 Mid-Gulf buoy data with something closer to the May means rather than the January 2020 data.