* Elaborate on the nature of the response/target/Y to predict? Is it numeric or categorical? Is this a classification problem, or a regression problem?
* What is the machine-learning approach, and how can it be used to solve the problem you have identified?
* Elaborate on the features/X that you curated for each observation you collected? Are the features numeric or categorical? Speculate as to why they may be good predictors of the target.
* Detail your collection process. How many examples did you collect? What tools did you used to collect and curate the examples?
* Show some examples of observations that you collected.
* Present some basic summary statistics of the features.
* Compute a histogram of the response. If the response is categorical, comment on whether there appears to be a balanced number of observations from each class.
* Are there any missing values? If the response is missing from an observation, can the observation be utilized by a supervised machine-learning approach?
* Describe how, or which, observations you would set aside from the set you collected to be used to evaluate a trained machine-learning model.

Hey everyone, I’m Jimmy Lozano and today I’ll be focusing on wave height prediction using buoy and weather data. As I’m sure you all know, weather has a huge impact on just about any maritime activities, including navigation and safety of both the vehicle and personal. For my work, it primarily affects the capabilities of an unmanned vehicle and knowing wave heights with better accuracy would be ideal in the planning of testing these vehicles.

Machine learning algorithms are ideal for this as they can learn complex, non-linear relationships between input features (e.g., wind speed, atmospheric pressure) and wave height. This can lead to more accurate and nuanced predictions. Machine learning will also greatly benefit from the enormous amount of historical data that is available. Although for my project, I’ve curated just a year’s worth of data. However, continuously updating the algorithm with historical data will only serve to create a more accuarate and reliable model.

The target variable, wave height, is numeric and continuous in nature. Though, I really doubt we’d get tsunami level waves around here. From the observations it typically ranges between 0 and 6 feet as you’ll see in my code soon. Since Regression involves predicting a continuous numeric value based on input features, regression would be ideal here.

**1. Temperature**

**Why It May Be a Good Predictor:**

* **Impact on Wave Formation**: Temperature can influence the density of water and atmospheric pressure, which in turn affects wave formation. Higher temperatures might lead to warmer water and potentially impact wave dynamics.

**2. Wind Speed**

* **Type**: Numeric
* **Description**: The speed of wind at the buoy location, typically measured in meters per second (m/s) or kilometers per hour (km/h).
* **Range**: Wind speed can vary widely depending on weather conditions.

**Why It May Be a Good Predictor:**

* **Direct Influence on Waves**: Wind speed directly affects wave height. Higher wind speeds generally lead to higher waves due to increased energy being transferred to the water surface.

**3. Wind Direction**

* **Type**: Categorical (often encoded as angles)
* **Description**: The direction from which the wind is coming, measured in degrees (0° to 360°) relative to true north.

**Why It May Be a Good Predictor:**

* **Wave Propagation**: Wind direction affects the orientation and propagation of waves. Waves generally travel in the direction of the wind, so knowing the wind direction helps in understanding the direction in which waves are growing.
* **Interaction with Current**: Wind direction in relation to ocean currents can influence how waves form and interact, affecting their height.

**Why These Features Are Valuable**

1. **Complex Interactions**: Each of these features affects wave height through complex physical interactions. Machine learning models can learn these interactions and predict wave height more accurately than traditional methods.
2. **Correlation with Wave Height**:
   * **Temperature**: Although less directly correlated than wind speed, temperature still affects the physical properties of water and atmospheric conditions, which can impact wave height.
   * **Wind Speed and Direction**: These features are directly related to wave dynamics. Wind speed influences the energy transferred to the waves, while wind direction helps predict the wave's direction and how it may evolve.
3. **Historical Relevance**: Historical data on temperature, wind speed, and direction has shown a consistent relationship with wave height, making these features reliable predictors when used in conjunction with machine learning algorithms.
4. **Modeling Complexity**: Machine learning models can capture the nonlinear and complex relationships between these features and wave height. For example, neural networks or ensemble methods can model interactions between wind speed and direction, and temperature effects, which might be difficult to capture with simple regression models.