**Storm Predictor**

**Project Members:**

Blas Antunez \_\_\_\_\_

Robert Barnes – 8333-5538

Ivana Morales-Martinez \_\_\_\_\_

Marian Morales-Martinez \_\_\_\_\_

**Goal:**

The project will use historical data from the dataset the Atlantic Hurricane Dataset – HURDAT2 (available from <https://www.kaggle.com/datasets/jayc00009/atlantic-hurricane-dataset-hurdat2>) to predict the likelihood of a hurricane that passes through a given location will also pass over the University of Florida.

This project uses data science in a predictive manner to solve the real life problem of tropical storm preparation. Preparations for a major storm create significant cost for individuals and organizations (like the University of Florida) in terms of time and lost productivity. The team will use data science techniques from CAP4770 to analyze and manipulate the data set to provide relevant information to decision-makers at the University of Florida so these decision-makers can make decisions relation to storm preparation (e.g., filling sandbags, securing doors and windows, relocating vehicles) or class cancellation in a more accurate and timely manner. This will enhance the primary mission of the University (providing education to tens of thousands of students) by not unnecessarily distracting operations given a low-probability storm, and not delay actiosn for a high-probability storm.

The team will certainly use data preprocessing techniques to ensure the data set is useable and relevant. Furthermore, the team will leverage pattern mining to identify frequent patterns, correlations, and associations between the attributes of historical storms (e.g., location, intensity, direction) and a notional current storm, as well as clustering techniques to provide an actionable probability to University decision makers.

**Details:**

From the source website: “This dataset is a cleaned and structured version of the NOAA HURDAT2 Hurricane Dataset (1851-2023). The original dataset is provided in an unstructured .txt format and has been converted into a standardized .csv format to facilitate ease of use in data analysis and machine learning applications.” The data include the following relevant fields, in addition to several not used:

Storm ID……………………………………A reference number unique to each storm (primary key)

Storm Name………………………………A clear text name commonly used for the public

Number of Observations……………..The number of storm measurements

Date…………………………………………Date of observations

Time…………………………………………Time of observations

Status of System………………………..Is it a Tropical Storm or Hurricane?

Latitude/Longitude……………………..Location of the storm

Maximum Sustained Wind Knots…..The speed of storm winds

**Possible Innovations:**

Some possible innovations include clustering locations of observations, rather than taking each location naively, resulting in a vector-based approach to prediction. Furthermore, separating storms into eras (early 1900’s, 1950’s, 1960’s, etc.) would provide some clarity on evolution of storm tracks.