Nongjie Feng, Rhys Leahy, and Carter Rogers

Prof. Amir Jafari

DATS 6103

8 November, 2021

Final Project Proposal: Great Lakes Ice Concentration

For our final project, we will implement various data mining techniques to predict ice concentration for each of the Great Lakes (Huron, Ontario, Michigan, Erie, and Superior). We chose this topic because we are curious to see if the well-documented effects of climate change have impacted the ice levels in the Great Lakes in recent years. The data for this project will come from NOAA CoastWatch's Great Lakes statistics, where we can access surface water temperature data and ice concentration levels for the Great Lakes. The data will need to be cleaned as the website contains separate .csv files for each lake and year. One technique that we will use frequently is linear regression since we will be predicting a numeric variable. We intend to use several modeling and visualization Python packages such as Pandas, Matplotlib, and Sklearn. For background on this topic, we will need to research normal ice concentration levels for each of the lakes in order to understand their unique behaviors. To judge the accuracy of our models, we will lean heavily on linear regression model summary statistics such as Mean Squared Error. The schedule for the completion of the project is as follows:

Proposal and Data Cleaning: completed by 11/10

Modeling: completed by 11/17

GUI: completed by 11/24

Link to database: https://coastwatch.glerl.noaa.gov/statistic/statistic.html