Sea It To Believe It: Machine Learning-Based Prediction of Harmful Algal Bloom (HAB) Intensity

Ye, Olivia (School: College Park High School)
Wang, Ada (School: College Park High School)

In recent years, insurgences of red tide, harmful algal blooms (HAB), caused by concentrates of the dinoflagellate Karenia brevis, have yielded millions of marine wildlife deaths, respiratory issues in humans, and an annual loss of 82 million dollars in commercial activity. Yet, contemporary prediction measures fail to recognize the environmental conditions that may aggravate blooms and employ unreliable, computationally expensive models. The researchers' means of achieving a comprehensive red tide prediction were two-fold: to determine the most accurate machine learning model and influential environmental factors. The collected historical data from the four factors of nutrient concentration, river discharge, wind velocity, and difference in sea surface height, were applied to three supervised machine learning classification algorithms: SVM, KNN, and Random Forest. 21 different models were trained on a twenty-year span of combinations of factors and K.brevis counts and tested on recent HAB blooms. The role of each factor was evaluated in predicting bloom intensity through statistical analysis; a two-tailed paired t-test indicated the SVM algorithms performed significantly better than the KNN (p-value <0.020). The most successful model developed utilized the SVM algorithm, nutrient concentration, and river discharge with an accuracy of 96.65% and a 0.95 F1 score. The model's novel focus on a variety of factors produced an average accuracy 10% greater than models in existence. With these developments, a mobile application was developed for the public to view forecasted conditions of toxicity based on the model's real-time predictions to ensure public safety.