Course: Erdos Deep Learning - Summer 2024
Team Name: Ocean Optimizers (New Atlantis)
Team Members: Margaret Swerdloff, Junaid Hasan, Meltem Uyanik, Maitituerdi Aihemaiti, Aryama Singh

Executive Summary:
Our project investigates the impact of ocean variables such as water temperature, salinity, and pH on dissolved oxygen levels and their subsequent effects on phytoplankton concentrations in the Gulf of Mexico. Phytoplankton, crucial to marine ecosystems as both the base of the food chain and major oxygen producers, are directly influenced by changes in dissolved oxygen, which reflect overall marine health. Our goal is to predict oxygen depletion in marine ecosystems to identify potential risks to phytoplankton sustainability, fishery health, marine biodiversity and guide conservation efforts. We used the NCCOS Coastal Pollutants Dataset from 2003 to 2017 to study the role of ocean variables in oxygen concentration.

To prepare the data for analysis, we applied K-Nearest Neighbors (KNN) imputation to handle missing values, ensuring a complete dataset without losing critical information. To understand the relationship between the various ocean variables, we used an autoencoder architecture to generate parameters that arise from the data. We explored different autoencoder architectures, including a 3-dimensional embedding generated through an autoencoder with hidden layers of sizes 8, 5, and 3. This architecture allowed for the extraction of complex patterns within the data, enriching the feature set and improving model performance. XGBoost was employed as our primary model due to its robustness and efficiency in handling large, complex datasets. After cross-validation and hyperparameter tuning, our best model, using a 7-dimensional embedding without hidden layers, achieved a Mean Squared Error (MSE) of 0.0077 and an R^2 score of 0.9991, demonstrating strong predictive capabilities.

Our SHAP analysis revealed that water temperature is the most important factor influencing dissolved oxygen levels. Air temperature and barometric pressure also play major roles, with higher temperatures likely leading to lower oxygen levels. Salinity and pH contribute to the predictions as well, though their impact is less pronounced compared to temperature and pressure. These findings highlight the complex interplay between environmental variables and their combined impact on dissolved oxygen, which is crucial for sustaining phytoplankton populations and, in turn, the overall health of marine ecosystems.

Our findings highlight the significant role of environmental changes on ocean oxygen levels, which in turn affects the health of marine ecosystems. Dissolved oxygen levels are closely linked to phytoplankton populations, which are essential for ocean health. The enhanced predictive power of our models, achieved through the use of both simple and complex autoencoder architectures, can help identify regions at risk of oxygen depletion, guiding targeted conservation strategies. While the study has limitations, including potential biases from imputed data, it offers a robust framework for understanding and mitigating the effects of climate change on marine life. Future work will focus on further expanding the dataset and refining the model architecture to improve accuracy over time.