**Project Topic: Predictive Analysis of Sea Level Rise**

*Hongxin Wu*

**What are we trying to solve?**

The project aims to develop a predictive model to forecast the sea level rise on *California's coastline* over the upcoming years, utilizing historical and current climate data. This model will help identify potential impacts on coastal communities, infrastructure, and ecosystems due to rising sea levels, including increased flooding, coastal erosion, and loss of habitat.

**Why is this important?**

Understanding and predicting the rise of sea levels is crucial for preparing and adapting to climate change. It allows for informed decision-making in urban planning, conservation efforts, and policy-making to mitigate the adverse effects on human and natural systems. Accurate predictions are vital for developing strategies to protect vulnerable communities and ecosystems and for ensuring sustainable development along coastlines.

**Where will the data come from?**

Data will be sourced from:

* Satellite observations for sea surface height (e.g., from NASA's satellite missions like Jason-3).
* <https://podaac.jpl.nasa.gov/dataset/JASON_3_L2_OST_OGDR_GPS>
* Historical sea level records from tide gauges (e.g., NOAA's National Water Level Observation Network).
* <https://tidesandcurrents.noaa.gov/sltrends/>
* Climate models and projections (e.g., from the IPCC's Assessment Reports).
* <https://www.ipcc.ch/site/assets/uploads/2018/02/ar4-wg1-chapter8-1.pdf>
* <https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter09_FINAL.pdf>
* Ocean Surface Topography from Space
* <https://sealevel.jpl.nasa.gov/data/get-data/>
* <https://sealevel.jpl.nasa.gov/resources/?page=0&per_page=25&order=pub_date+desc&search=&condition_1=1%3Ais_in_resource_list&category=211>
* <https://www.epa.gov/sites/default/files/2021-04/documents/sea-level_td.pdf>

**How will it need to be refined?**

The data refinement process will include:

* Cleaning: Removing inaccurate or irrelevant data points.
* Normalization: Standardizing the scale of different datasets to enable meaningful comparisons.
* Interpolation: Filling in missing values in historical records to maintain consistency.
* Outlier Detection: Identifying and handling statistical outliers to improve model accuracy.

**What method will be used?**

The project will implement a two-pronged approach to forecasting sea level rise:

* **Time Series Analysis with ARIMA (Autoregressive Integrated Moving Average):** This method will be used for linear trend analysis, suitable for datasets showing stationarity or where linear trends have been identified.
* **Deep Learning with LSTM (Long Short-Term Memory) Networks:** For capturing complex, non-linear relationships in the data, especially useful for integrating various predictors of sea level rise, such as temperature anomalies and ice melt rates.

The choice of models allows for a comprehensive analysis, leveraging ARIMA for its efficiency in handling time series data and LSTM for its ability to process sequences of data and capture long-term dependencies.

**What metrics will be used to evaluate success?** The model's performance will be evaluated using the following metrics:

* **Root Mean Square Error (RMSE):** Measures the model's prediction accuracy by quantifying the square root of the average squared differences between predicted and actual values.
* **Mean Absolute Error (MAE):** Provides an understanding of the average magnitude of the prediction errors, which is easy to interpret.
* **R-squared (Coefficient of Determination):** Indicates the proportion of the variance in the dependent variable that is predictable from the independent variables, useful for assessing the goodness of fit for linear models like ARIMA.

Additionally, model validation will involve comparing the forecasts against observed sea level changes not used in model training and against projections from authoritative sources like the IPCC. This step ensures that the model's predictions are realistic and aligned with current scientific understanding.

**Contribution and Impact**

The project's success is not solely determined by its predictive accuracy but also by its contribution to the broader scientific and policy-making community. A key outcome could be the development of a publicly available dataset or repository that combines historical sea level data, satellite observations, and climate model outputs in a refined, ready-to-use format for future research.

Moreover, documenting the carbon cost of the project – like embedding *CodeCarbon*, from data processing to model training and evaluation, aligns with the growing emphasis on sustainable AI practices. This aspect underlines the environmental impact of computational methods used in climate change research, contributing to a more holistic understanding of the trade-offs involved in deploying DS/ML/AI solutions.

By addressing these elements, the project not only aims to provide actionable insights into sea level rise but also to enhance the resources available to the climate change research community and to underscore the importance of sustainability in scientific research methodologies.