**Proposal: Analysis of Arctic Sea Surface Height Anomalies Using Machine Learning and Deep Learning Techniques**

**Introduction and Problem Definition**

The Arctic region is experiencing profound impacts from climate change, with sea levels rising at a rate faster than the global average. This project aims to analyze and predict Sea Surface Height Anomalies (SSHA) in the Arctic using advanced machine learning and deep learning techniques. Understanding SSHA patterns is crucial for predicting and mitigating the effects of climate change, as they can indicate shifts in ocean currents and impact global climate patterns. Recent studies highlighting accelerated Arctic ice melt and its global implications underscore the importance of this research (Arctic Monitoring and Assessment Programme, 2017). The primary objectives are to identify and characterize spatial patterns in Arctic SSHA using machine learning techniques, develop predictive models for SSHA using both traditional machine learning and deep learning approaches, and assess the potential of these techniques in enhancing understanding of Arctic sea level changes.

**Data Collection and Refinement**

For this project, I will utilize the ICESat-2 ATL21 dataset, which provides gridded sea surface height anomaly measurements for the Arctic region. This dataset is particularly suitable for research due to its high-resolution spatial data (448 x 304 grid), inclusion of both daily and monthly aggregations, and provision of additional relevant variables such as geoid and mean sea surface measurements. These features will allow us to conduct a comprehensive analysis of SSHA patterns across the Arctic.

The data refinement process will involve several critical steps. First, I will address the issue of missing or invalid data (currently represented by the value 3.4028235e+38) using spatial interpolation techniques. Specifically, I will employ kriging interpolation, which is well-suited for geospatial data as it accounts for the spatial correlation structure of the measurements. This method will help preserve the spatial continuity of the SSHA patterns while filling in missing values. For grid cells surrounded by missing values, I will use a combination of temporal interpolation (using data from adjacent time periods) and spatial interpolation to ensure robust estimates. I will then normalize the SSHA values to account for regional variations, allowing for more accurate comparisons across different areas of the Arctic. To enrich the analysis, I plan to create derived features directly from the ATL21 dataset, including:

* Distance from the pole (calculated from provided grid\_lat and grid\_lon)
* Proximity to land (derived from the included land\_mask\_map)
* Temporal indicators from the delta\_time information

**Implementation Methodology**

The approach to analyzing the Arctic SSHA data will involve a multi-stage implementation process. I will begin with exploratory data analysis, using tools like matplotlib and Basemap to visualize SSHA patterns and calculate basic statistics. This initial exploration will help us identify potential anomalies and guide subsequent machine learning approaches.

The next stage will involve applying initial machine learning techniques. I plan to use K-means clustering to identify distinct SSHA pattern regions within the Arctic. This unsupervised learning approach will help us understand the natural groupings of SSHA patterns. I will also employ Random Forest regression to predict SSHA based on spatial and temporal features available in the ATL21 dataset. Spatial features will include grid coordinates, latitude/longitude positions, and derived metrics such as distance from coastlines, while temporal features will utilize the daily and monthly aggregations provided in the dataset. This analysis will provide insights into the most influential factors affecting sea surface height in different Arctic regions.

Building on these foundational techniques, I will then move to more advanced deep learning approaches. I plan to implement Convolutional Neural Networks (CNNs) for spatial pattern recognition in SSHA data. CNNs are particularly well-suited for identifying complex spatial relationships, which could be crucial in understanding the intricate patterns of sea surface height variations in the Arctic. If I am able to acquire sufficient temporal data from the ATL21 dataset, I will also explore the use of Long Short-Term Memory (LSTM) networks for temporal analysis, allowing us to capture time-dependent patterns in SSHA variations. Additionally, I will investigate the use of autoencoders for dimensionality reduction and feature learning, which could help us uncover latent patterns in the data that might not be immediately apparent through traditional analysis methods. Throughout the project, I will use Python, including libraries such as scikit-learn for traditional machine learning techniques, and TensorFlow and Keras for deep learning models.

**Evaluation Approach**

To ensure reliability of the findings, I will do a comprehensive evaluation strategy using various metrics appropriate for different aspects of the analysis. For the clustering and unsupervised learning approaches, I will use the Silhouette score and Calinski-Harabasz index to assess the quality of the identified clusters. These metrics will help us determine how well-separated and distinct the SSHA pattern regions are.

For regression tasks, such as predicting SSHA values, I will use Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to quantify the accuracy of the predictions. Also, I will calculate the R-squared (R²) score to understand how much of the variance in SSHA the models can explain. In cases where I develop classification models, I will evaluate their performance using metrics such as Accuracy, Precision, Recall, and F1-score, as well as the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) for a more comprehensive assessment of classification performance.

For deep learning models, I will monitor training and validation loss curves to ensure proper model fitting and to avoid overfitting. I will also evaluate the performance of these models on a held-out test set to assess their generalization capabilities. Beyond these quantitative metrics, I plan to conduct qualitative evaluations by visualizing the model predictions and comparing them with actual SSHA patterns. This visual assessment will be crucial in understanding how well the models capture the complex spatial patterns of sea surface height anomalies in the Arctic.

To measure and document the carbon footprint of this research, I will use CodeCarbon. CodeCarbon will be integrated into both the training and inference phases of all machine learning models, providing estimates of CO2 emissions based on power consumption and geographical location of the computing infrastructure.

By combining traditional machine learning techniques with advanced deep learning approaches, I aim to gain deeper insights into Arctic sea surface height anomalies. My hope is that this project will contribute to the understanding of climate change impacts in this region, potentially informing future research directions and policy decisions related to Arctic environmental management.