



REPORT

INNOVATORS

Team Members:

Shreya S Bhat

Bhavesb Mali

Aaditya Jadhav

PROBLEM STATEMENT - SOI 2024 - TWINS OF THE WINDS

Develop a Machine Learning Model which can take Sequential Data and Generate the Sea Surface Temperature. Sea Surface Temperature is one of the important factors in prediction of El Niño.

TASKS

- Train a Machine Learning Model on Labeled Data.
- Use the ML model to predict the Unlabelled Data.

STEP-BY-STEP PROCESS

LEARNING THE PROCESS

Learnt the process to predict required values from given data by watching some YouTube videos and reading related articles.

DATA UNDERSTANDING AND PREPARATION

Loaded the data and checked for any NULL values.

Tried various techniques to fill the NULL values (imputation): mean, median, mode, regression, MICE (Multivariate Imputation by Chained Equations)


Finally, filled the NULL values using IterativeImputer from sklearn(simpler implementation of MICE).

MODEL SELECTION AND TRAINING

Evaluated the performance of different models such as Linear Regression, K-Nearest Neighbors Regressor, Neural Network (Feedforward Neural Network) and Random Forests.

Models were compared using R2 score (coefficient of determination) and mean squared error metrics.

Selected Random Forests as its performance was better than the others.



Model was trained, evaluated and saved to use it for prediction of sea surface temperatures in evaluation.csv and data_1997_1998.csv.

CONCLUSION

The final model selected for predicting the sea surface temperatures was a **Random Forest Regressor**, which provided the best performance compared to other models.

Model Performance

R2 score = 0.9789543540955655

Mean Squared Error (MSE) = 0.09944798105641219