ChloroCast: Tracking Nutrient Trends in Coral Waters

Project Report

The health of coastal ecosystemsespecially coral reefsdepends heavily on water quality indicators such as chlorophyll concentration. Elevated chlorophyll levels often signal nutrient pollution, which can threaten marine biodiversity.

This project focuses on building predictive models for chlorophyll concentration using regression techniques. I have analyzed data from the Burdekin River, a key catchment in Northern Australia, to explore temporal and seasonal trends.

Objectives

- Build regression models to predict chlorophyll concentrations over time.
- Examine how seasonal variations (month of the year) influence chlorophyll levels.
- Evaluate model performance using Mean Absolute Error (MAE) and R-squared metrics.

Dataset

Source: Burdekin River monitoring data

Key Features:

- CHL_QA_AVG: Chlorophyll concentration (g/L)
- SAMPLE_DAY: Date of observation
- STATION_ID: Sampling station ID

Derived Features:

- idx: Sequential index representing each day
- MONTH: Month extracted from the sample date

ChloroCast: Tracking Nutrient Trends in Coral Waters

Methodology

I explored two models using the Python statsmodels library:

1. Simple Linear Regression
- Predictor: idx (time)
- Goal: Capture long-term chlorophyll trends

2. Multiple Linear Regression
- Predictors: idx, MONTH
- Goal: Add seasonal variation to improve predictions

Model evaluation included:
- R-squared: Proportion of variance explained
- MAE: Mean Absolute Error
- Residual Analysis: To detect bias or anomalies

Modeling

Simple Linear Regression:
Creating a simple linear regression model to predict chlorophyll concentration using the day number (idx) as the predictor.

Multiple Linear Regression:
Creating a multiple linear regression model using both the day number (idx) and the month (MONTH) as predictors to account for seasonal variation.

ChloroCast: Tracking Nutrient Trends in Coral Waters

Results

Model Type	MAE	R-squared	Key Observations
Simple Linear Regression	~0.257	Moderate	Captures overall trend but lacks seasonal sensitivity
Multiple Linear Regression	~0.2497	Improved	Effectively models both trend and seasonal fluctuations

Even minor improvements in MAE can significantly impact environmental policy and forecasting.

Conclusion

This project demonstrated that incorporating seasonal variables into regression models improves chlorophyll prediction accuracy. The results support the hypothesis that nutrient dynamics are strongly season-dependent.

Future Work

To build on this project, future enhancements could include:

- Adding features such as rainfall, temperature, or turbidity
- Exploring advanced models like Random Forests or LSTM networks
- Deploying the model as a REST API for real-time monitoring

Tools & Technologies

- Language: Python
- Libraries: pandas, numpy, statsmodels, matplotlib, seaborn, scikit-learn
- Notebook Interface: JupyterLab

ChloroCast: Tracking Nutrient Trends in Coral Waters

Author

Sagar Shahari

GitHub Profile: https://github.com/maverick4code