AquaData BS: Monthly Analysis of Black Sea Water Parameters

BIDUSA
Big Data Unites Sciences and Arts

Co-funded by the Erasmus+ Programme of the European Union

Erasmus+

Table of contents

01

PROJECT HISTORY AND AUTHORS

How our project came to be

We are students from the 3rd High School of mathematics and Natural Sciences "Acad. Metodii Popov" in Varna – a city located right on the Black Sea coast.

Being close to the sea, the topic of preserving the natural state of the Black Sea is important and familiar to us, therefore making it an obvious pick for our topic.

02

ABSTRACT

Objectives

This study demonstrates seasonal and short-term variations in:

- ➢ Temperature of the Black Sea
- ➢ Oxygen concentration of the Black Sea
- ➢ Salinity of the Black Sea

Which help to identify environmental trends and assess the region's vulnerability to climatic and ecological changes.

Method

Data was obtained from the Institute of Oceanology in Varna, utilizing French-manufactured oceanographic sondes.

The dataset was processed and analyzed using linear regression algorithm with WEKA's time series forecasting tool, training the model on the last 8 hours of January 2023 to predict the first 3 days of February 2023.

Results

The model achieved an average accuracy of approximately 96.47% across all three parameters.
However, while the numerical predictions were close to actual values (especially for salinity), the model did not fully capture short-term fluctuations, especially in oxygen concentration.

03

DATA DESCRIPTION

Justification

This dataset contains real environmental data measuring key indicators from the Black Sea. Understanding changes in these parameters is critical for assessing **marine health**, **ecosystem stability**, and the **effects of climate change** in the region as well as its importance as a strategic and an economic resource.

Dataset Source

The dataset was provided by the Institute of Oceanology and Hydrology in Varna, which conducts marine research and environmental monitoring within the Black Sea region.

Features

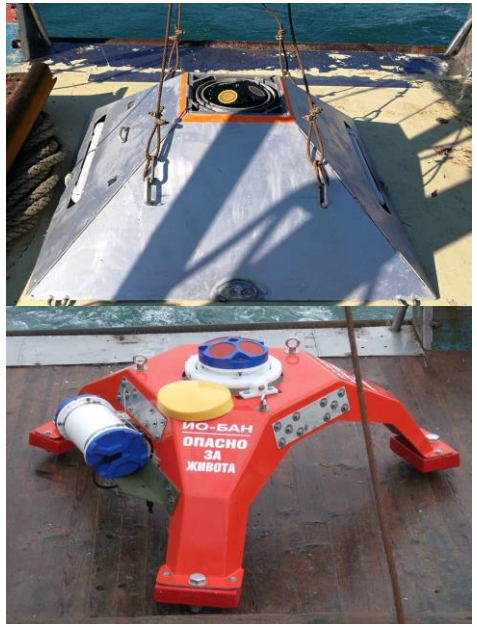
The dataset includes 53,000 entries with time-stamped values for:

- ➢ temperature (ºC)
- ➢ oxygen concentration (ml/L)
- ➢ practical salinity (PSU)

The data is structured in tabular format in ARFF files sorted by months with one measurement every 15 minutes.

Limitations

While working with the data, Weka detected a malfunction in part of the data, necessitating its removal. This step allowed Weka to visualize our dataset.
We encountered a few more issues while working with the program.



04

METHODOLOGY

Data Processing

File formats used:

- xlsx
- csv
- arff
- original raw data

The original data was provided in .xlsx format, which we converted into .csv and arff.

Analytical Techniques

We used WEKA's time series forecasting tool with the linear regression algorithm.

The model was trained on the final 8 hours of January 2023 to predict the first 3 days of February.

The achieved accuracy was ~96.47%.

Software

The following software was used:

- Microsoft Excel (for formatting)
- WEKA (for forecasting)
- Artificial Intelligence (for calculations).

Data was visualised using diagrams.

05

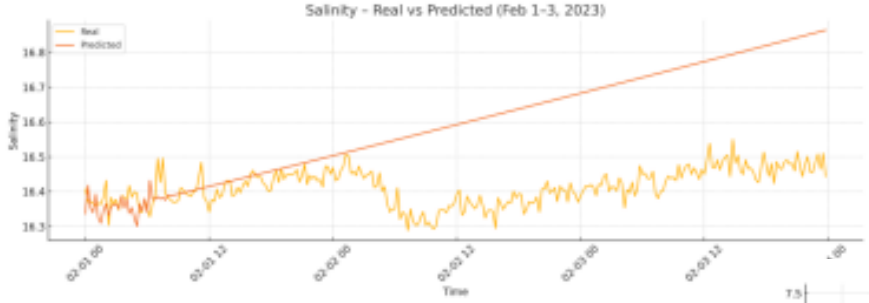
STATISTICS & DIAGRAMS

Key findings

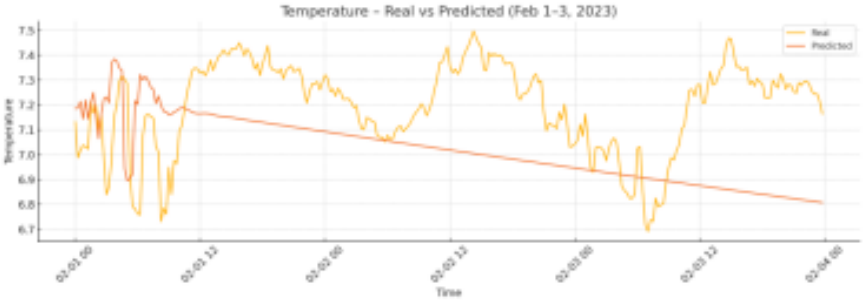
Overall model accuracy was 96.47% based on calculated MAPE (Mean Absolute Percentage Error);
(if the prediction has less than a 5% shift from the actual value it is counted for accurate prediction).
Although visual comparison showed that linear regression did not fully capture the real-time variation of oxygen concentration or the short-term temperature peaks.

Predictions

The model produced 288 forecasts for each variable (3 days of 15-minute intervals).
These predictions were compared to real February 2023 values, showing a strong correlation for salinity and temperature, but lower for oxygen. Visual charts confirmed these differences.



Salinity



Temperature

Oxygen

Implications

Machine learning tools like WEKA can help anticipate marine changes. Though simple models like linear regression lack precision in dynamic variables, they provide useful trend estimation that supports environmental monitoring and awareness in schools and local communities.

06

RESULTS

RESULTS

This project demonstrated how machine learning tools like WEKA can be used to analyze and forecast environmental changes in the Black Sea.While linear regression proved effective in capturing overall trends, it was less reliable for rapid or irregular changes, especially in dynamic parameters like oxygen. This reveals one of the key limitations of basic forecasting models when applied to complex natural systems.

07

CONCLUSIONS

Conclusion

This project demonstrated how machine learning tools like WEKA can be best utilised in order to analyze and forecast environmental changes in the Black Sea.

By working with real-world data from the “Quarantine” station near Varna, we were able to successfully predict temperature, oxygen concentration, and practical salinity values using a linear regression model.

Project goals

Collaborative efforts, such as the Erasmus project:
"Big Data Unites the Sciences and the Arts" (BIDUSA),
play a vital role in integrating scientific research with societal awareness and action.

The Project’s goal is to educate other countries on the importance of the Black sea.

08

REFERENCES

Sources used

The origin of our information was the Institute of Oceanology situated in Varna.
The data was collected and structured by the institute. We have only received the raw data.

Other sources:

- Bulgarian Academy of Sciences
  Institute of Oceanology – Varna

- Oguz, T. (2017)- The Black Sea Environment: A Regional Overview
- Journal of Marine Systems-European Environment Agency. (2024)
- European Sea Surface Temperature Report.
- Ginzburg, A. I., et al. (2022)
- Oxygen variability in the northwestern Black Sea
- Marine Ecology Progress Series. Zatsepin, A. G., et al. (2019).
- Acoustic Doppler Current Profilers in coastal sea monitoring.

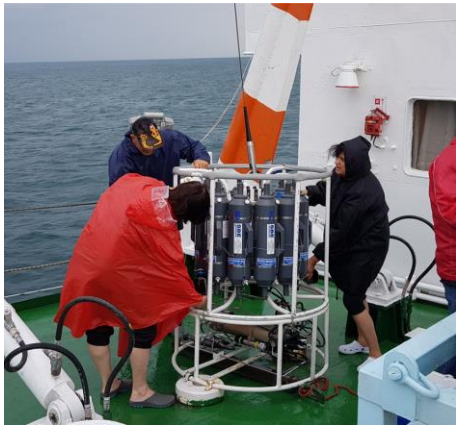
09

APPENDIX

As part of the Erasmus project ***"Big Data Unites the Sciences and the Arts" (BIDUSA)***, we used a variety of tools to collect and study data.
We worked with data gathered using: **two types of acoustic Doppler current profilers (ADCPs):**

- **ADCP Workhorse Sentinel 600kHz and**
- **SonTek ADP 1MHz.**

To process and explore the data, we used **Weka** which helped us sort, clean, and analyze everything we've collected.



Thank you for the attention!

Co-funded by the
Erasmus+ Programme
of the European Union