**A MACHINE LEARNING APPROACH TOWARDS OPTIMIZATION OF WATER QUALITY FOR CORAL GROWTH**

***A Project***

***Submitted in partial fulfillment of the requirements for***

***the award of the Degree of***

**BACHELOR OF COMPUTER APPLICATION**

**By**

**Adrija Sarkar**

**ROLL NO-12022004006010 AND REGISTRATION NO-223661010010**

**Naina Jha**

**ROLL NO-12022004006103 AND REGISTRATION NO-223661010102**

**Nitu Maity**

**ROLL NO-12022004006106 AND REGISTRATION NO-223661010105**

**Protyusha Pal**

**ROLL NO-12022004006121 AND REGISTRATION NO-223661010120**

**Rahi Dasgupta**

**ROLL NO-12022004006122 AND REGISTRATION NO-223661010121**



**DEPARTMENT OF COMPUTER APPLICATION**

**INSTITUTE OF ENGINEERING & MANAGEMENT**

**2025**

**DECLARATION CERTIFICATE**

This is to certify that the work presented in the thesis entitled **“A MACHINE LEARNING APPROACH TOWARDS OPTIMIZATION OF WATER QUALITY FOR CORAL GROWTH”** in partial fulfillment of the requirement for the award of degree of **Bachelor of Computer Application** of Institute of Engineering & Management is an authentic work carried out under my supervision and guidance.

To the best of my knowledge the content of this thesis does not form a basis for the award of any previous Degree to anyone else.

Date: 11.04.2025 Prof. Ankita Mandal

Dept. of Computer Application

Institute of Engineering & Management

Prof. Dr. Rupam Bhattacharya

Head of the Department

Dept. of Computer Application

Institute of Engineering & Management

**CERTIFICATE OF APPROVAL**

The foregoing thesis entitled **“A MACHINE LEARNING APPROACH TOWARDS OPTIMIZATION OF WATER QUALITY FOR CORAL GROWTH”** is hereby approved as a creditable study of research topic and has been presented in satisfactory manner to warrant its acceptance as prerequisite to the degree for which it has been submitted.

It is understood that by this approval, the undersigned do not necessarily endorse any conclusion drawn or opinion expressed therein, but approve the thesis for the purpose for which it is submitted.

**(Internal Examiner) (External Examiner)**

**Acknowledgements**

We would like to express our special thanks of gratitude to our Guide Prof. Ankita Mandal who helped us a lot in this project, her valuable suggestions helped us to solve tough challenges and without her help this project could not have been completed in time. Secondly, we would like to thank all our respected faculty members and our friends who helped us a lot in finalizing this project within the given time frame.

**Name of Student: Adrija Sarkar**

**Enrollment Number: 12022004006010**

**Name of Student: Naina Jha**

**Enrollment Number: 12022004006103**

**Name of Student: Nitu Maity**

**Enrollment Number: 12022004006106**

**Name of Student: Protyusha Pal**

**Enrollment Number: 12022004006121**

**Name of Student: Rahi Dasgupta**

**Enrollment Number: 12022004006122**

**Contents**

**Abstract**………………….…………………………………………………………………………………..v

**Chapter 1**

1.1 Introduction……….*…………………………..……………………………………….…………….*1

**Chapter 2**

2.1 Literature Survey…….…*………………………………………………………………….…….…*2

**Chapter 3**

3.1 Experimental Dataset….………………………………….*……………………………………..*4

**Chapter 4**

4.1 Proposed Methodology………………………………………….……………………….…..…5

**Chapter 5**

5.1 Results and Discussion…*…..……………………………………………………………….……*8

**Chapter 6**

6.1 Conclusions………………………………………….…………………………………………….……11

6.2 Future Work……………………………………….……………………..…………………………….12

**References** ………………………..…………………………………..………………..……………………14

**Abstract**

Coral reefs are among the most diverse and essential ecosystems on Earth, yet they are increasingly threatened by environmental changes and human activities. Water quality plays a crucial role in coral growth, and monitoring key parameters is essential for conservation efforts. This study focuses on analyzing water quality data using machine learning classification models to assess their effectiveness in predicting conditions suitable for coral growth. The research employs a comparative analysis of eight to ten supervised classification algorithms, evaluating their performance based on key water quality indicators such as pH, dissolved oxygen, temperature, and turbidity. The dataset used in this study is sourced from Indian water bodies, ensuring its relevance to regional coral conservation efforts. By applying machine learning techniques, the study aims to determine the most accurate model for classifying water quality suitable for coral ecosystems.

The evaluation of these models is based on metrics such as accuracy, precision, recall, and F1-score. Comparative analysis provides insights into the strengths and limitations of each algorithm in classifying water quality data. The results highlight which models perform best in predicting suitable conditions for coral survival, aiding in informed decision-making for ecological management. By leveraging data-driven insights, this study contributes to the advancement of machine learning applications in environmental conservation. The findings offer a scientific basis for improving coral reef management strategies, helping researchers and policymakers implement more effective conservation practices. Ultimately, this research aims to support sustainable coral growth and mitigate the impact of environmental stressors through predictive analytics.

**Chapter 1**

**1.1 Introduction**

Coral reefs are vital marine ecosystems that offer habitat, food, and coastal protection for a wide range of marine species. Despite their importance, these ecosystems are rapidly declining due to climate change, pollution, and human activities. Water quality plays a significant role in coral health, with parameters like temperature, pH, dissolved oxygen, and turbidity directly influencing coral survival. Monitoring these factors is essential for ensuring the sustainability of coral reefs.

Traditional water quality assessment methods involve manual observations and laboratory testing, which can be time-consuming and limited in coverage. With advancements in technology, machine learning provides a more efficient, data-driven alternative for analyzing and classifying water quality. Supervised learning models can predict whether a water body supports optimal conditions for coral growth, allowing for faster decision-making and improved conservation strategies.

This study aims to compare the performance of various classification algorithms in assessing water quality for coral survival. Using a dataset from Indian water bodies, eight to ten models are evaluated based on accuracy, precision, recall, and F1-score. The goal is to identify the most effective algorithm to aid marine biologists, conservationists, and policymakers in coral reef protection. This research highlights the potential of machine learning in supporting sustainable marine conservation efforts.

**Chapter 2**

**2.1 Literature Survey**

**Limitations of Traditional Water Quality Assessment Methods:**  
Conventional water quality monitoring relies heavily on manual sampling, laboratory testing, and observational analysis. While effective for precise measurements, these methods are resource-intensive, time-consuming, and not suited for large-scale or real-time assessments. Umanandini et al. [1] and Modasshir et al. [2] emphasize the inefficiencies of traditional models in predicting coral reef productivity or mapping coral species across vast regions. Additionally, Raghuraman et al. [11], [15], [17], and [18] outline how the lack of high-resolution temporal data and automated systems limits the early detection of environmental stressors impacting coral biodiversity in regions like the Andaman Islands and other Indian coral zones. These constraints hamper proactive conservation efforts, particularly in the face of rapid environmental changes.

**Adoption of Machine Learning for Automated Water Quality Assessment:**  
To address these challenges, researchers have turned to machine learning (ML) for improved prediction and classification of water quality. ML algorithms can process large datasets and evaluate complex relationships among parameters such as pH, temperature, turbidity, and dissolved oxygen. Pavithra and Kumar [3], along with Verma and Singh [4], successfully implemented ML and ANN models to predict water quality in the River Ganga. These approaches offer faster and more scalable solutions compared to manual assessments. Similar methods were applied by Singh and Tiwari [5] and Jha and Singh [6], demonstrating ML’s ability to generate reliable water quality indices for conservation planning. Moreover, Wang et al. [14] conducted a scientometric review outlining the evolution of coral reef studies under changing climates, stressing the growing role of data-driven methodologies.

**Role of Machine Learning in Environmental Monitoring:**  
Machine learning enables predictive analytics and scalable environmental monitoring, particularly for aquatic systems crucial to coral survival. Sahoo and Sahu [7] used ML models for monitoring the Mahanadi River, achieving high predictive accuracy. Kumar and Srivastava [8] applied ANN to the Yamuna River, highlighting ML's flexibility across different water bodies. Gupta and Pandey [9] and Sharma and Mittal [10] also utilized neural networks to assess and model pollution trends in major Indian rivers. These studies underscore ML’s adaptability and effectiveness in different geographic and ecological contexts. Furthermore, Martell et al. [12] and MacNeil et al. [13] emphasize that integrating complex environmental datasets with advanced algorithms is essential for understanding multifaceted stressors affecting coral ecosystems worldwide.

**Real-Time Prediction and Adaptive Learning Models:**  
The implementation of real-time and adaptive ML models marks a critical advancement in water quality assessment. Umanandini et al. [1] developed ensemble-based predictive models for coral productivity, capable of processing continuous data streams. These adaptive systems learn from new data over time, enhancing the model's ability to respond to evolving environmental conditions. Raghuraman et al. [11], [17], and [18] stress the need for dynamic monitoring tools in India’s coral-rich coastal regions, while Kleypas et al. [16] and Hoegh-Guldberg et al. [19] discuss the urgency of mitigating ocean acidification and climate change impacts on coral reefs. Hughes et al. [20] further support real-time systems, noting their importance in responding to coral bleaching events and other rapid environmental shifts.

The shift from manual water quality assessment to machine learning-driven approaches represents a significant step toward sustainable coral reef conservation. By improving speed, accuracy, and scalability, machine learning supports better-informed decision-making for ecosystem protection. This research builds on the foundation of existing studies to identify the most effective classification models for assessing water quality and coral viability. Ultimately, the integration of machine learning into environmental monitoring bridges the gap between data science and marine conservation, offering proactive solutions for preserving these fragile ecosystems.

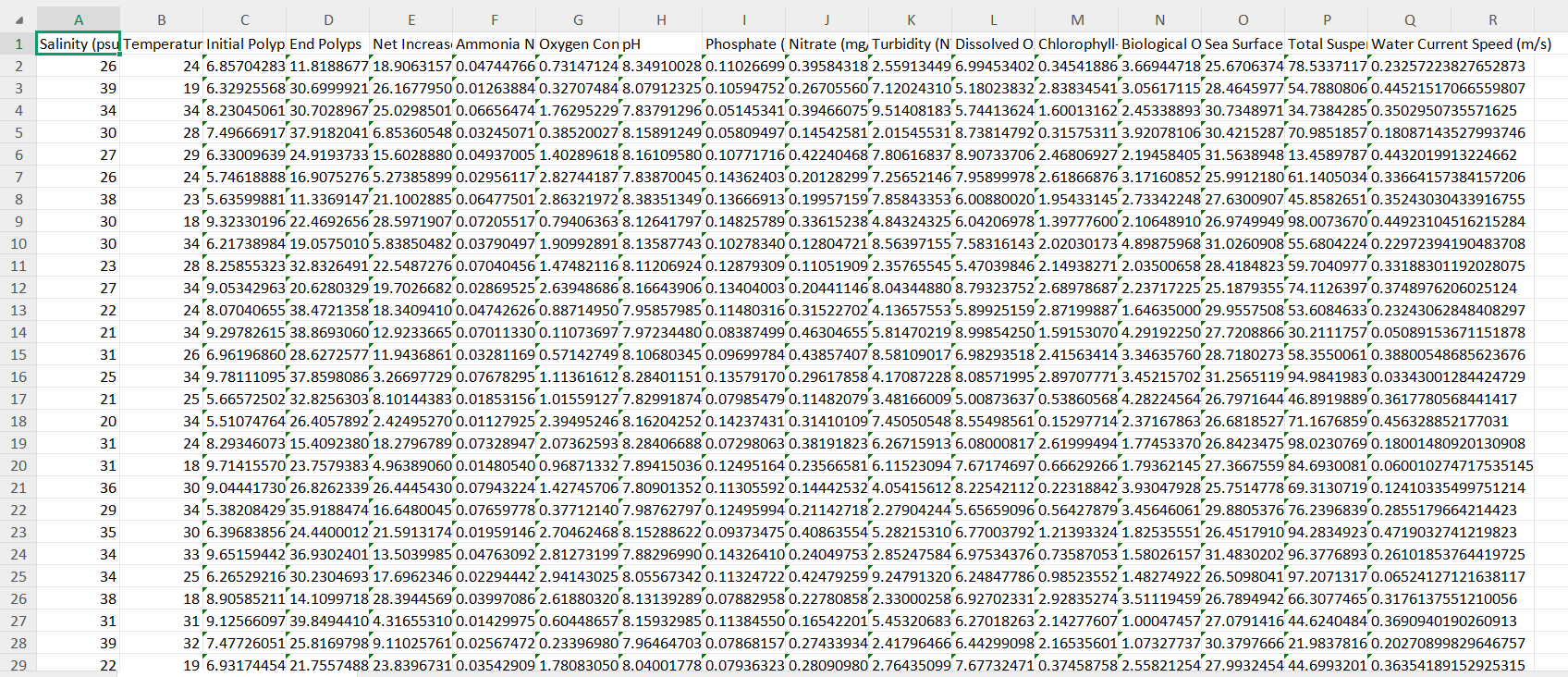
**Chapter 3**

**3.1 Experimental Dataset**

**Dataset:** <https://drive.google.com/file/d/1CXX4DTEQRFuU-PjssD7qSsXIl9ycP1wM/view?usp=sharing>

**Platform Used:** Windows

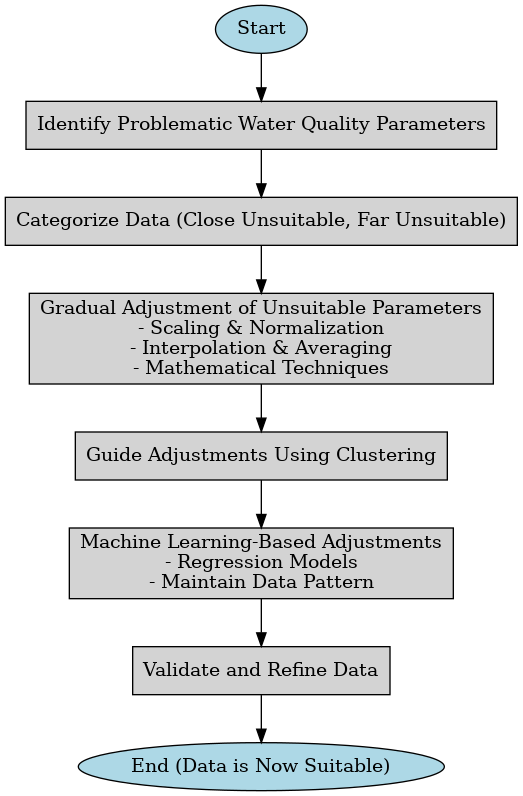
**IDE Used:** Visual Studios Code

****

**Fig. 1.** Experimental Water Quality Dataset used in the Project.

**Chapter 4**

**4.1 Proposed Methodology**

****

**Fig. 2.** Flowchart of Proposed Methodology for Converting Near Suitable Data into Suitable Data.

### **1. Identify Problematic Water Quality Parameters**

To begin, the water quality parameters in the dataset should be compared with known threshold values that define suitability for coral growth. This comparison helps identify which specific parameters are outside the acceptable range. By analyzing the degree of deviation from the suitable range, we can determine which values are potentially problematic and need further attention.

### **2. Categorization of Data Based on Suitability**

Once the problematic parameters are identified, the dataset can be divided into two categories based on how unsuitable the values are. The first category includes data points that are only slightly outside the ideal range, termed as far unsuitable. The second category contains data points that show significant deviation from the threshold, labeled as close unsuitable. This categorization allows for a more targeted approach to data correction.

### **3. Gradual Adjustment of Unsuitable Parameters**

To improve the quality of the data, unsuitable parameter values should be adjusted gradually. Scaling and normalization techniques can be used to nudge these values closer to the acceptable range while maintaining the original distribution of the data. Additionally, interpolation and averaging based on neighboring data points can provide better estimates for adjustments. Statistical and mathematical methods can also be applied to ensure that the modified values remain realistic and contextually appropriate.

### **4. Guiding Adjustments Using Clustering**

The close unsuitable cluster serves as a useful reference for understanding how slightly unsuitable data appears in comparison to suitable data. By studying the characteristics of this cluster, we can guide the transformation of the far unsuitable data. This guidance helps shift the severely deviated values more effectively toward the suitable range, using patterns observed in the less problematic cluster.

### **5. Machine Learning-Based Adjustments**

Regression models can be employed to predict the specific changes needed for each unsuitable parameter in the dataset. These models learn from existing patterns and relationships in the data to suggest appropriate modifications. It's important that these adjustments preserve the overall structure and trends of the original dataset, ensuring the modified data remains meaningful and consistent.

### **6. Validation and Refinement**

After making the adjustments, it is crucial to validate the results by checking if the new values fall within the acceptable range for water quality. If some values still remain outside the threshold, further refinements should be made. This process continues iteratively until all previously far unsuitable data points are successfully transformed into suitable values, ensuring the dataset meets the desired standards for coral health.

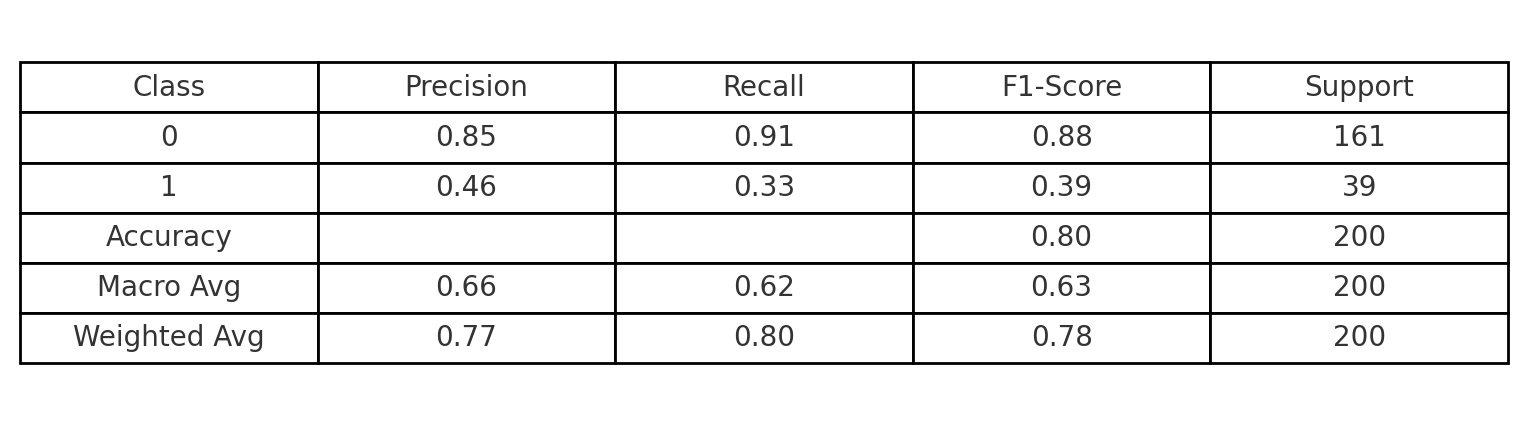
**Chapter 5**

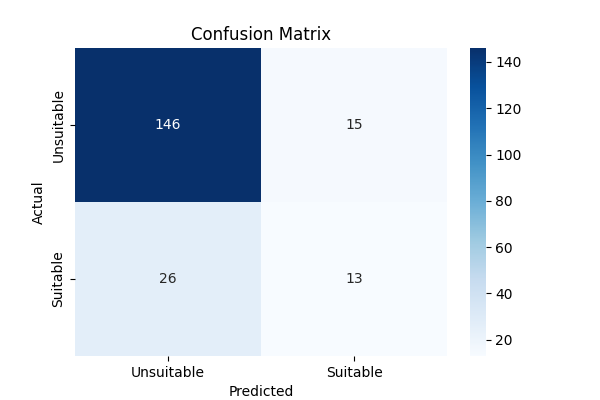
**5.1 Result and Discussions**

The proposed methodology was implemented and tested on a water quality dataset to assess its effectiveness in classifying and improving suitability for coral growth. In the first stage, the dataset was categorized into suitable and unsuitable classes using various machine learning classifiers. AdaBoost Classifier showed the highest accuracy and was selected as the primary model for classification. The unsuitable data was then further divided using unsupervised clustering methods. Both K-Means and Hierarchical Clustering were tested, with K-Means achieving a higher Silhouette Score. Hence, K-Means was chosen for effectively distinguishing between close unsuitable and far unsuitable data. To enhance the far unsuitable data, regression models were applied to suggest necessary parameter adjustments. Linear Regression performed poorly, while Decision Tree and Gradient Boosting showed perfect scores but indicated overfitting. Random Forest and Support Vector Regressor offered high accuracy and better generalization, making them the preferred models. The adjustments were applied using scaling, normalization, and statistical methods to ensure realistic and ecologically consistent changes. Validation confirmed the transformed data met suitability requirements for coral growth. A key strength of this approach is its ability to apply intelligent, data-driven adjustments by referencing close unsuitable data. This prevents overcorrection and supports gradual, ecologically sound improvements. The framework also allows iterative updates to adapt to environmental changes.

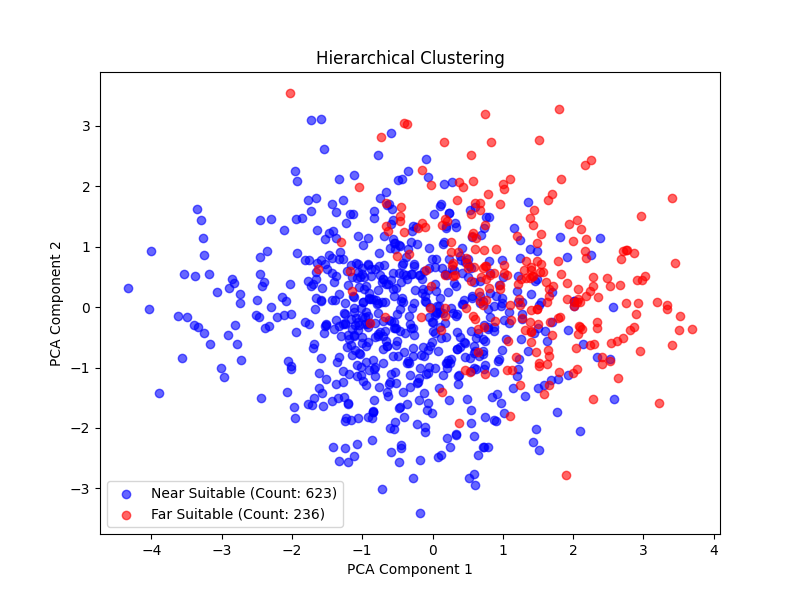
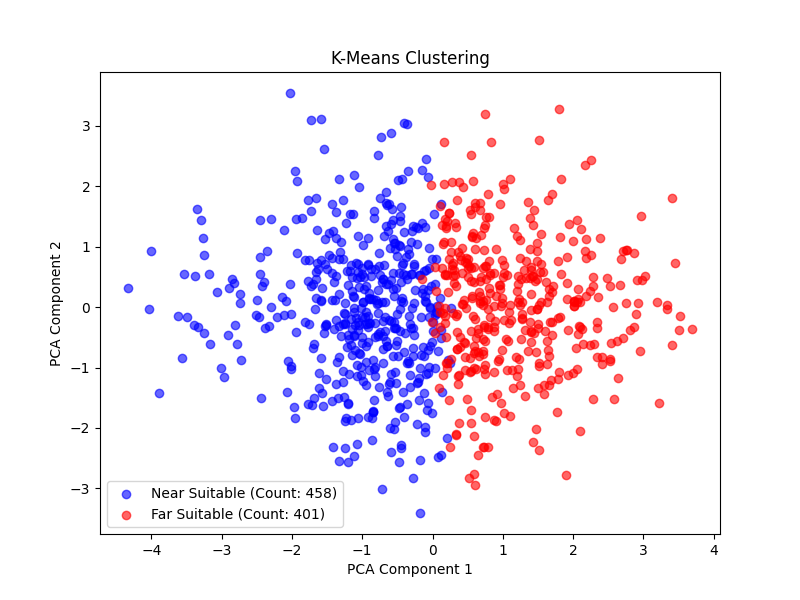
However, the system's effectiveness depends on the quality and diversity of the training data. Regular data collection and model updates are needed for real-world application. Overall, this study highlights the potential of machine learning in water quality management. By combining classification, clustering, and regression in a single pipeline, the approach offers a scalable and efficient solution for supporting coral reef conservation. Future work may include deep learning models and real-time monitoring for even greater impact.

**Table 1.** Classification Table of AdaBoost Classifier

****

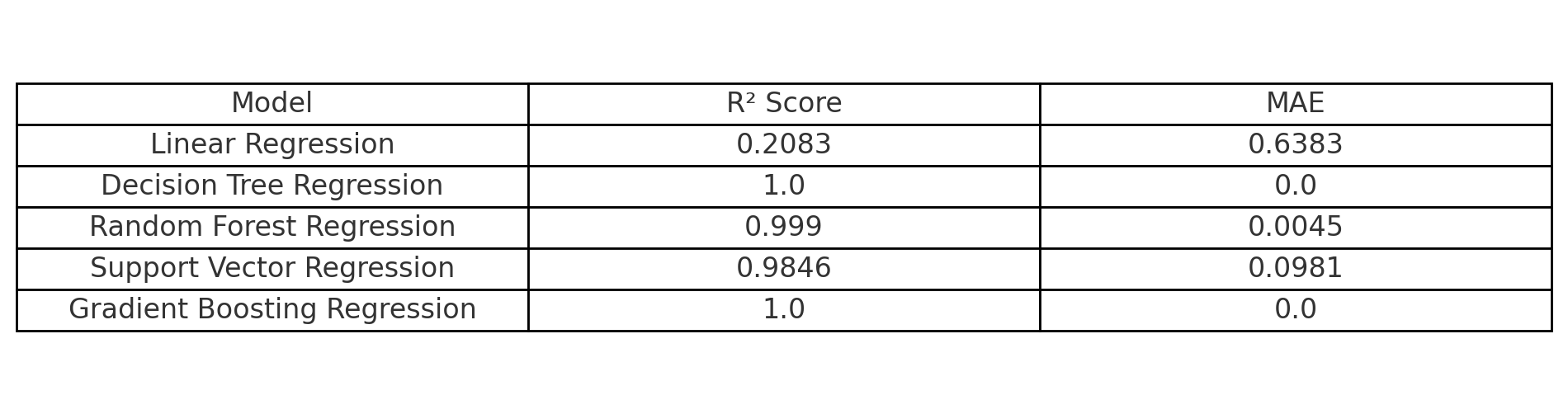
****

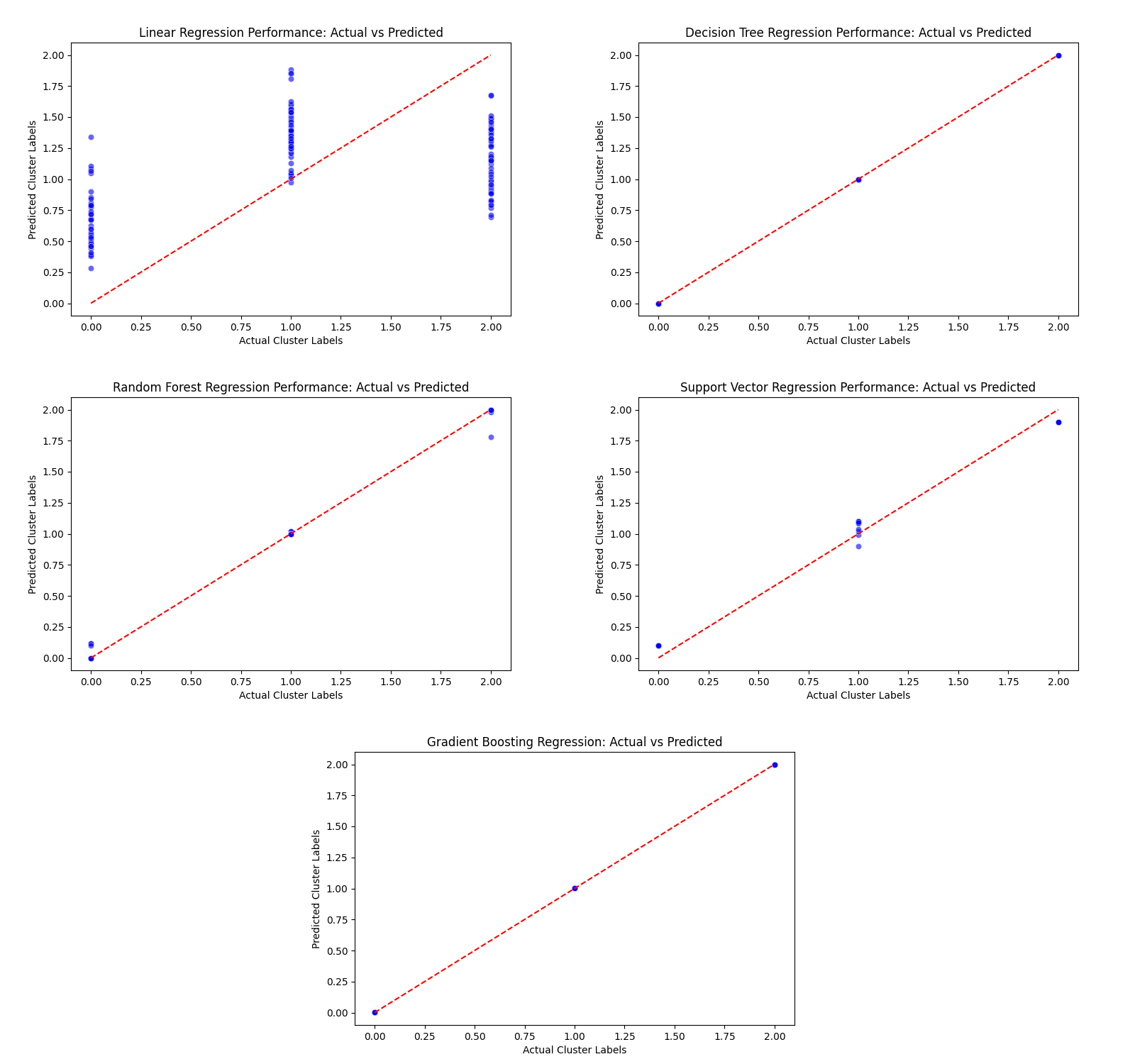
**Fig. 3.** Confusion Matrix of AdaBoost Classifier



**Fig. 4.**  PCA-based Scatter Plot for K-Means Clustering. **Fig. 5.** PCA-based Scatter Plot for Hierarchical Clustering.

**Table 2.** Comparison Table of Regression Algorithms Performance in Converting Near Suitable Data to Suitable.

****

****

**Fig. 6.**  Performance Comparison of Regression Models: Actual vs Predicted Cluster Labels

**Chapter 6**

**6.1 Conclusion**

In this study, we proposed a machine learning-based approach to optimize water quality parameters for coral growth by classifying and adjusting unsuitable data. By leveraging clustering techniques and regression models, our methodology effectively categorized and refined unsuitable water quality parameters, transforming them into more favorable conditions for coral ecosystems. The results demonstrated the effectiveness of gradual parameter adjustments, ensuring a balanced and data-driven improvement process.

Among the clustering techniques evaluated, **K-Means clustering outperformed hierarchical clustering**, achieving a **higher silhouette score**, which indicates more well-defined and distinct clusters. This made K-Means a more suitable choice for grouping water quality data based on coral suitability.

In the regression phase, multiple models were employed to predict and adjust unsuitable parameter values. **Linear regression yielded the lowest R² score**, indicating poor predictive performance and underfitting. On the other hand, **Decision Tree Regressor and Gradient Boosting Regressor achieved an R² score of 1**, suggesting overfitting to the training data and raising concerns about generalizability. **Random Forest Regressor and Support Vector Regressor emerged as the most reliable and balanced models**, delivering strong predictive performance without overfitting, making them the most suitable choices for the task.

Our approach provides a more efficient and scalable alternative to traditional water quality management methods by minimizing manual intervention and enabling automated, data-driven refinements. The adaptability of the model supports continuous learning and improvement, making it a promising solution for long-term water quality monitoring and optimization.

While the methodology has proven effective in a controlled setting, further enhancements—such as integrating advanced deep learning techniques and incorporating real-time monitoring—could significantly improve predictive accuracy and system responsiveness. Future studies could also investigate the ecological outcomes of these parameter adjustments in real coral reef environments to validate the approach in practical applications.

Overall, this research underscores the potential of machine learning in promoting environmental sustainability and highlights its transformative role in advancing water quality management practices for coral conservation.

**6.2 Future Work**

While this study has demonstrated the effectiveness of machine learning-based approaches in optimizing water quality for coral growth, there are several areas for further improvement and exploration.

**Integration of Real-Time Monitoring Systems**  
Future enhancements could include the implementation of IoT-enabled real-time water quality monitoring systems. This would allow for continuous data collection and automated adjustments based on real-time environmental changes, ensuring a more responsive and adaptive approach to water quality management.

**Incorporation of Advanced Deep Learning Models**  
Deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), could be explored to improve the predictive accuracy of water quality classification and adjustment. These models can capture complex relationships between parameters more effectively, leading to more precise recommendations.

**Validation with Real-World Coral Ecosystems**  
The proposed methodology should be tested and validated in real-world coral reef environments to assess its practical impact. Collaborations with marine biologists and environmental researchers could provide valuable insights into the biological effectiveness of the suggested improvements.

**Expansion of Dataset and Parameter Optimization**  
Collecting more diverse and extensive datasets from different marine ecosystems will help improve the generalizability of the model. Additionally, further research can be conducted on parameter optimization techniques to fine-tune the adjustment process for different environmental conditions.

**Development of a User-Friendly Application**  
A web-based or mobile application could be developed to allow researchers and environmental agencies to easily input water quality data and receive recommendations for necessary adjustments. Such a tool would improve accessibility and practical usability.

By addressing these future directions, the proposed methodology can be further refined and expanded, making it an even more effective tool for sustainable water quality management and coral conservation.

**References**

1. **S. Umanandini, M. Rishivardhan, A. Barathwaj S.R.Y., J. Jasline Augusta, S. Shrirang Sapate, S. Reenasree, M. Vigneash, "Predictive Model for Gross Community Production Rate of Coral Reefs using Ensemble Learning Methodologies," arXiv preprint arXiv:2111.04003, 2021.**

**[2] S. Modasshir, Y. Li, I. Rekleitis, "Coral Species Identification and Mapping Using Unsupervised Learning," OCEANS 2018 MTS/IEEE Charleston, 2018, pp. 1-6.**

**[3] P. Pavithra, S. S. Kumar, "Water Quality Prediction Using Machine Learning Algorithms: A Case Study of the River Ganga," International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 8, 2020, pp. 1046-1050.**

**[4] A. K. Verma, A. K. Singh, "Application of Artificial Neural Networks for Water Quality Prediction of River Ganga," Journal of Water Resource and Protection, vol. 9, no. 1, 2017, pp. 1-11.**

**[5] S. K. Singh, R. K. Tiwari, "Prediction of Water Quality Index Using Artificial Neural Network Approach for River Ganga at Kanpur," Environmental Claims Journal, vol. 32, no. 1, 2020, pp. 37-53.**

1. **M. K. Jha, V. K. Singh, "Water Quality Assessment and Prediction Modeling of Ganga River: A Review," International Journal of Environmental Sciences, vol. 5, no. 6, 2015, pp. 1394-1407.**

**[7] S. K. Sahoo, S. K. Sahu, "Water Quality Prediction of Mahanadi River Using Machine Learning Techniques," International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 7, 2020, pp. 1206-1210.**

**[8] A. Kumar, R. K. Srivastava, "Water Quality Assessment of River Yamuna Using Artificial Neural Network Approach," International Journal of Environmental Sciences, vol. 6, no. 1, 2015, pp. 57-68.**

**[9] N. Gupta, S. K. Pandey, "Water Quality Modeling of River Ganga Using Artificial Neural Network," International Journal of Environmental Sciences, vol. 5, no. 3, 2014, pp. 536-544.**

**[10] R. Sharma, A. K. Mittal, "Application of Artificial Neural Networks for Water Quality Prediction of River Yamuna," International Journal of Engineering Research & Technology (IJERT), vol. 3, no. 7, 2014, pp. 1871-1876.**

**[11] R. Raghuraman, C. Raghunathan, K. Venkataraman, "Threats to coral reef diversity of Andaman Islands, India: A review," Regional Studies in Marine Science, 2018.**

**[12] S.J.D. Martell, et al., "Multiple Stressors and Ecological Complexity Require a New Approach to Coral Reef Research," Frontiers in Marine Science, 2016.**

**[13] M.A. MacNeil, et al., "Coral Reef Monitoring, Reef Assessment Technologies, and Ecosystem-Based Management," Frontiers in Marine Science, 2019.**

**[14] Y. Wang, et al., "The Evolution of Coral Reef under Changing Climate: A Scientometric Review," Animals, 2023.**

**[15] R. Raghuraman, et al., "Coral Reef: Their Importance, Threats and Conservation Strategies," ResearchGate, 2013.**

**[16] J. Kleypas, et al., "Impacts of Ocean Acidification on Coral Reefs and Other Marine Calcifiers: A Guide for Future Research," Report of a workshop, 2006.**

**[17] R. Raghuraman, C. Raghunathan, K. Venkataraman, "Present Status of Coral Reefs in India," Book Chapter, 2013.**

**[18] R. Raghuraman, et al., "Coral Reefs in India," Journal Article, 2011.**

**[19] O. Hoegh-Guldberg, et al., "Impacts of Climate Change on Coral Reefs: A Global Perspective," Science, 2007.**

**[20] T.P. Hughes, et al., "Coral Bleaching and the Future of Coral Reefs," Science, 2017.**