Catfish Growth Rate and Average Weight Prediction Through Water Quality Parameters in Aquaponics System

Leveraging Data-Driven Approaches for Sustainable Aquaculture

John Kyle N. Junsay

Introduction

Aquaculture growth accounts for **53%** of global fish production by 2030 (FAO, 2024). Consequently, thhis industry also contributes significantly to the Philippine economy **(4.125 million metric tons annually)**.

Introduction

Why is Catfish farming vital as an aquaculture commodity?

- Cost effective
- High demand
- Easy to culture

Introduction

However, traditional farming methods often fall short in addressing critical factors such as water quality management and its impact on fish growth and average weight. Inefficiencies in monitoring and responding to water quality fluctuations can lead to resource waste, inconsistent growth rates, and reduced productivity.

Introduction

The primary objective of this study is to identify key water quality parameters and their temporal trends that influence catfish growth, providing actionable insights for sustainable aquaculture practices. By leveraging predictive modeling, this research aims to contribute to the development of data-driven tools for optimizing aquaculture productivity.

Introduction

Methodology

Results and Discussion

Conclusion

Recommendation

Methodology
Dataset Overview

- **12 aquaponics catfish ponds with over 170,000 records in total.**
- **Data collected at five-second intervals (temperature, turbidity, pH, ammonia, nitrate, dissolved oxygen, and biophysical attributes).**
- **Bi-weekly fish measurements for weight and length.**

Methodology
Workflow

Data Cleaning:

- **Removed missing values, standardized names, validated timestamps.**

Feature Engineering:

- **Created rolling mean, variance, and differences for water quality parameters.**
- **Derived growth rate from fish weight.**

Methodology
Workflow

Preprocessing::

- Normalization, standardization, and 80:20 data split.

Modeling:

- Linear Regression, Random Forest, and XGBoost.

Modeling:

- Metrics: RMSE, MAE, R^2.
- Validation: 5-fold Cross-Validation, LOO-CV.

Results and Discussion
Growth Rate Prediction
Model Performance (Train-Test)

Metrics	RMSE	MAE	R-squared
Linear Regression	1.499	1.201	1.0
Random Forest	0.3909	0.2627	0.9992
XGBoost	0.1371	0.1159	0.9999

Results and Discussion
Growth Rate Prediction
Model Performance (Train-Test)

Metrics	RMSE	MAE	R-squared
Linear Regression	1.499	1.201	1.0
Random Forest	0.3909	0.2627	0.9992
XGBoost	0.1371	0.1159	0.9999

Results and Discussion

Growth Rate Prediction

Model Performance (Train-Test)

Metrics	RMSE	MAE	R-squared
Linear Regression	1.499	1.201	1.0
Random Forest	0.3909	0.2627	0.9992
XGBoost	0.1371	0.1159	0.9999

Results and Discussion
Growth Rate Prediction
Model Performance (K-Folds Cross Validation)

K-Fold CV	RMSE	MAE
Linear Regression	1.4951	1.1283
Random Forest	0.4555	0.2749
XGBoost	0.1859	0.1468

Results and Discussion
Growth Rate Prediction
Model Performance (K-Folds Cross Validation)

K-Fold CV	RMSE	MAE
Linear Regression	1.4951	1.1283
Random Forest	0.4555	0.2749
XGBoost	0.1859	0.1468

Results and Discussion

Growth Rate Prediction

Model Performance (Leave One Out CV)

LOO-CV	RMSE Mean	RMSE Std Dev
Random Forest	0.1246	0.1596
XGBoost	0.1202	0.0847

Results and Discussion
Growth Rate Prediction
Model Performance (Leave One Out CV)

LOO-CV	RMSE Mean	RMSE Std Dev
Random Forest	0.1246	0.1596
XGBoost	0.1202	0.0847

Results and Discussion

Growth Rate Prediction

The findings clearly indicate that XGBoost is the most suitable model for predicting catfish growth rate, given its consistent performance and ability to generalize across different validation techniques. Random Forest also demonstrated strong performance and stability, making it a reliable alternative, especially when interpretability and simplicity are prioritized.

Results and Discussion
Average Weight Prediction
Model Performance (Train-Test)

Train-Test Validation	RMSE	MAE	R2
Linear Regression	4.2618	3.2205	1.0
Random Forest	0.4476	0.2267	0.9999
XGBoost	0.7584	0.6966	0.9998

Results and Discussion
Average Weight Prediction
Model Performance (Train-Test)

Train-Test Validation	RMSE	MAE	R2
Linear Regression	4.2618	3.2205	1.0
Random Forest	0.4476	0.2267	0.9999
XGBoost	0.7584	0.6966	0.9998

Results and Discussion
Average Weight Prediction
Model Performance (Train-Test)

Train-Test Validation	RMSE	MAE	R2
Linear Regression	4.2618	3.2205	1.0
Random Forest	0.4476	0.2267	0.9999
XGBoost	0.7584	0.6966	0.9998

Results and Discussion

Average Weight Prediction

Model Performance (K-Folds Cross Validation)

K-Fold CV	RMSE	MAE
Linear Regression	4.4418	3.4254
Random Forest	0.9503	0.3783
XGBoost	0.7021	0.5286

Results and Discussion

Average Weight Prediction

Model Performance (K-Folds Cross Validation)

K-Fold CV	RMSE	MAE
Linear Regression	4.4418	3.4254
Random Forest	0.9503	0.3783
XGBoost	0.7021	0.5286

Results and Discussion
Average Weight Prediction
Model Performance (Leave One Out CV)

LOO-CV	RMSE Mean	RMSE Std Dev
Random Forest	0.2412	0.3496
XGBoost	0.3585	0.2926

Results and Discussion
Average Weight Prediction
Model Performance (Leave One Out CV)

LOO-CV	RMSE Mean	RMSE Std Dev
Random Forest	0.1246	0.1596
XGBoost	0.1202	0.0847

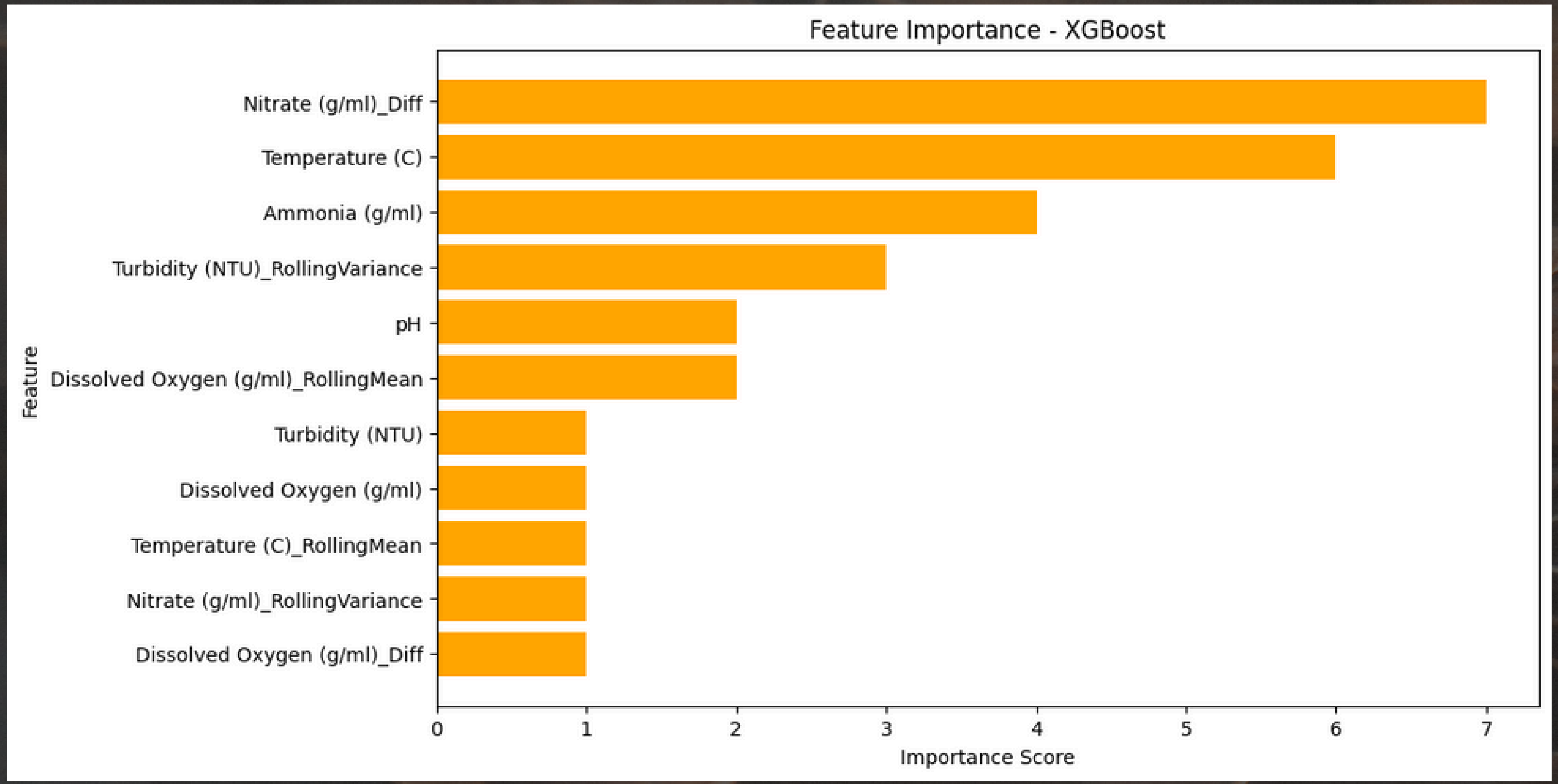
Results and Discussion
Average Weight Prediction

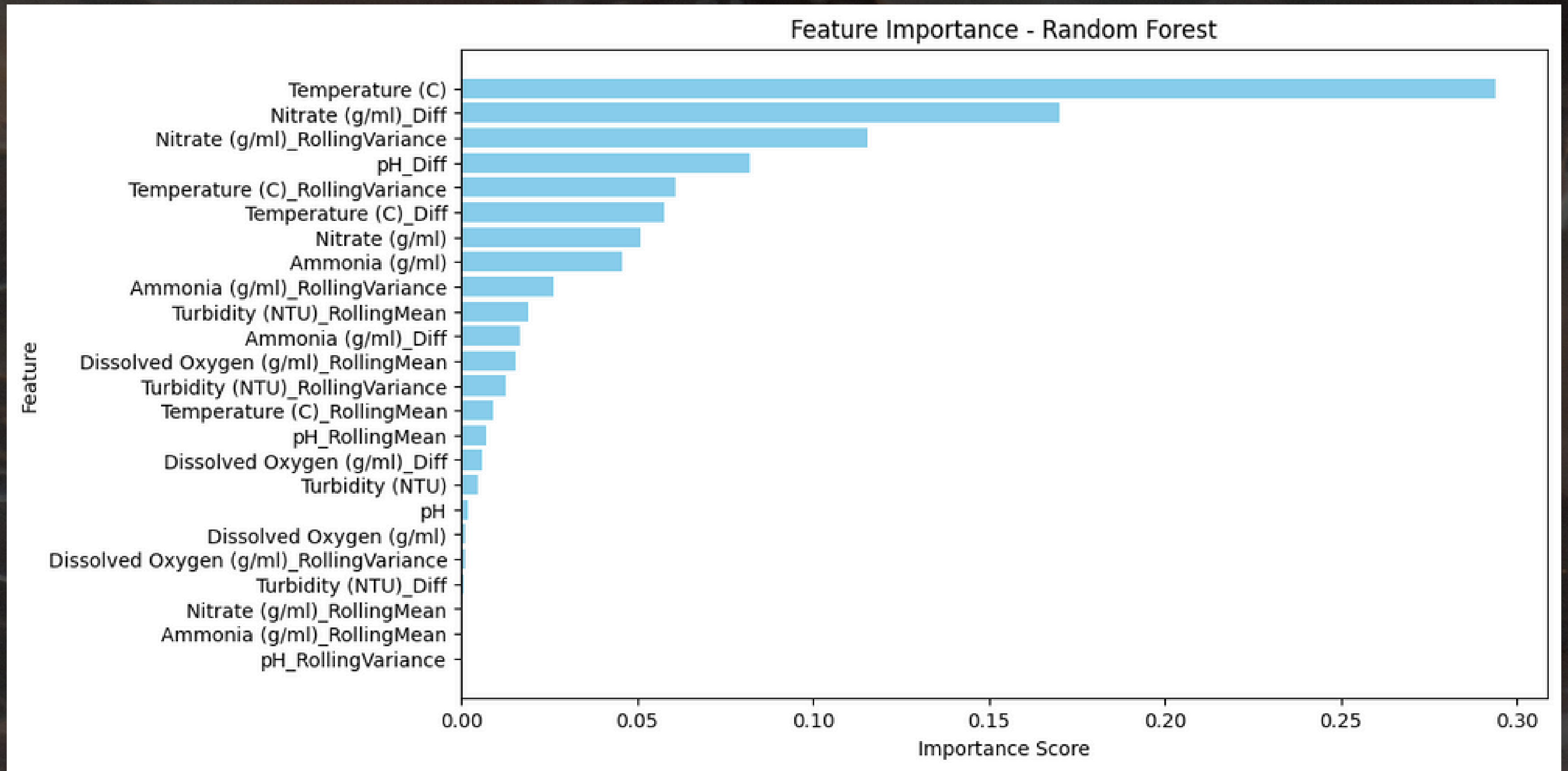
Random Forest outperformed XGBoost in direct train-test validation, achieving lower RMSE and MAE values. However, **XGBoost demonstrated better generalization in cross-validation**, indicating its potential for higher reliability in unseen data. Both models were effective, but Random Forest slightly edges out XGBoost for average weight prediction based on these results.

Methodology

Growth Rate Prediction

Comparative Feature Importance

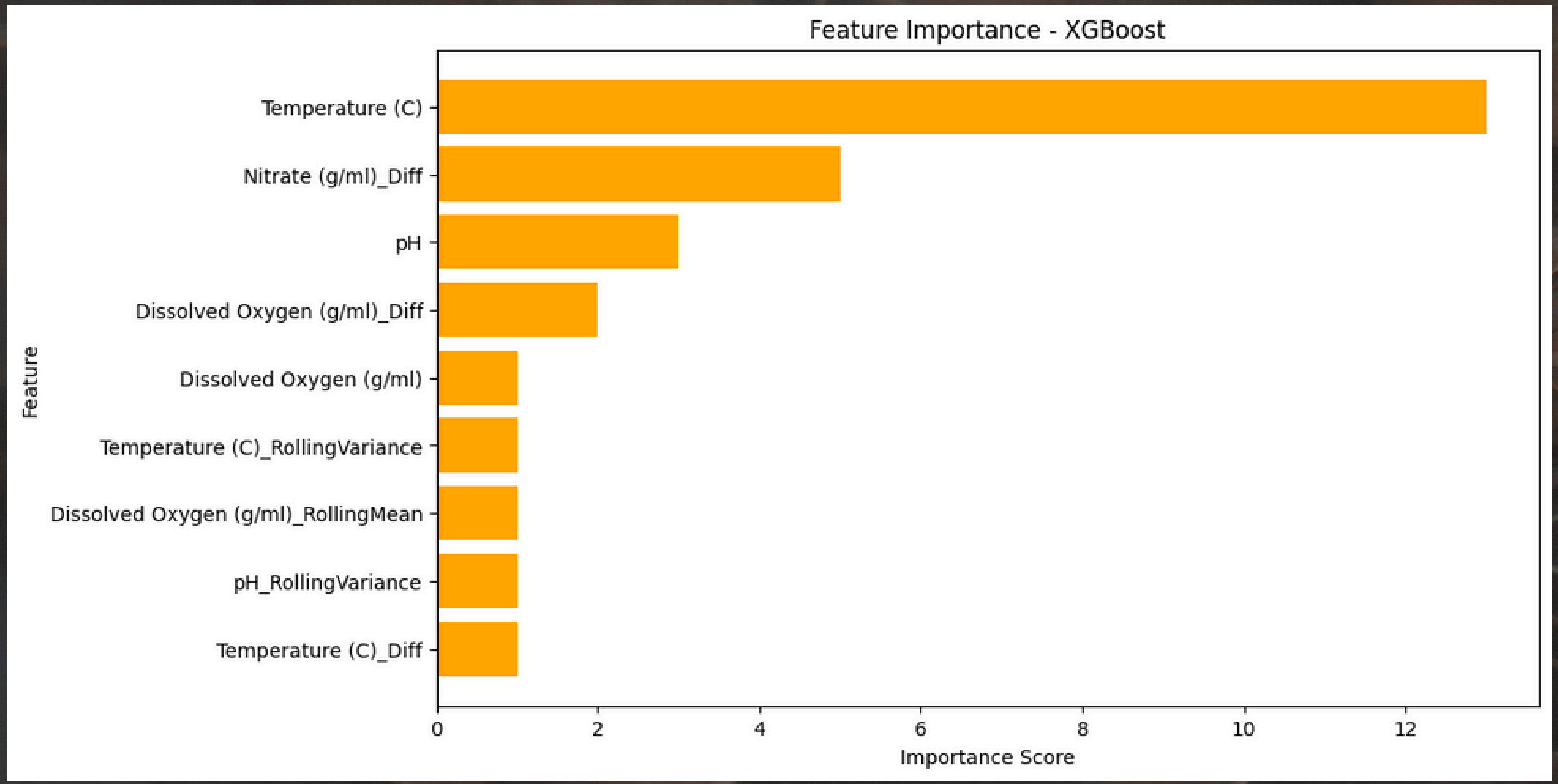




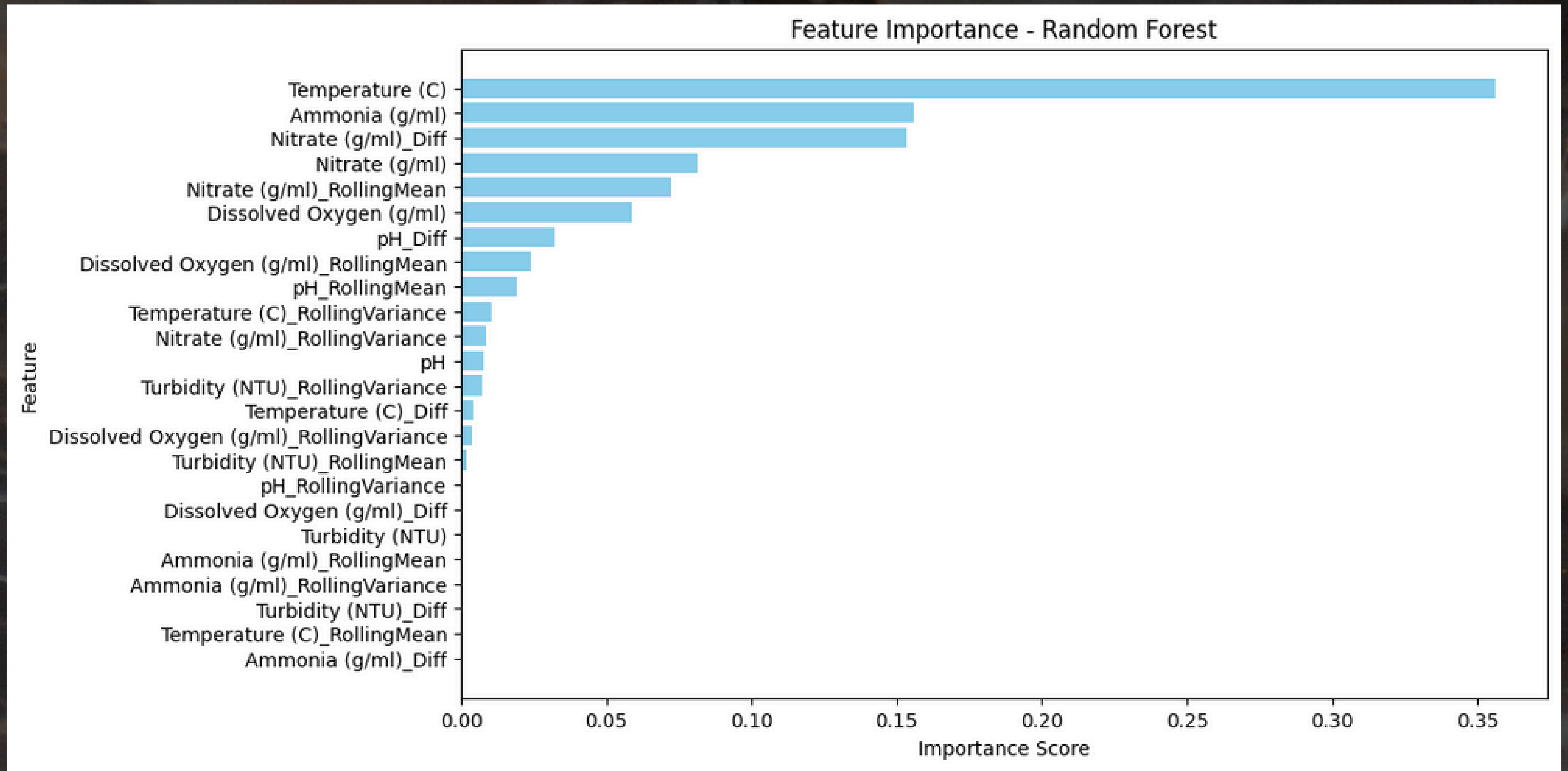
XGBoost Feature Importance

Random Forest Feature Importance

Results and Discussion
Average Weight Prediction
Comparative Feature Importance



XGBoost Feature Importance



Random Forest Feature Importance

Results and Discussion
Average Weight Prediction
Comparative Feature Importance

Growth Rate Prediction:

- **Top Features: Temperature, Nitrate Differences, Ammonia.**

Average Weight Prediction:

- **Top Features: Temperature Variance, Turbidity Trends.**

Conclusion

Key Findings

- **XGBoost is the most reliable for predicting growth rate due to accuracy and generalization.**
- **Random Forest performed best in average weight prediction, particularly in train-test evaluations.**
- **Engineered temporal features, such as nitrate rolling differences and temperature trends, significantly improved model performance, while static values like pH and turbidity contributed minimally.**

Recommendations

- **Expand datasets to include more ponds and diverse conditions.**
- **Incorporate real-time monitoring systems for continuous validation.**
- **Explore additional machine learning techniques for further improvement.**

Catfish Growth Rate and Average Weight Prediction Through Water Quality Parameters in Aquaponics System

Leveraging Data-Driven Approaches for Sustainable Aquaculture

John Kyle N. Junsay