Fish Growth Prediction, Optimal water Identification and Fish Growth and Water Quality Forecasting for Aquaponic Fish Ponds

Adam Osborne u21529583

ABSTRACT

Aquaponic Fish Ponds are a vital source of food and income for people. A disruption to yielding fish that have good size and length in these ponds due to bad water qualities will have dramatic consequences to the people that use the fish for food and the people that sell them as a source of income. Hence, this projects develops models that will predict the length and size of the fish in the ponds, identify which water qualities promote fish growth in both length and weight and forecast the water qualities, fish length and fish weight for the next month. This will allow the people managing these ponds to produce good size fish so that people will be able to have a good source of food and ensure that their source of food will not be disrupted. This is also good for business as the owner will be able to identify which water qualities will promote growth in the fish and therefore yield bigger fish that can be sold for more. They will also be able to plane their yields of fish using the forecaster.

1. INTRODUCTION

For many years, aquaponic fish ponds have been a source of income and food for many people. This serves as a sustainable way of farming fish that can be used as a source of food for many people. Many people have reaped the benefits of aquaponic fish ponds both as a source of food and as a source of income , but have also faced challenges with which these aquaponic fish ponds present. One of the major challenges is the providing good fish in terms of weight and length to people as a source of food, this is greatly due to the water qualities in these ponds in which the fish live in [4]. They also pose the challenge of knowing what the water qualities and fish length and size will be in the future.

With the advancement of technology, sensors were developed to take the readings of the different water quality parameters. The data these sensors capture can be used to solve these challenges presented by aquaponic fish ponds if used correctly. However, these readings are not being leveraged to their full capacity and leaves aquaponic fish ponds vulnerable to the challenges.

2. PROBLEM STATEMENT

Aquaponic fish ponds are a crucial source of income and food for many people. If a disruption occurs this could lead to food shortages and bankruptcy. There are not proper mechanisms in place to avoid these issues. Therefore, the problem this project addresses is that due to the absence of prediction and forecasting models that leverage the data captured by the sensors, it leaves the aquaponic fish ponds vulnerable to a food shortage or the business failing. The list below outlines the sub problems that this project addresses:

- One sub problem of this project is the lack of prediction models that will predict the fish weight and length based off the water quality. Aquaponic fish ponds owners do not have the correct models in place to be able to ensure that their fish grow to suitable lengths and weight to be used as a source of food.
- Another sub problem is being able to identify the optimal water qualities that promote the most fish growth in length and weight. Being able to identify the optimal water qualities is crucial as you can then modify you current water quality values to match the optimal to ensure the fish have optimal growth.
- The last sub problem is being able to predict the water quality values, fish length and fish weight. This is crucial as aquaponic fish ponds owners are able to then see if there fish will be big enough and if their water qualities will be good next month so immediate action can be taken.

3. METHODOLOGY

The problem this project addresses is the lack of effective models for actively monitoring and predicting water quality, along with fish length and weight, to ensure optimal yields of sizable fish from ponds to avoid food shortages and potential business failure.

To tackle this challenge, a literature review will be employed as part of the research methodology. This review will enrich my project and help shape the model. It will provide insights into previous research and highlight essential concepts in this field, such as related models or algorithms that could enhance my approach to addressing this issue. Additionally, a model will be developed based on findings from the literature review, integrating identified key concepts and adapting existing models where appropriate.

4. BACKGROUND

There is a lot of research that has been done on using the water qualities for different areas, not a lot of research has

been done on using historical water qualities to predict and forecast the water qualities, fish length and weight.
The paper by *Md. Monirul Islam et al.* [2] provides a model that uses water quality variables captured by sensors to classify what that pond will be suitable for cultivating fish based on the water quality variables. Although it is a very good to know if a pond is suitable for fish farming based off the water quality it does not predict or forecast the growth of the fish based on the water quality.
The paper by *Budy et al.* [1] performed research to see the affects of thermal habitat suitability and fish performance. They interfaced bioenergetics, field measurements, and Thermal Remote Imaging to measure fish growth potential. Although temperature is a water quality that does need to be measured its not the only measurement that plays a role in fish growth [4]. This leaves a gap in the work for using all the water quality variables available to be able to accurately predict fish length and weight as-well as the water quality variables.
The paper by *Parker et al.* [3] focused on the Illinois River, examining how improvements in water quality have impacted fish assemblages over time. They employed techniques like Random Forests and multiple linear regressions to link fish metrics with water quality and weather variables. Understanding the abundance and biomass of predatory and native fish, along with species richness, is certainly valuable. However, without tracking how fish growth is influenced by water quality, a significant gap remains in their research. They need to utilize the available water quality variables to make accurate predictions regarding fish length and weight, as well as the associated water quality factors.

5. EXPLORATORY DATA ANALYSIS (EDA)

EDA was performed on one of the Sensor Based Aquaponics Fish Pond Datasets that will have the data mining methods and analysis performed on it. The first few rows are printed to see how the dataset is structured and to give us more information of the features. This also ensures that the dataset was read in correctly. in table 1 and 2 you can see we have 11 features:

- created at
- entry id
- Temperature (C)
- Turbidity (NTU)
- Dissolved Oxygen (g/ml)
- PH
- Ammonia (g/ml)
- Nitrate (g/ml)
- Population
- Fish Length (cm)
- Fish Weight (g)

An initial overview of the dataset is performed to give us information about the features such as the count of the values, data types and missing values of each of the features. These are shown in the tables 1 and 2. As we can see there are missing values in the features. A statistical analysis is also performed on the features of the dataset and can be found in table 3.
Histograms were then created for each feature to see the distribution of values for each feature, this can be found in figure 5. A box plot was created to see if there are any outliers in any of the features, this can be found in figure 1. A pair-plot was then created to analyze the relationships between the features, this can found in figure 4. A correlation heat-map as also created to show the correlation between the features, this can be found in figure 2.
After analyzing all the figures, the water there is some correlation between the water quality and the fish length and weight. The trends shows that as the water quality features change so does the fish weight and length. There is also quite a big distribution of the water quality values.

6. DATA PREPROCESSING

After doing the EDA, there are missing values in the features and any data mining methods to the dataset if there are missing values. The rows with missing values were dropped. After inspecting the features the created at feature contains dates but is an object data type. Therefore, created at is transformed to a datetime datatype. The year and month is also extracted from the created at feature to make using the feature easier.
The dataset is then standardized so that the features are on all the same scale. This will improve the performance of machine learning algorithms since they all will be equal and wont have a just few features dominating when training. PCA is also applied to the dataset so that the dimensionality of the dataset is reduced but retaining as much information as possible. The variance of the PCA components can be found in figure 3.

7. DATA MINING METHODS AND ANALYSIS

I applied 3 different data mining methods and analysis for the Sensor Based Aquaponics Fish Pond Dataset. The three different methods are a fish length and weight prediction model based off the water quality of the pond, an optimal water quality analyzer and a water quality, fish length and weight prediction model. The three methods are discussed below.

7.1 Fish length and weight prediction based off the water qualities

This data mining method makes use of a Random Forest Regression algorithm to predict the fish length and weight based off the different water quality features in the dataset. The water quality features used are: Temperature (C), Turbidity (NTU), Dissolved Oxygen (g/ml), PH, Ammonia (g/ml) and Nitrate (g/ml).
Two separate models were made fro the fish weight and length. One model predicted the fish length off the water quality features and the other predicts the fish weight. Both Random Forest Regression algorithms use 100 estimators with a random state of 42. It used a test size of 0.2.
The Random Forest Regression algorithm for predicting length of the fish performed exceptionally and had a Mean squared

error of 0.02 and a R^2 of 1.00. You are able to find the plot of the prediction in figure 6. As you can see its very accurate in predicting fish length based off the water qualities.

The Random Forest Regression algorithm for predicting weight of the fish performed well and had a Mean squared error of 13.96 and a R^2 of 1.00. You are able to find the plot of the prediction in figure 7. As you can see its very accurate in predicting fish weight based off the water qualities.

7.2 Optimal Water Qualities

This method doesn't make use of any machine learning algorithm. However, it is really important as it shows what water qualities promoted the best growth in length and weight of the fish over a month. It works by subtracting the growth from the end of the month from the start of the month and therefore we can see which month had the biggest fish growth by comparing the differences between each month. We are then able to average the water quality values for that month and return them. The optimal water quality variables fr growth of weight and length of the fish can be found in table 4.

7.3 Water quality, Fish length and weight forecaster

This data mining method makes use of a Long Short-Term Memory (LSTM) algorithm. It is able to identify patterns in the data which makes it perfect for projecting the water qualities, fish length and weight. The model trains itself on the following features: Temperature (C), Turbidity (NTU), Dissolved Oxygen (g/ml), PH, Ammonia (g/ml), Nitrate (g/ml), Fish Length (cm) and Fish Weight (g).

The model looks back at the days of the previous month and sequences the data for it to be able to identify patterns between each day. It then splits the data into the test and train sets with a test size of 0.2. The model is then run for 50 epochs and is then evaluated using Mean Squared Error. It achieved a Mean Square Error of 0.07. The forecast is then plotted for each feature for the next month, you will find the predictions in figure 8 .

8. CONCLUSION AND REFLECTION

This project addresses the problem of aquaponic fish ponds being vulnerable to food shortages and potential business failure. This project solves each sub problem that was presented in Section 2. The processes on how the problem was solved and extent to which each sub problem is solved is explained below.

The problem was solved by applying EDA, data preprocessing and data mining methods to the Aquaponics Fish Pond Dataset predict the fishes length and weight based on the water qualities of the pond, identify the water qualities that promoted the best growth for fish weight and length and forecast the water qualities of the pond, fish length and weight for the next month. This involved cleaning up the dataset of missing values, scaling and reducing the dimensionality of the dataset using PCA, training the models on historical trends and leveraging this to predict and forecast. Reflecting on this project, it is clear that good EDA and data preprocessing is essential in order to be able to use the data effectively in data mining methods and analysis. Due to the extensive EDA and data preprocessing my models using the data were able to be efficient and accurate in their predictions and forecasting. Fish farmers are now able to avoid food shortages and potential business failure by predicting fish length and weight from the water qualities, obtain the optimal water qualities that promote fish growth and forecasting the water quality values, fish length and weight.

9. REFERENCES

[1] P. Budy, M. Baker, and S. K. Dahle. Predicting fish growth potential and identifying water quality constraints: A spatially-explicit bioenergetics approach. *Environmental Management*, 2011.

[2] M. M. Islam, M. A. Kashem, S. A. Alyami, and M. A. Moni. Monitoring water quality metrics of ponds with iot sensors and machine learning to predict fish species survival. *Microprocessors and Microsystems*, 102:104930, 2023.

[3] J. Parker, J. Epifanio, A. Casper, and Y. Cao. The effects of improved water quality on fish assemblages in a heavily modified large river system. *River Research and Applications*, 32(5):992–1007, 2016.

[4] M. M. Rahman, L. A. Nagelkerke, M. C. Verdegem, M. A. Wahab, and J. A. Verreth. Relationships among water quality, food resources, fish diet and fish growth in polyculture ponds: A multivariate approach. *Aquaculture*, 275(1):108–115, 2008.

APPENDIX

Column	Non-Null Count	Dtype
created_at	172249 non-null	object
entry_id	172249 non-null	int64
Temperature (C)	172249 non-null	float64
Turbidity (NTU)	172249 non-null	int64
Dissolved Oxygen (g/ml)	172249 non-null	float64
PH	172249 non-null	float64
Ammonia (g/ml)	172159 non-null	float64
Nitrate (g/ml)	172249 non-null	int64
Population	172249 non-null	int64
Fish_Length (cm)	172249 non-null	float64
Fish_Weight (g)	172249 non-null	float64

Table 1: Feature Descriptions

Column	Missing Values
created_at	0
entry_id	0
Temperature (C)	0
Turbidity (NTU)	0
Dissolved Oxygen (g/ml)	0
PH	0
Ammonia (g/ml)	90
Nitrate (g/ml)	0
Population	0
Fish_Length (cm)	0
Fish_Weight (g)	0

Table 2: Missing Values

Column	mean	min	25%	50%	75%
created_at	NaN	NaN	NaN	NaN	NaN
entry_id	148725.739702	1889.000000	62717.000000	147763.000000	226310.000000
Temperature (C)	24.982849	-127.000000	24.375000	24.937500	25.500000
Turbidity (NTU)	90.974665	1.000000	94.000000	100.000000	100.000000
Dissolved Oxygen (g/ml)	9.708503	0.007000	3.200000	3.283000	11.739000
PH	3.971857	-3.137450	-0.173180	7.099040	7.516670
Ammonia (g/ml)	3.112110e+08	6.590000e-03	5.693500e-01	8.470560e+00	8.070516e+01
Nitrate (g/ml)	719.891448	45.000000	189.000000	890.000000	1050.000000
Population	50.0	50.0	50.0	50.0	50.0
Fish_Length (cm)	23.428782	6.960000	14.220000	20.970000	32.540000
Fish_Weight (g)	166.470532	3.360000	22.890000	65.480000	302.500000

Table 3: Descriptive Statistics for Each Feature

Condition	Optimal for Highest Weight Growth	Optimal for Highest Length Growth
Temperature (C)	24.809527	24.809527
Turbidity (NTU)	100.000000	100.000000
Dissolved Oxygen (g/ml)	3.158294	3.158294
PH	-0.036547	-0.036547
Ammonia (g/ml)	9.071701	9.071701
Nitrate (g/ml)	1043.939128	1043.939128

Table 4: Optimal Conditions for Highest Weight and Length Growth

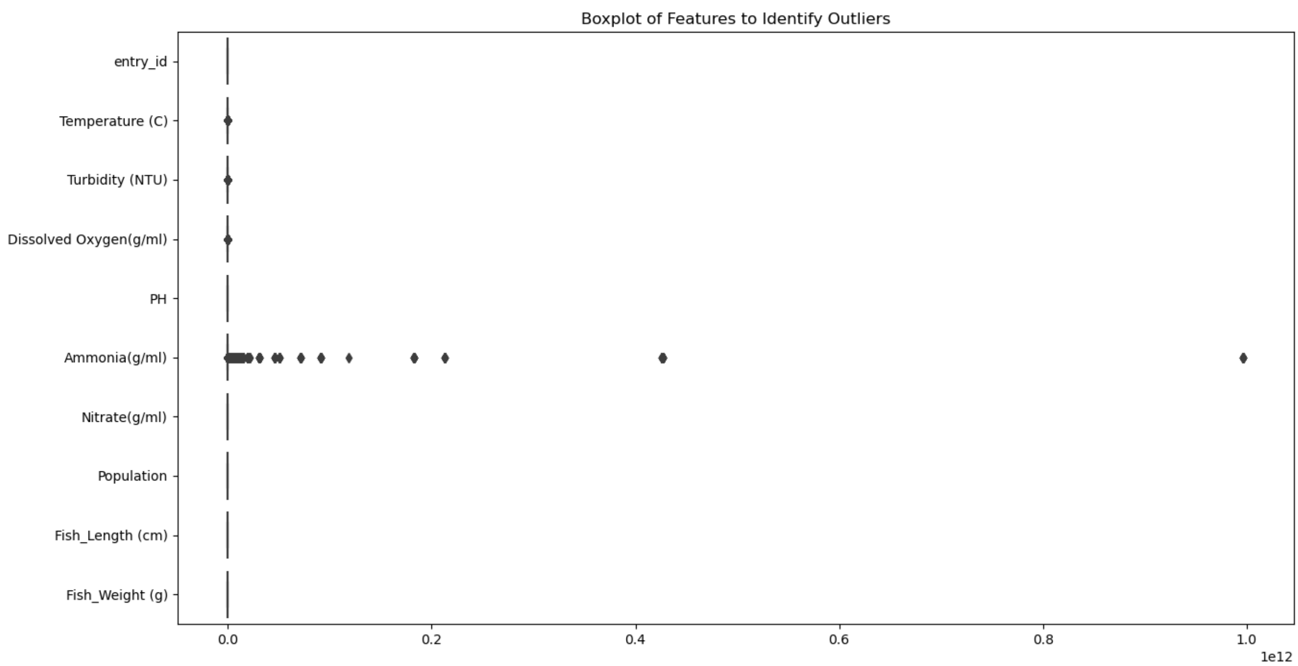

Figure 1: Box Plot of the Features

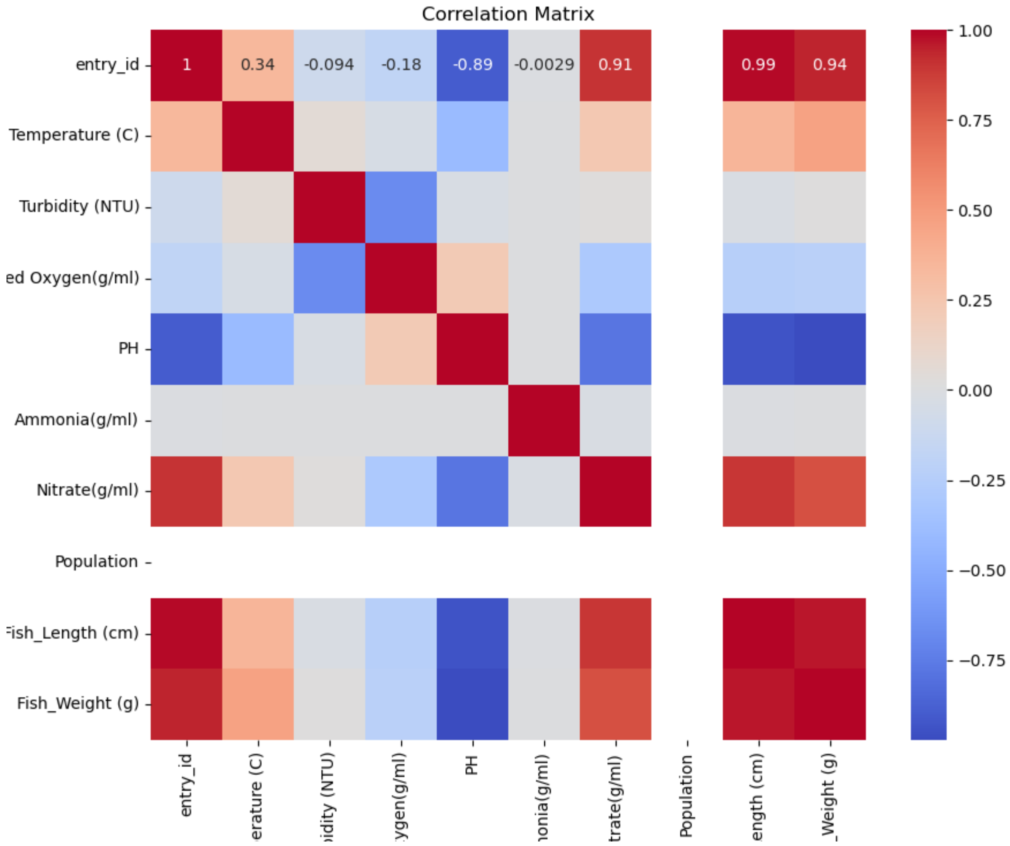

Figure 2: Correlation matrix of the Features

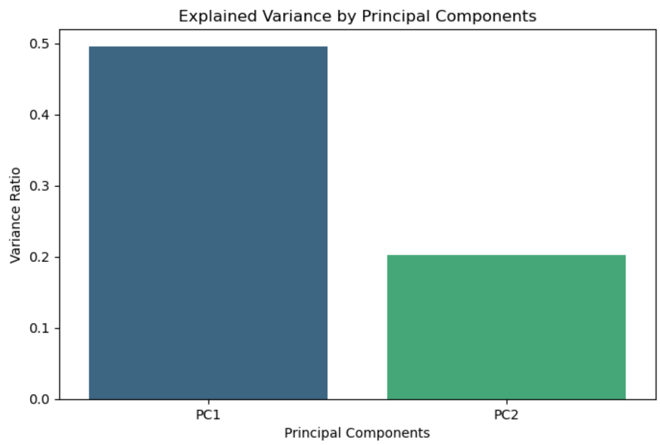

Figure 3: PCA

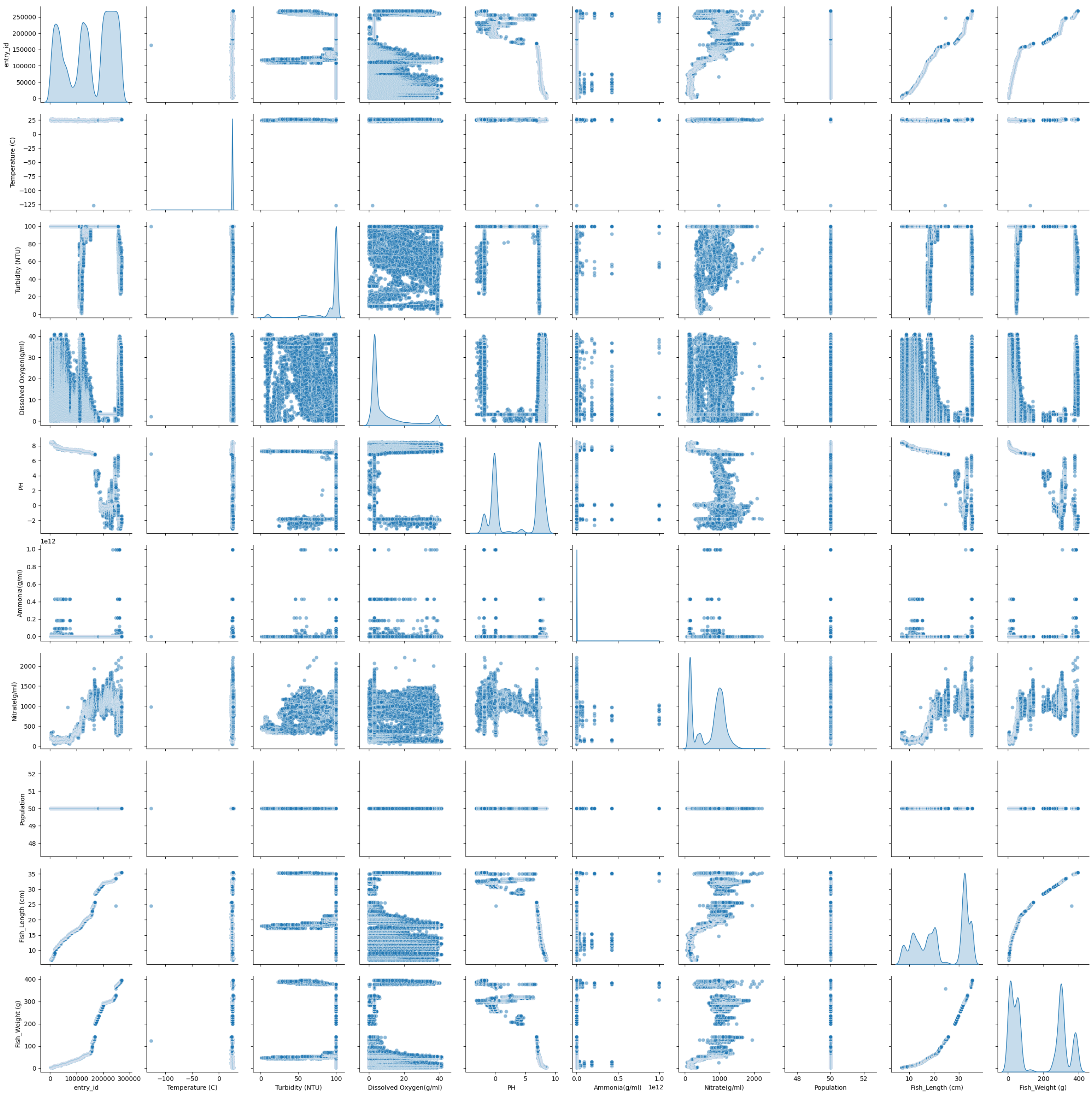

Figure 4: Pair Plot of the features

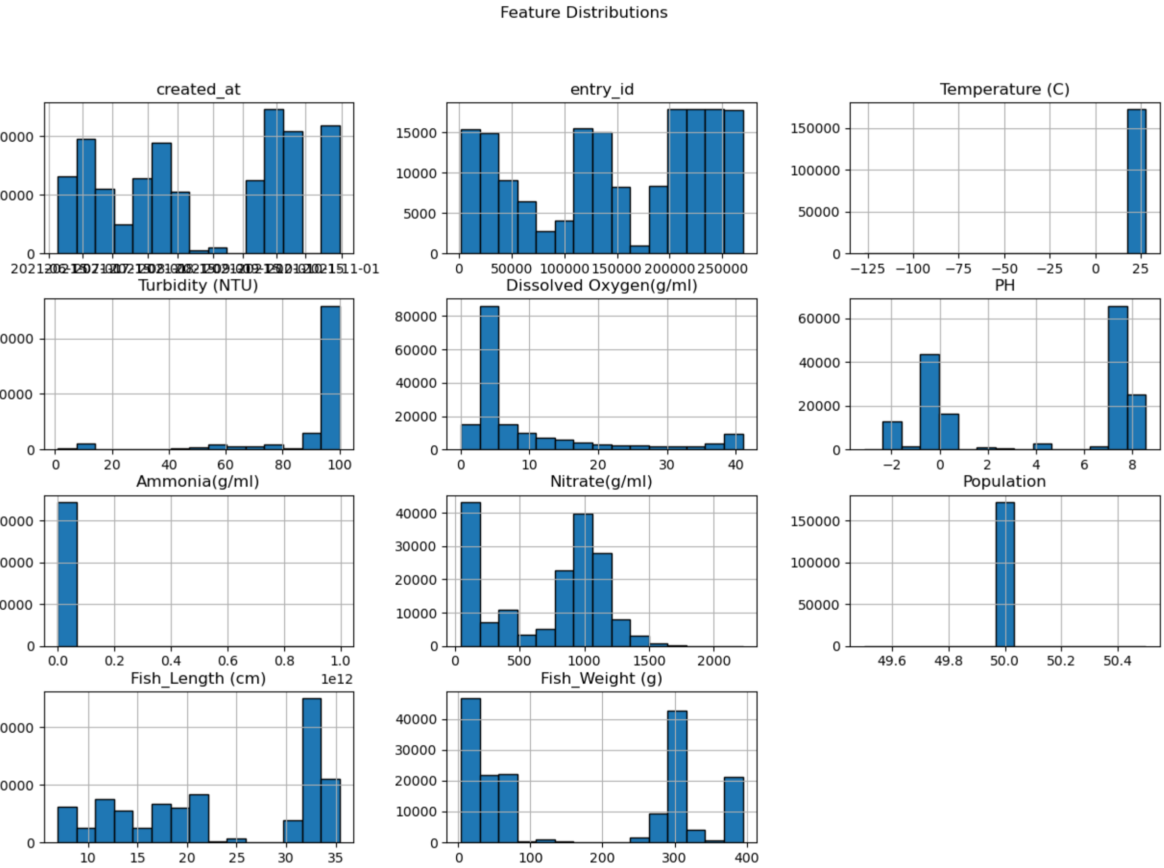

Figure 5: Histogram of the features

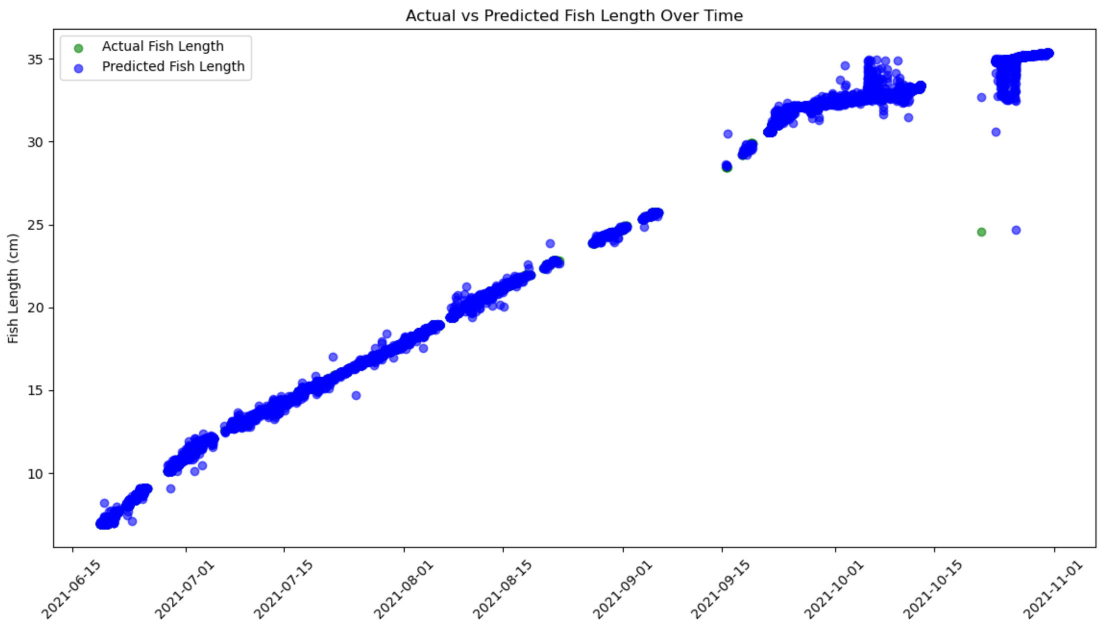

Figure 6: Fish Length Prediction

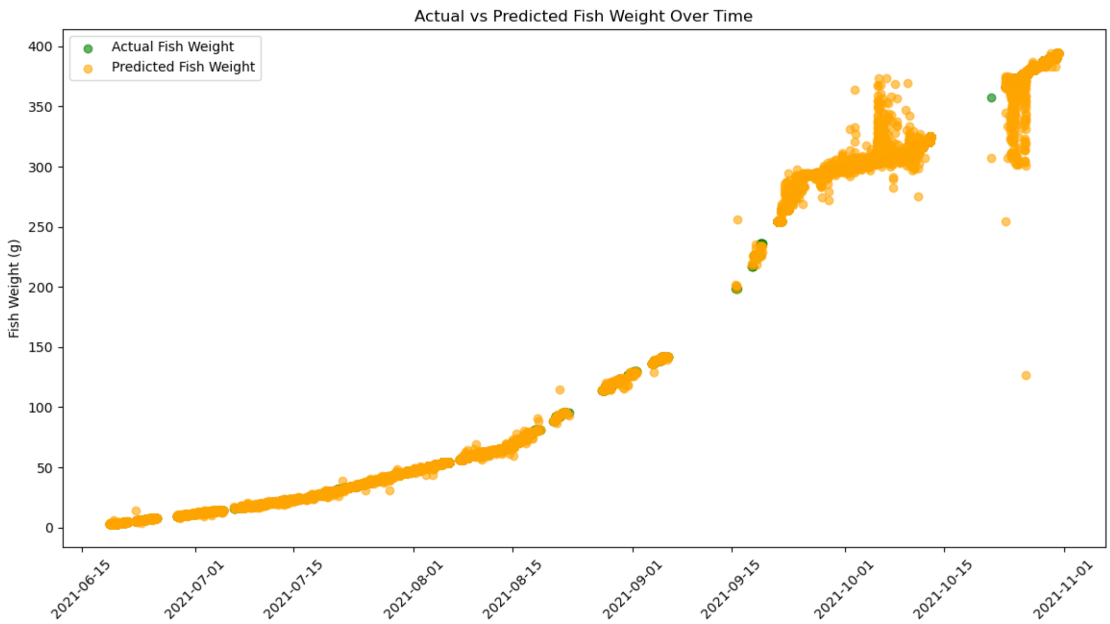

Figure 7: Fish weight Prediction

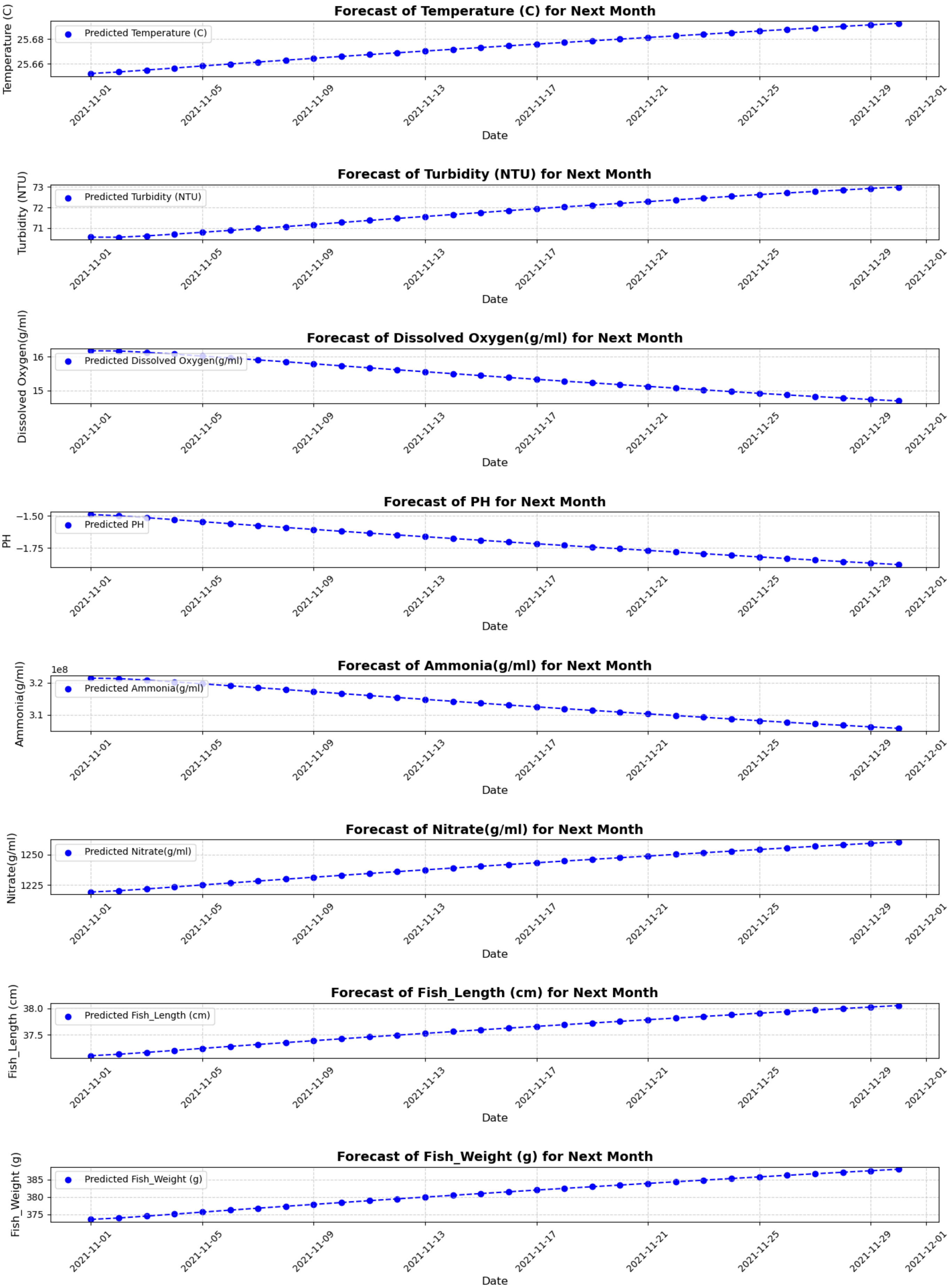

Figure 8: Forecast of the next month