Faculty of Electronic and Computer Technology and Engineering

PREDICTION MODEL FOR FISH GROWTH RATE USING MACHINE LEARNING IN AQUACULTURE TECHNOLOGY

NURUL ALYAA IZZATI BINTI MEOR KHAZAMUDDIN

Bachelor of Computer Engineering Technology (Computer Systems) with Honours

2024

PREDICTION MODEL FOR FISH GROWTH RATE USING MACHINE LEARNING IN AQUACULTURE TECHNOLOGY

NURUL ALYAA IZZATI BINTI MEOR KHAZAMUDDIN

A project report submitted
in partial fulfillment of the requirements for the degree of
Bachelor of Computer Engineering Technology (Computer System) with Honours

Faculty of Electronic and Computer Technology and Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2024

UNIVERSITI TEKNIKAL MALAYSIA MELAKA
FAKULTI TEKNOLOGI DAN KEJUTERAAN ELEKTRONIK DAN KOMPUTER

BORANG PENGESAHAN STATUS LAPORAN
PROJEK SARJANA MUDA II

Tajuk Projek : Prediction Model For Fish Growth Rate Using Machine Learning In Aquaculture

Sesi Pengajian : 2023/2024

Saya Nurul Alyaa Izzati Binti Meor Khazamuddin Mengaku membenarkan laporan Projek Sarjana Muda ini disimpan di Perpustakaan dengan syarat-syarat kegunaan seperti berikut:

1. Laporan adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan laporan ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (✓):

☐	**SULIT***	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
☑	**TERHAD***	(Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
☐	**TIDAK TERHAD**	

Disahkan oleh:

alyme

(TANDATANGAN PENULIS)

(COP DAN TANDATANGAN PENYELIA)

Alamat Tetap:

No 54, Jalan Pelangi 3,
Taman Pelangi Fasa 2,
36700 Langkap,
Perak.

TS. NADZRIE BIN MOHAMOOD
Jurutera Pengajar
Fakulti Teknologi dan Kejuruteraan Elektronik dan Komputer (FTKEK)
Universiti Teknikal Malaysia Melaka (UTeM)

Tarikh: 12.02.2024

Tarikh: 12.02.2024

*CATATAN: Jika laporan ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali tempoh laporan ini perlu dikelaskan sebagai SULIT atau TERHAD.

DECLARATION

I declare that this project report entitled Prediction Model for Fish Growth Rate Using Machine Learning in Aquaculture Technology is the result of my own research except as cited in the references. The project report has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature :

Student Name : Nurul Alyaa Izzati Binti Meor Khazamuddin

Date : 12.02.2024

APPROVAL

I hereby declare that I have checked this project report and, in my opinion, this project report is adequate in terms of scope and quality for the award of the degree Bachelor of Computer Engineering Technology (Computer Systems) with Honours.

Signature	:	
Supervisor Name	:	Ts. Nadzrie Bin Mohamood
Date	:	12.02.2024
Signature	:	
Co-Supervisor Name (if any)	:	
Date	:	

DEDICATION

I dedicate this project to my esteemed parents, whose unwavering support and encouragement have been a constant presence throughout this journey. Their guidance, fortitude, and provision of moral, emotional, and financial support have been instrumental.

I extend my sincere gratitude to my Bachelor's Degree Project supervisor, Ts. Nadzrie Bin Mahamood, for his invaluable guidance and expertise, which significantly contributed to the success of this project. His patience and words of encouragement throughout this endeavour have served as an incredible source of inspiration.

I wish to express my heartfelt thanks to all my friends who played crucial roles as supporters, sources of strength, advisers, and motivators during the course of the Bachelor's Degree Project. Your moral support and guidance throughout this semester at Universiti Teknikal Malaysia Melaka are deeply appreciated.

ABSTRACT

Grouper fish farming is gaining popularity among entrepreneurs in Malaysia due to high demand and income return. However, in general this sector requires constant effort and commitment as well as knowledge complemented by skills. Current practice in fish farming still rely on traditional or manual methods to estimate the physical dimensional measurement and the weight of fish to forecast those fish growth rates. However, the concern arises here where manual methods could affect the health of the fish itself by applying unnecessary pressure and injury threat every single time when manual measurement took place. With that, developing new technology-based methods to improve fish performance and growth rate is a priority. Based on the problems stated above, the main objective of this project is to develop a prediction model for fish growth rate realizing with machine learning regression method, thus enabling fish farming to be carried out more accurately and efficiently. This project also determines the prediction accuracy across several potential machine learning algorithms, which are selected from conducted literature on past research studies. The proposed algorithms for prediction model will be trained and tested using selected datasets. In general, realization of entire project uses PyCharm as an IDE platform with Python programming language, including Graphical User Interface (GUI) design implementation. Random Forest Regression model outperformed the Linear Regression and Polynomial Regression models in terms of accuracy in predicting fish weight, according to comprehensive analysis of the project. The Random Forest Regression was determined to be the most predictive capability that achieve a high accuracy. This model shown the lowest value of MSE, MAE, and RMSE which has been determined to be 6540986.24, 1468.15 and 2557.54 respectively. In conclusion, machine learning is very important to predict accuracy of fish weight which are cost effective, precise and less strain on the fish that could harm it.

ABSTRAK

Penternakan ikan kerapu semakin mendapat sambutan di kalangan pengusaha di Malaysia berikutan permintaan yang tinggi dan pulangan pendapatan. Walau bagaimanapun, secara umumnya sektor ini memerlukan usaha dan komitmen yang berterusan di samping pengetahuan dan juga kemahiran. Amalan semasa dalam penternakan ikan masih bergantung kepada kaedah tradisional atau manual untuk menganggar ukuran dimensi secara fizikal dan juga berat ikan dalam menentukan kadar pertumbuhan ikan tersebut. Walau bagaimanapun, kebimbangan timbul di sini di mana kaedah manual boleh menjejaskan kesihatan ikan itu sendiri dengan mewujudkan tekanan yang tidak perlu terhadap ikan selain ancaman kecederaan setiap kali ketika pengukuran manual dilakukan. Dengan itu, kaedah penyelesaian berasaskan teknologi baharu untuk meningkatkan prestasi dan kadar pertumbuhan ikan adalah menjadi keutamaan. Berdasarkan masalah yang dinyatakan di atas, objektif utama projek ini adalah untuk membangunkan sebuah model bagi meramalkan kadar pertumbuhan ikan dengan menggunakan kaedah regresi pembelajaran mesin, sekali gus membolehkan penternakan ikan dijalankan dengan lebih tepat dan cekap. Projek ini juga bertujuan untuk menentukan ketepatan ramalan melalui beberapa algoritma pembelajaran mesin yang berpontensi, yang mana telahpun dipilih melalui kajian kepustakaan kajian penyelidikan lepas. Algoritma yang dicadangkan untuk ramalan akan dilatih dan diuji menggunakan set data terpilih. Secara umum, keseluruhan projek direalisasikan menggunakan *PyCharm* sebagai platform *IDE* dengan mengaplikasikan bahasa pengaturcaraan *Python*, termasuk pelaksanaan reka bentuk Antaramuka Pengguna Grafik (GUI). Model Regresi Hutan Rawak mengatasi model Regrasi Linear dan Regresi Polinomial dari segi ketepatan dalam meramal berat ikan, menurut analisis komprehensif projek. Regrasi Hutan Rawak ditentukan sebagai keupayaan ramalan yang mencapai ketepatan yang tinggi. Model ini menunjukkan nilai terendah MSE, MAE, dan RMSE yang telah ditentukan masin-masing ialah 6540986.24, 1468.15 dan 2557.54. Kesimpulanya pembelajaran mesin adalah sangt penting untuk meralmalkan ketepatan berat ikan yang kos efektif, tepat dan kurang ketegangan pada ikan yang boleh membahayakanya.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my supervisor, Ts. Nadzrie Bin Mohamood for his precious guidance, words of wisdom and patient throughout this project.

I am also indebted to Universiti Teknikal Malaysia Melaka (UTeM) and Faculty Technology and Engineering and Computers for the financial support which enables me to accomplish the project. Not forgetting my colleagues for the willingness of sharing his thoughts and ideas regarding the project. My highest appreciation goes to my parents, and family members for their love and prayer during the period of my study.

Finally, I would like to thank all the staffs at the Universiti Teknikal Malaysia Melaka (UTeM) and classmates, the faculty members, as well as other individuals who are not listed here for being co-operative and helpful.

TABLE OF CONTENTS

		PAGE
DECLARATION		
APPROVAL		
DEDICATIONS		
ABSTRACT		**i**
ABSTRAK		**ii**
ACKNOWLEDGEMENTS		**iii**
TABLE OF CONTENTS		**iv**
LIST OF TABLES		**vi**
LIST OF FIGURES		**vii**
CHAPTER 1	**INTRODUCTION**	**1**
1.1	Background	1
1.2	Problem Statement	2
1.3	Project Objective	3
1.4	Scope of Project	3
CHAPTER 2	**LITERATURE REVIEW**	**4**
2.1	Introduction	4
2.2	Overview of Machine Learning	4
	2.2.1 Supervised Learning Algorithm	5
	2.2.2 Unsupervised Learning Algorithm	6
2.3	Prediction Models	8
	2.3.1 Machine Learning Classification	8
	2.3.2 Machine Learning Regression	8
2.4	Machine Learning in Aquaculture Application	12
2.5	Data Collection and Preprocessing	13
	2.5.1 Data Collection for Fish Growth Methods	14
	2.5.2 Data preprocessing and features for Fish Growth Methods	14
2.6	Element of Accuracy for Fish Growth Rate in Aquaculture	15
	2.6.1 Feeding Controlling	15
	2.6.2 Biomass Estimation	16
	2.6.3 Water Quality	17
2.7	Dataset in Machine Learning Aquaculture	17
2.8	Python in Artificial Intelligent	19
2.9	Figure of Machine Learning	20
2.10	Table of Related Work	21
2.11	Summary	31

CHAPTER 3 METHODOLOGY 32
3.1 Introduction 32
3.2 Machine Learning Model 32
3.2.1 Data Collection 32
3.2.2 Data Preparation 33
3.2.3 Data Splitting 33
3.2.4 Model Development 34
3.2.5 Model Performance Evaluation 34
3.3 Software Architectural Design 35
3.4 PyCharm Integrated Development Environment (IDE) 39
3.4.1 Python 40
3.4.2 Python GUI development using Python 3 41
3.5 Software Flow Chart 42
3.6 Equipment 46
3.7 Summary 47

CHAPTER 4 RESULTS AND DISCUSSIONS 48
4.1 Introduction 48
4.2 Result and Analysis 48
4.3 Datasets Setup and Preparation 49
4.4 Regression Model Evaluation and Selection 50
4.5 Performance Model Machine Learning Evaluation 51
4.5.1 Correlation Result 52
4.5.2 Model Performance Result 53
4.5.3 Regression Analysis and Model Evaluation 54
4.6 Graphical User Interface 56
4.7 Summary 60

CHAPTER 5 CONCLUSION AND RECOMMENDATIONS 61
5.1 Introduction 61
5.2 Conclusion 61
5.3 Potential for Commercialization 62
5.4 Future Works 62

REFERENCES 64

LIST OF TABLES

TABLE	TITLE	PAGE
Table 2.1 Summary of Related Work		21
Table 3.1 Software Application		46
Table 4.1 Correlation Result Against all Biometric Measurement’s variables.		50
Table 4.2 Regression Score of Model Evaluation		51
Table 4.3 Correlation Between Variables		53
Table 4.4 Performance Score for Machine Learning Model		54

LIST OF FIGURES

FIGURE	TITLE	PAGE
Figure 2.1 Type of Supervised Learning		6
Figure 2.2 Type of Unsupervised Learning		7
Figure 2.3 Linear Regression		10
Figure 2.4 Logistic Regression		10
Figure 2.5 Polynomial Regression		11
Figure 2.6 Random Forest Regression		11
Figure 2.7 Process of Machine Learning		20
Figure 2.8 Dataset for Machine Learning		20
Figure 3.1 Software Architectural Design of Machine Learning Regression		36
Figure 3.2 Model Diagram of Machine Learning Regression		37
Figure 3.3 Model Diagram GUI		38
Figure 3.4 PyCharm Integrated Development Environment (IDE)		39
Figure 3.5 Python Programming Language		40
Figure 3.6 Tkinter of GUI		41
Figure 3.7 Flowchart of System Prediction Model for Fish Growth Rate		43
Figure 3.8 Flowchart of Prediction Weight in GUI		45
Figure 4.1 Total Count among Grouper's Species		49
Figure 4.2 Scatter Comparison of Actual vs. Predicted Weight		55
Figure 4.3 Main Page of GUI		57
Figure 4.4 Fish Weight Prediction Page		58
Figure 4.5 Prediction Weight of the Fish		58
Figure 4.6 Fish Weight Prediction in Month Page		59

Figure 4.7 Prediction Weight in Months 60

CHAPTER 1

INTRODUCTION

1.1 Background

Fish farming is a task that requires continuous effort and commitment, and farmers need to be equipped with proper knowledge, skills, and most importantly patience. In addition, fish farms face challenges in knowing the maturity of fish for sale, increasing knowledge about the size of fish to classify according to grade, and unpredictable factors such as environmental factors. Significant developments in fish management, nutrition, culture methods, and genetics have improved fish performance and increased production worldwide. To increase fish performance and growth, farming needs to understand the feeding, environment, and interaction among genotypes [1]. By adapting technology to the aquaculture field, fish farms are becoming more intelligent and able to access a large amount of sensor data, which will help in cost reduction and further increase production capacity, thus maximizing income. Taking advantage of the advancement of Artificial Intelligence (AI) will catch the attention among researchers to understand on how far this technology could possibly drive and redesign the aquaculture industry. This project will focus on developing the prediction model for fish growth rate using machine learning in aquaculture technology.

1.2 Problem Statement

Nowadays, Grouper fish catch an attention among farmers in Malaysia due to high demands, faster harvesting, fewer defect, and easy to raise. The market price for grouper fish can reach up to RM 60.00 per kg. However, there are challenges among farmers in raising this fish species since it is greatly influenced by the quality of the seawater present in their habitat. The optimum saltwater quality needs to be maintained at the level of pH 7.5 to pH 8.5 with suggested salinity is at 30 to 33 ppt within 24°C to 31°C of water temperature [2]. On the other hand, the quantity, and the frequency of feeding are also crucial in producing good quality grouper fish. Overfeeding fish will cause water pollution, increased bacterial load, increased biological oxygen demand, and low dissolved oxygen levels. Current practice, farmers still rely on manual methods in periodically examining the fish biomass, to ensure their fish's growth rate is on track. To measure the physical dimension, fish need to be caught by fishing net thus will cause the unnecessary pressure to the fish itself and could potentially cause an injury. Looking into this matter, all necessary parameters for grouper fish growth rate estimation are currently determined based on observation and experiences.

However, conducted research has outlined three important elements that greatly influence the grouper fish growth rate; 1) water quality; 2) feeding; and 3) biomass estimation. Since the manual method could not produce a precise measurement towards all those parameters, this project will propose a technological solution in providing better accuracy in determining grouper fish growth rate. Thus, potentially can improve productivity and increase profitability and at the same time minimize human intervention in grouper fish farming.

1.3 Project Objective

The aim of this project is to design a project to control the workflow of the prediction model for fish growth rate using machine learning in aquaculture technology. Specifically, the objective are as follows:

a) To develop a prediction model for fish growth rate using machine learning regression method.

b) To determine the prediction accuracy of selected machine learning algorithm.

1.4 Scope of Project

One of the purposes of this project is to focus on fish farming areas that specialize in grouper fish. This grouper fish breeding area is in the coastal area. In addition, this grouper fish is raised from small seedlings to maturity. This project also determines the prediction accuracy across several potential machine learning regression algorithms. In providing the training inputs to the prediction model, this project will not apply the image processing techniques in extracting fish dimensions but will be replaced with the available datasets instead. As the outputs, the model will predict the weight based on the 1) measured fish dimension: width and length, 2) given time in the future.

This project focuses on the development of an algorithm for grouper fish growth rate predictions model. The study intends to improve the accuracy and efficiency of predicting the growth rates of grouper fish by identifying the best machine learning regression algorithms.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter explains the machine learning for fish growth rate overview regarding feeding, size or biomass estimation, and water quality that plays an important role which influences in fish growth rate. Furthermore, previous studies or research in smart aquaculture will also be stated. A review of the literature states a portion of activities related to decisions on support systems in aquaculture operation. This section will explore how the project to predict fish growth rates using machine learning can contribute to our understanding of analysis, including data, algorithms, and performance to reduce and facilitate fish farms. Later in this chapter, further investigation and discussion are conducted on the elements considered in the development of aquaculture

2.2 Overview of Machine Learning

Artificial intelligence (AI) and machine learning (ML) have recently received a lot of attention. However, machine learning is a subfield of artificial intelligence that can define as the process of developing algorithms and models that allow computers to collect data and make predictions which is commonly to solve problems and decisions based on observation. However, machine learning became its own domain in the 1980s. Machine learning began to develop in the 1990s, where most of AI inspired strategies and techniques derived from probability and statistics [2].

Nowadays, machine learning has become an established discipline in various ways. The methodology for applying ML involves the selection of relevant data and its pre-processing, the selection of appropriate algorithms, and the assessment of solution quality. Machine learning, capable of handling massive amounts of data, has the potential to outperform humans with higher accuracy [3]. By creating models with the of a training data set, the ML algorithm "learn." In general, there are two main categories of learning such as supervised learning and unsupervised learning.

The training data is fed into the chosen algorithm to start the machine learning process. The training data used to create the final machine learning algorithm may be known or unknown data. The method is influenced by the type of training data input, and that idea will be discussed in more detail shortly. The machine learning is given new input data to see if it works correctly. Then, predictions and results are cross-checked. Algorithms are repeatedly retrained until data achieves the desired result if the predictions and do not align. This is because machine learning can continuously train itself and generate the best solutions by improving their accuracy significantly[4].

2.2.1 Supervised Learning Algorithm

Supervised learning is the process of labeling a new observation from a test set using a dataset that has been taught using a training set to create machine learning. Regarding the training set, the attribute that will affect the accuracy of the predicted variable is the input variables. It consists of both qualitative variables, and the label class that supervised learning will use to categorize new observations in is the output variable. Supervised learning can be divided into two types of issues such as regression and classification. Depending on the number of output variables. Both the output variables for the classification and regression are categorical variables. It is also referred to [5] as the purpose of building an artificial system that can learn the relationship between the input and output of the system given new input is the goal of supervised learning. A learned mapping produces a categorization of the input data if the output accepts a limited number of discrete values representing the input class labels. If the output has a constant value, the input will be regressed. In Figure 2.1, supervised learning is depicted and can be further classified into different types.

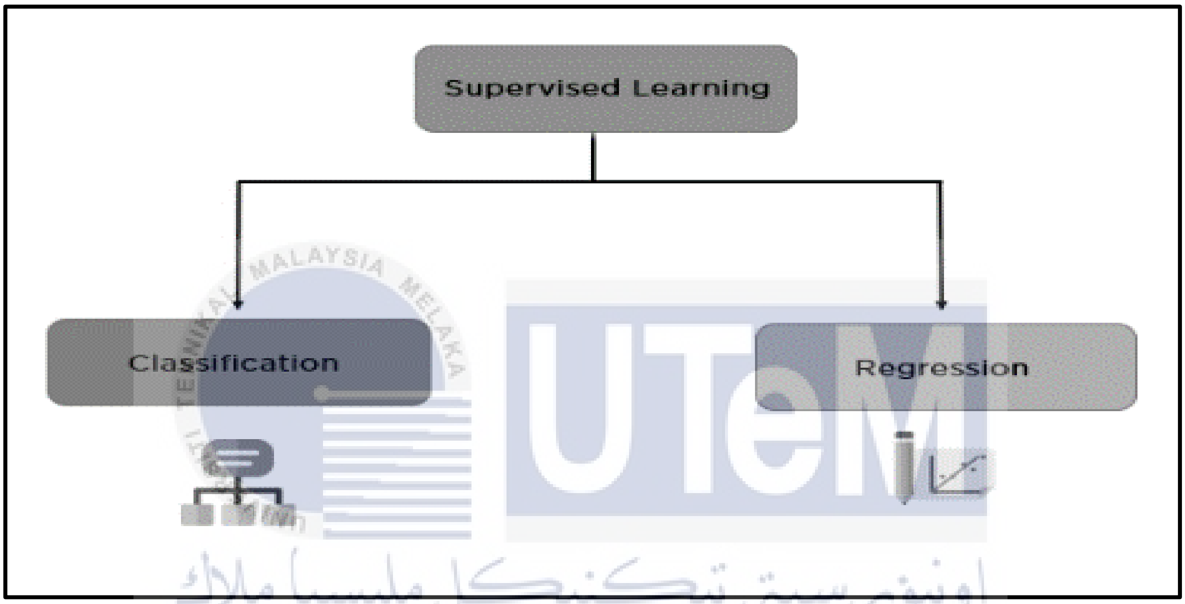


Figure 2.1 Type of Supervised Learning

Source: https://www.simplilearn.com/tutorials/machine-learning-tutorial/supervised-machine-learning

2.2.2 Unsupervised Learning Algorithm

Unsupervised learning is a term used to describe an algorithm that looks for patterns in data sets consisting of data points that are neither categorized or labeled. As a result, algorithms are allowed to categorize, label, and group data points within the dataset without the need for external supervision. Besides that, the learning algorithm is not given any label. Instead, it is leaving to its own device to identify structure in the input collection. However, there are no specific categories in unsupervised learning. AI systems will group unsorted material based on differences and similarities. In Figure 2.2, the main objective of unsupervised learning is to find hidden and interesting patterns in unlabeled data. Principal component analysis, anomaly detection, clustering, and autoencoders are the four of unsupervised learning [6].

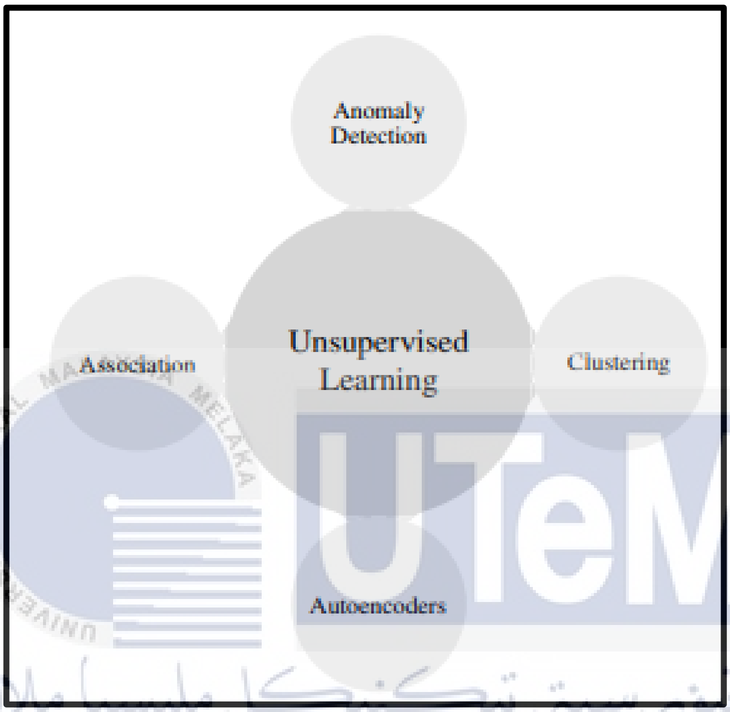


Figure 2.2 Type of Unsupervised Learning

Source: Dridi S 2021

The type of requirement for predicting fish growth rate in aquaculture is supervised learning. Supervised learning uses datasets with both observed values of the dependent variable output and explanatory variable input [7]. Explanatory variables are used in the model's building to produce predictions of fish growth rates that are as close as feasible to the observed values of the dependent variable [8].

2.3 Prediction Models

This purpose is to perform existing machine learning trends and techniques for predictive analytics [9]. Aquaculture technology and predictive models for fish growth rates have become popular. Machine learning approaches have shown a tool for predicting fish growth rates due to the increasing demand for fish consumption and the need for effective control of fish development. Collecting information on fish development and performing the feature engineering and necessary preprocessing is part of the data collection and preprocessing phase. To accurately estimate fish growth rates, many predictive models, classification model including regression, have been created. Here is about machine learning classification and machine learning regression to obtain which process is more suitable for fish growth rate prediction.

2.3.1 Machine Learning Classification

Machine learning classification is the process of identifying a model that best captures and distinguishes data classes or ideas with the goal of using the model to forecast the class of objects whose class label is anonymous [10]. The aim of a collection of training records, data items with well-known class labels, served as the foundation for the resulting model of the main ideas in data mining is classification, which describes the process of labeling events with predefined classes based on their properties. Classification methods were preselected to facilitate efficient analysis of large data sets [11].

2.3.2 Machine Learning Regression

Machine learning regression is a statistical technique used to study how independent variables and dependent variables relate to each other. It is important for machine learning predictive modeling because it makes it possible to predict continuous outcomes based on hypothetical connections. Regression requires minimizing the distance between each data point and the line to draw a line that best fits the data point. In comparison, regression is one of the main applications of supervised machine learning, along with classification. Regression focuses on predicting continuous outcomes, while classification categorizes objects based on learned attributes. The model training, both strategies require labeled input and output training data [12].

Regression analysis facilities the creation of models for forecasting trends, forecasting results, or detecting gaps in historical data by illuminating the relationship between independent factors and dependent variables. To avoid overestimation and inaccurate predictions, it is important to ensure that the labeled training data accurately reflect the entire population. In order to fully understand the relationship between variables, it is taken when selecting the appropriate features for regression analysis [12]. Regression is performed in machine learning using various methods and algorithms to produce regression modeling. These methods may use varying numbers of independent variables or handle different types of data. In addition, many machine learning regression models may posit multiple relationships between independent and dependent variables. When applied to data sets that contain nonlinear relationships, linear regression techniques assume that there is a linear relationship [13]. Both linear will obtain high accuracy in predicting fish growth rates.

Now we may ed interest to investigating further machine learning regression, especially on three types, which are linear regression, logistic regression, polynomial regression, and random forest regression in Figure2.3, Figure 2.4, Figure 2.5, and Figure 2.6 respectively [14]. As linear regression is to establish a correlation between an input independent variable and an output dependent variable using a linear model. Whereas, logistic regression is used when the output is categorized. It’s more like a categorization issue. Then, only when the data’s regression is linear can the linear regression procedure function. One of the rare instances of multiple linear regression models is polynomial regression [15].

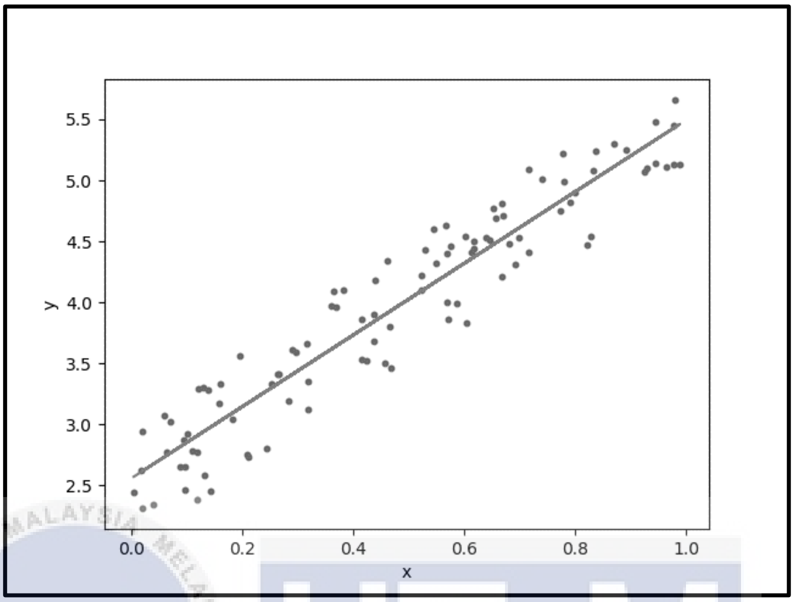


Figure 2.3 Linear Regression

Source: https://towardsdatascience.com/linear-regression-using-python-b136c91bf0a2

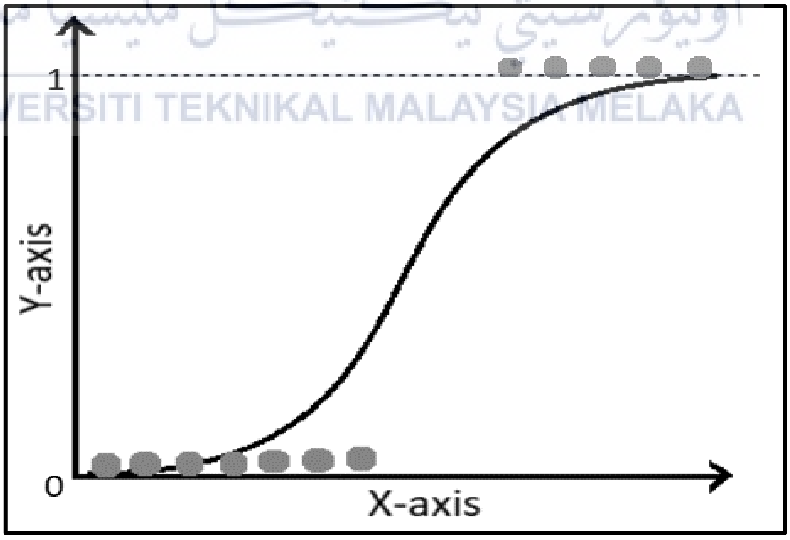


Figure 2.4 Logistic Regression

Source: https://medium.com/analytics-vidhya/the-math-behind-logistic-regression-c2f04ca27bca

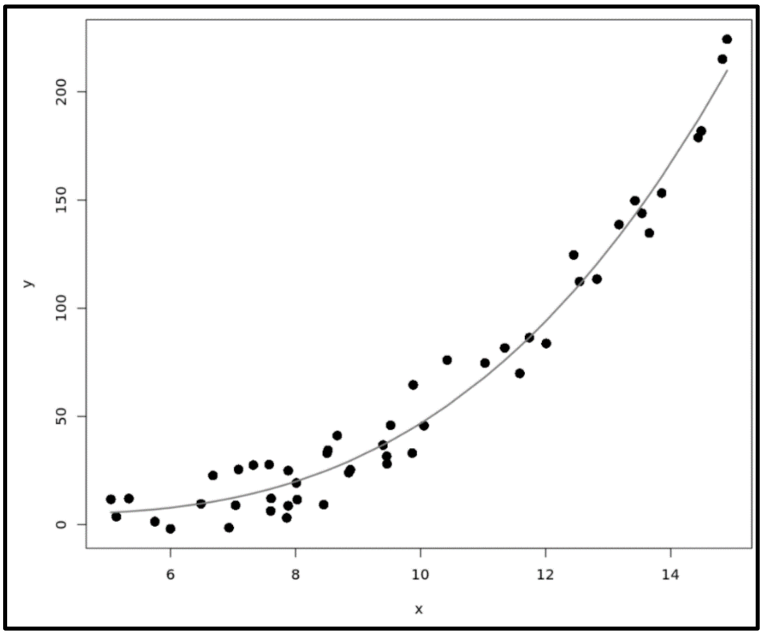


Figure 2.5 Polynomial Regression

Source: https://www.statology.org/plot-polynomial-regression-in-r/

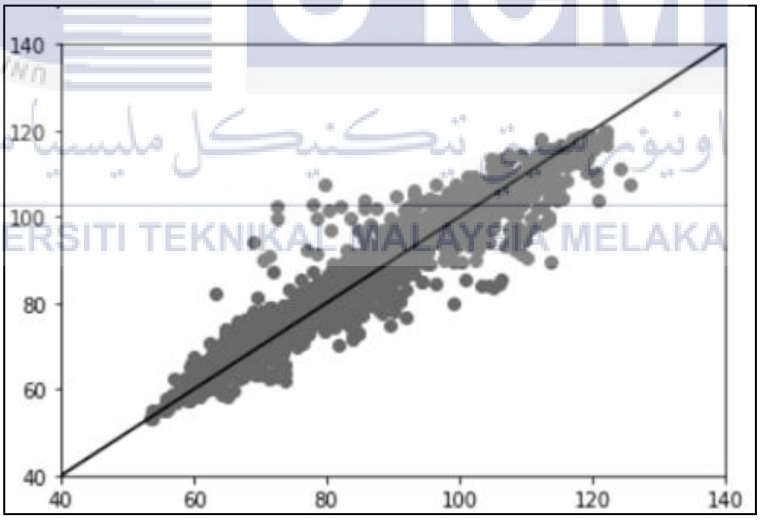


Figure 2.6 Random Forest Regression

Source: https://towardsdatascience.com/random-forest-regression-for-continuous-well-log-prediction-61d3ec1c683a

A regression model was specially created to modal and estimate the relationship between input factors and continuous output since the fish growth rate is a continuous output. The prediction model requirement for predicting fish growth rate in aquaculture is machine learning regression using machine learning regression to study the differences in graphs such as linear regression, logistics regression, polynomial regression, and random forest regression based on datasets that have been used as input to obtain the accuracy of fish growth rates in aquaculture.

2.4 Machine Learning in Aquaculture Application

Years back ago, most of the smart aquaculture care was handled manually by humans, from ponds management to fish growth. However, aquaculture, including feeding, size or biomass estimation, and water quality, will require lots of human energy, especially for large-scale aquaculture. The limitation here is that humans might not always manage and monitor their fish farms 24 hours a day. Then it comes to the situation where it is hard to manually feed, size or biomass estimation and water quality is providing for the raising for their aquaculture industry. Fish farming technology innovation is not a recent development. Aquaculture production quality could potentially increase, enhancing aquaculture production and reducing labor. It will make fish farms more intelligent and connect those realizing precision aquaculture and smart farming [16].

Aquaculture has grown and has been the focus of many researchers, which involves numerous studies in a particular area. According to [17] proposes developing an ensembled machine learning prediction, mainly focusing on a prediction accuracy model for marine fish and aquaculture production. The climatic data from Malaysia's five largest states are considered to assess the critical feature relevance in forecasting three distinct forms of fish production, which are freshwater aquaculture, marine landing, and brackish water aquaculture prediction of fish production using various machine learning algorithms. Therefore, this system creates an ensembled machine learning model using various ML techniques. Data on the highest and lowest sea surface, air temperatures, sea surface temperatures, rainfall duration, rainfall, and humidity have been gathered. The system displays various fish production rates and data for climate variables in five major states of Malaysia. The system included the data set with annual data on fish productivity and statistical data on several climatic variables. Then, [18] proposes the development of a smart fish farm system prototype based on the Internet of Things and artificial intelligence regarding current aquaculture issues.

They deployed several smart managements to help farmers collect-real-time data, achieve peak fishing performance and revive the metabolic processes linked to limited fisheries Next, refer to [19] purposed to develop applications for intelligent fish aquaculture in machine learning. The system, compared with traditional machine learning, neural network, and deep learning advancements, has increased the potential for intelligent applications in fish aquaculture and increased the effectiveness and value of breeding. The machine learning applications in aquaculture focused mostly on data collection, algorithm development, and data processing. There are two main components of machine learning and artificial intelligence data and algorithms.

2.5 Data Collection and Preprocessing

Preparing data for analysis and model training in machine learning requires procedures such as data collection and preprocessing. The required variables and characteristics are determined, and relevant data sources are selected during the data collection phase. Model performance, generalization, and prediction accuracy can all be improved with transformation, formatting, and proper data cleaning. While preprocessing is raw, data and images from the world are often unfinished, unreliable, and have no behavior or trend. There are several techniques of preprocessing, such as data imputations, oversampling, data cleaning, data normalization, and data integration [20].

2.5.1 Data Collection for Fish Growth Methods

In order to create predictive models, machine learning techniques for fish growth data collection require the application of relevant data on fish growth rates and related variables. The techniques used to acquire data vary. There are many several methods that can be applied for data collection, such as measures taken directly. Its approach requires taking accurate measurements of each fish growth characteristics, such as its weight, length, and biomass. Then, real-time information on fish development and environmental variables such as dissolved oxygen and water temperature is provided through sensor technology, such as telemetry systems or acoustic tags. Studies under controlled conditions change variables such as temperature and feeding schedules to analyze their influence on fish development. Data are preprocessed after collection to ensure quality, including cleaning management of missing information and appropriate formatting. A machine learning model is then trained using structured data set to predict fish growth rates based on the collected information and related factors. Knowledge of fish development and aquaculture systems may be enhanced using reliable data collection and machine learning techniques [21].

2.5.2 Data preprocessing and features for Fish Growth Methods

A processing stage that includes the segmentation and noise reduction techniques discussed earlier. Processing methods performed on fish images. Preprocessed fish images are used to extract features, calculate feature value, and combine feature value with fish mass to create a predictive model that can be used to estimate fish mass [22]. This may involve estimating gain or growth rates for fish based on weight and length data. To guarantee that numerical features are on the same scale, normalization, and scaling are used. Techniques of pre-processing can be used to deal with unbalanced data when the number of samples for various growth rates is unequal. Fish growth data were cleaned, normalized, and modified learning methods as part of the preparation stage. Data quality is improved during pre-processing, which helps create models to predict fish growth reliably and accurately.

2.6 Element of Accuracy for Fish Growth Rate in Aquaculture

To estimate the accuracy of fish growth rate are feeding, biomass estimation, and water quality prediction plays an important role and determining the quality of fish farms. These three elements will directly influence the accuracy process of the prediction for fish growth rate. In the scope of overview machine learning, due prediction for fish growth rate, close monitoring and control into these elements will determine the accuracy prediction. Therefore, this section will briefly discuss these three stated elements.

2.6.1 Feeding Controlling

The purpose is to improve the size and health of the fish by feeding fish. The amount of time between feedings affects how quickly the fish grows. Since fish often have a higher metabolism than other animals, they usually benefit from multiple feedings throughout the day. However, depending on the age, size, and species of the fish, the frequency of feeding may change. The method of feeding fish affects its growth. While underfeeding or overfeeding may have adverse effects. To avoid waste and water contamination and to ensure that fish are properly maintained, it is important to offer the appropriate amount of food [23]. In aquaculture, nutrition is one of the main issues. Farmers will spread the food, according to the conventional approach, either across the pond or in specific areas within the pond, according to the nutritional needs of the species being planted. Therefore, it is important to use IoT in feeding systems to automatically manage feeding and feed quantities, which has many advantages, such as reducing labor costs and food waste, as well as improving water quality in aquaculture [16].

An accurate understanding of fish feeding control can help direct processes, achieve optimal feed control, reduce feeding costs, and improve economic efficiency [24]. Refer to [25] have conducted extensive research on fish feeding control, using near-infrared imaging technology to evaluate fish feeding control and measure feeding control index using SVM and grey gradient symbiotic matrix. In addition, to categorize fish feeding control, a deep convolution neural network was used, and 90% accuracy was achieved in “none”, “weak”, “moderate”, and “strong” level [26].

2.6.2 Biomass Estimation

The purpose of this method is to create the next generation of fish sorting equipment using modern hardware, real-time length measurement, and programming techniques to identify species. This can reduce the expense of any mechanical feeding system while increasing the productivity of the Catch Meter. Fish length measurements are usually carried out on research vessels to the nearest 1cm accuracy using electronic measuring boards such as Scantrol in Norway or manual measuring boards to enter the data into the computer. A computer vision system that can autonomously measure the length of fish in the laboratory with an error of less than 1cm has been presented [27]. Classification has two several ways to estimate fish which are biological and geometrical characteristics. Biological characteristics such as weight, length, and gender for accurate estimation of the age of the fish. While geometrical characteristics such as otolith, vertebrae, and scales. On the other hand, the length of the perimeter and the area between the two axes of the otolith are calculated in interpretations based on age. But geometrical is challenging to acquire the sample and extract in some species of fish. However, these types of reading and observation are considered too dependent on factors that can affect the fish age estimation. This study evaluates the effectiveness of three underutilized machine learning algorithms for fish age estimation [21].

2.6.3 Water Quality

The water quality index shows the water quality of an environment. Therefore, real-time monitoring of water quality parameters of the aquaculture environment is important to detect reproductive biological anomalies, prevent disease and reduce related risks [28]. Refer to [29] environmental water quality characteristics are influenced by several factors, which make it difficult to make predictions of water quality measurements. Feeding frequency is widely recognized to reduce aquaculture production costs and reduce water quality degradation due to overfeeding. Fish development and survival depend highly on water quality, which is affected by several variables, including pH, temperature, ammonia levels, and dissolved oxygen. Most fish demands are met by the water in which they live. Therefore, it is important to understand the needs of fish for water quality. The system proposed a potential approach to smart fisheries that would support the monitoring of water quality indicators, making choices based on collected data, and faster adaption to changing conditions [30]. The water quality affects the fish growth rates. Therefore, water quality care must be taken seriously to understand its impact on fish health.

2.7 Dataset in Machine Learning Aquaculture

Machine learning datasets are crucial for learning from machine learning algorithms. A dataset, which contains labels that indicate whether a prediction was successful or unsuccessful, is an example of how machine learning aids with prediction. Datasets for machine learning are available in a wide range of formats and origins. The three most typical categories of machine learning datasets are textual, image, video, audio, and numeric data. A dataset is essentially a collection of data that may be used to predict future occurrences or results using data from the past. Machine learning algorithms generally label datasets before using them so the system understands what outcomes to predict or categorize as an anomaly. A set of data divided into validation, training, and test sets is referred to as a machine dataset. These datasets are generally used in machine learning to instruct algorithms on how to spot patterns in data [31]. A dataset's purpose is to train the model and evaluate the model's accuracy after training. There are three types of machine learning datasets which are the training dataset, validation dataset, and test dataset [32].

According to [33], the purpose was introduced, and a group of experimental water quality monitoring results gathered directly from a fish pond were given. Therefore, water quality monitoring is a significant problem, especially in the fish farming business. A computerized fish farming monitoring system is used. Then, refer to [34] the purpose of developing a machine learning system to predict ideal fish species for aquatic environments. The system used a random forest model to utilize a dataset of aquatic environments for 11 different fish species to validate the model. The system used various aquatic environmental variables, such as temperature, pH, and turbidity, to predict fish species. Depending on their similarity and physiognomy, fish can be identified using fish categorization. Then, computerized fish categorization can speed up the process and increase the accuracy of fish species identification. The system introduces several strategies for computerized fish species identification. In this project the dataset that will be used is text data. The natural language documents are categorized based on content using text classification. While images and video dataset are not covered in this project.

2.8 Python in Artificial Intelligent

The goal of artificial intelligence (AI), as a subfield of computer science, is to create computer programs and devices capable of performing activities that humans naturally master, such as image recognition, natural language, and speech recognition understanding. Python is a well-known programming language that has gained a lot of attention for its strong skills in the field of AI. Furthermore, using Python's advanced features is essential to create accurate and effective AI models that can work with large amounts of data. Advanced methods such as machine learning, natural language processing, and deep learning can help developers in creating AI models that can correctly predict and categorize data in various sectors [35].

Due to its independent platform and widespread use in the programming community, the Python programming language is the most suitable for machine learning. Machine learning research and application development have long used a variety of environments and programming languages. However, with the rapid growth of the general-purpose Python language in popularity over the past ten years among scientific computing, most current machine learning packages are Python based. Python is popular high-level programming language that can access the power of system-level programming language when needed while still being easy to learn. There are many different algorithms and methods available in Python that can be used for different AI tasks [36].

2.9 Figure of Machine Learning

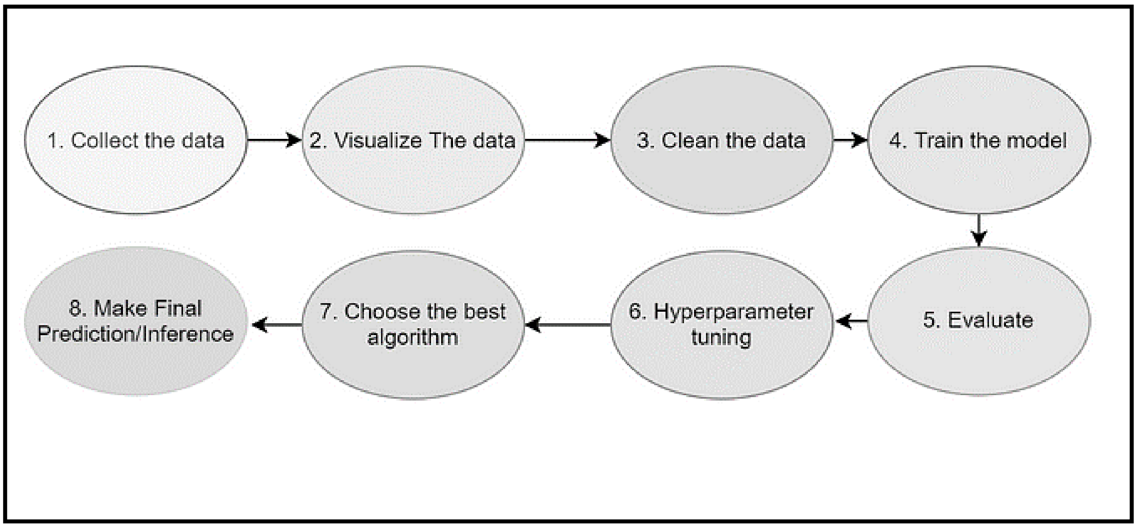


Figure 2.7 Process of Machine Learning

Source: https://towardsdatascience.com/regression-analysis-for-beginners-using-tree-based-methods-2b65bd193a7

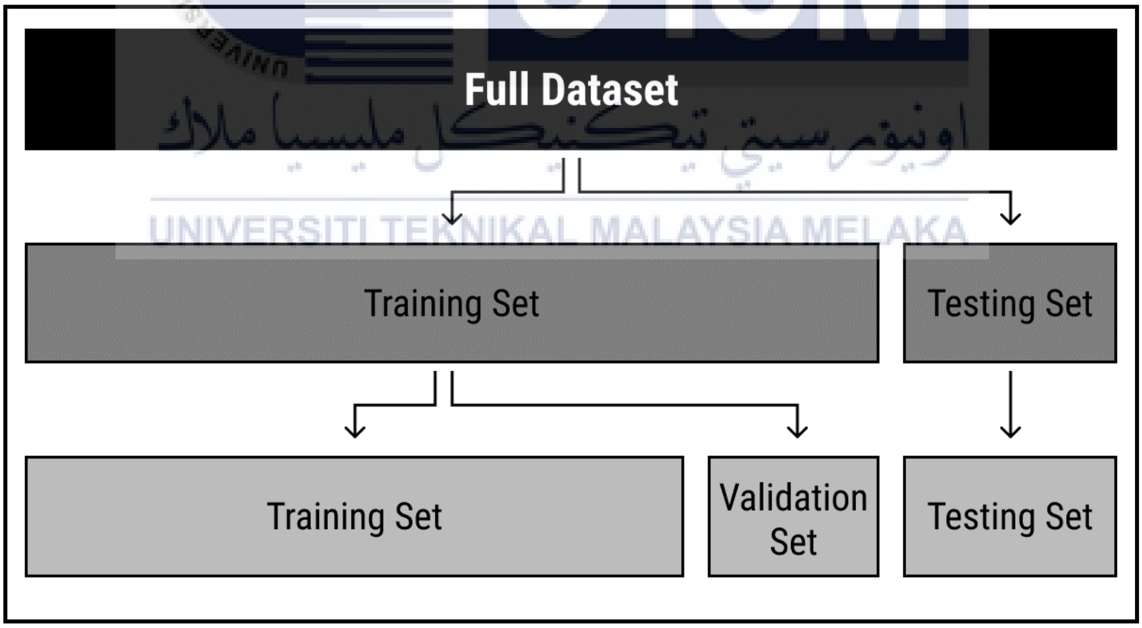


Figure 2.8 Dataset for Machine Learning

Source: https://labelyourdata.com/articles/what-is-dataset-in-machine-learning

2.10 Table of Related Work

Table 2.1 Summary of Related Work

NO	TITLE	AUTHOR	PLATFORM	PURPOSE	ADVANTAGE	DISADVANTAGE
1	Application Of Machine Learning in Intelligent Fish Aquaculture.	[38]	Machine Learning	The progress of machine learning in intelligent fish farming has been reviewed, and analyzed of the machine learning technique performed. In comparison, deep learning and neural network have expanded and possibilities for intelligent applications and improved productivity in fish aquaculture.	- Increase efficiency in aquaculture -	- Reduce labor workers

2	Introduction to Machine Learning	[2]	Machine Learning	The machine learning control developed out of the broader filed of Artificial Intelligent (AI), which attempts to replace human intelligence in machines. ML can be used for a variety of task. Classification is a traditional method that uses known qualities of a subject to assign it to predefined set of categories.	- Easily identifies patterns - Can handle variety of tasks.	- Time and resources - High error accuracy
3	Machine Learning - Regression	[37]	Machine Learning Regression	The idea of goal taking in. definitions of supervised and unsupervised learning is introduced, then followed by a discussion of the data separation algorithms. Regression analysis is a	- Easier to use, understand and train effectively - Useful for linearly separable data.	- Outliner sensitivity in linear regression is high

				statistical procedure that consists of various machine learning techniques, including data separation and regularization. After that, a significant amount of time was spent discussing regression techniques, with a concentration on linear regression.		
4	Machine Learning for Classification	[41]	Machine Learning Classification	Machine learning classification well in categorization challenges. It is essential to summaries and com pare the results of different classifiers in applying their unique classification tasks. The other classifier works similar on other datasets, namely	- The design and efficiency of machines. - Assist in classifying data into different classes.	- The algorithm does not immediately offer probability effective

				distinct specific classification task. The performance of various machine learning approaches in the application of classification problems.		
5	Application of Machine Learning and Computer Vision in Aquaculture	[16]	Machine Learning, Computer Vision	By making it possible to assess several factors, including weight, size, species categorization, and disease detection, smart aquaculture seeks to improve husbandry practices. The idea of "smart aquaculture" aims to develop the aquaculture sector sustainable and environmentally friendly way while increasing production. In addition,	- Reduce labor by automating culture systems - More accurate and faster than manual method by using algorithms or images	- Take much time to process water on time of water quality in ponds or tanks. - Difficult to accurate estimate the amount of food left in the pond

				using advanced technology, smart aquaculture aims to address issues related to traditional aquaculture.		that can affect the water quality. - Decrease the frequency of feed supply to the cultured systems
6	Developing an Ensembled Machine Learning Prediction Model for Marine and	[17]	Machine Learning	Environmental studies have widely used machine learning methods, which are effective for solving complex time series data. This research aims to create predictive models for marine fish production and aquaculture using machine	- Accurate estimation of fish production. - To predict and analyses fish landings of the coastal area.	- Limited dataset

	Aquaculture Production			learning. Three machine learning models were used to predict fish production based on a collection of climate factors linear regression, gradient enhancement, and random forest regression. Machine learning models are tested and trained using information using fish productivity and climate information.		
7	Applications Of Data Mining and Machine Learning Framework in	[39]	Data Mining and Machine Learning	With the growing demand for fish, the aquaculture and fisheries industries are implementing technology, data analysis and artificial intelligence. To increase productivity and ensure	- Reduces cost - Expedited of decision making	- Limitations of data - Scalability of issues

	Aquaculture A Fisheries.			sustainability. System to process complex data sets and deliver smart fisheries and agriculture solutions being created using data mining and machine learning techniques.		
8	Using Machine Vision to Estimate Fish Length from Image Using Regional Convolution Neural Network	[40]	Convolution Neural Network, Machine Vision	The accurate length estimates are made from an image by using a fiducial marker. However, the estimates are only accurate with controlling the imaging system. The methodology of this article is to estimate the total length, which uses machine vision. There are three regional convolution neural networks that trained from the public image.	- Increase the type and volume of data collected. - The accuracy of the data has been extracted in the machine vision system. - Real time detection will be volunteer-	- It difficult to divide species without a large number of high-resolution images using machine vision

				Machine vision can derive measurements of species and classify from the image without specialist equipment.	based data collection applications that can provide immediate feedback.	
9	Investigate of Some Machine Learning Algorithms in Fish Age Classification	[10]	Machine Learning Algorithms	The estimated age of fish can be used to predict their growth rate, which is important for managing fish populations. There are two methods for estimating in both marine and freshwater. Fish age can be estimated using biological characteristics such as length, weight and sex, while morphological characteristics such as	- Increase the accuracy of the estimate fish age. - Its ability to separate the training and test dataset. - The accurately of the machine	- The lack of data when the issues more data is available in the training process.

				otoliths, vertebrae and scales are also helpful. This study aims to evaluate three machine learning techniques for fish age estimation that currently need to be more appreciated for categorization. Additionally, this study will contrast traditional statistical approaches with machine learning techniques for age categorization and estimation.	learning algorithms classifying attributes in dataset.	
10	Automatic Measurement of Fish Weight and Size by Processing	[42]	Image Processing Software, Regression Algorithm	In order it explores on automated measurement systems that employ underwater camera systems, image enhancement, and segmentation algorithms. The method that includes	- It can improve the accuracy of fish length and weight measurements.	- Cost of installing underwater camera system

	Underwater Hatchery Images.			acquiring images with a single camera in a controlled environment enhancing fish images with a combination of homomorphic filtering, contrast limited adaptive histogram. The method is proven to be efficient in accurately and non-intrusively obtaining information from fish hatcheries.	- Allowing hatchery managers to determine necessary food amounts and population parameter	

2.11 Summary

The overview of machine learning carried out in this chapter is done mainly for the comprehensive study and further investigate the nature of this aquaculture. It is also to identify important elements that will influence the increase in fish growth rates. Realizing that local farmers are increasingly using machine learning for aquaculture, even though they continue to use traditional farming methods until they are motivated to adopt new practices. This project proposing the prediction model for fish growth rate using machine learning in aquaculture technology. Smart aquaculture is realized through the implementation of datasets, which will enable the precision fish farming and reduce human intervention.

CHAPTER 3

METHODOLOGY

3.1 Introduction

The methodology generally refers to the process or procedures used to identify, process, analysis and select the information connected to the research. Due to its effective methods to meet the growing world demand, aquaculture technology has revolutionized the fish farming sector. The project was built by using software component. The use of software for this project is to perform the program that has been created. The tools and flow of the project were clarified in this chapter where methods of the project were explained in details. A software development will take into consideration user interface design and arrangement, as well as the features. In this phase depends mainly on the Python programming language and PyCharm Integrated Development Environment (IDE).

3.2 Machine Learning Model

The projects to build machine learning regression model to prediction model fish growth rates. In machine learning there are many regressions model that could be used to predict the growth rate of fish. Therefore, various methods need to evaluated to compare the models and choose the high accuracy for prediction fish weight.

3.2.1 Data Collection

A dataset is load from CSV file using the panda’s library as a part of data collection. The effectiveness and reliability of the subsequent prediction model largely depends on the quality and usability of the data obtained using the CSV files. Biometric measurements (width, length, and height) were collected for grouper fish.

3.2.2 Data Preparation

The data are preprocessed for analysis. Python programming by PyCharm IDE was used as a tool. The project will proceed with reading a dataset from a CSV file using the panda's library. Data preparation is to generate structured, standardized, and clean data sets that can used to train machine learning model.

The direction and intensity of the related linearity will be determined through correlation analysis between variables. Correlation analysis requires measuring and explaining of continuous variables on linear or non-linear interactions. Correlation coefficient values range from negative (-1) to uncorrelation (0) to positive correlation (+1). The tendency of the association is indicated by the sign of the correlation coefficient, which may vary either positive or negative.

3.2.3 Data Splitting

The dataset was separate into two portion which is training and testing set. The project was split into the training and testing dataset represented ratio 80:20, respectively, use the (train_test_split) function to get the Scikit-learn Python is trained using the training set, and its accuracy is tested using test set. The test data is used to evaluated of machine learning techniques once ML builds the relationship with independent and dependent to predict or identify alternatives.

3.2.4 Model Development

Developing prediction models for fish growth rates using machine learning in aquaculture technology be verified to ensure the requirements and specifications. Performance testing and requirements validation are the two testing methodologies that will be used. This project developing prediction based on fish features such as length, width, and month to predict weight. This project implemented regression algorithm as a way for predictions. The project performed the tree-based models which is random forest regression and linear regression which is polynomial.

3.2.5 Model Performance Evaluation

Model performance evaluation is comprehensive assessment using key metric such as Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and coefficient of Determination (R^2) the four metrics used to evaluated the accuracy of the model predictions.

MSE measures the level of error in the model. Then, MAE is referred to the average size of the error from the model prediction, the others RMSE shows the average difference between the observed and estimated data. Then R^2 calculates how well predictor predicts the variation of the response variable.

requirement verifications will be separated from the unsuccessful requirement verification.

3.3 Software Architectural Design

The ability to prediction model for fish growth rates in software accurately. This is due to the applications of machine learning techniques on any dataset to predict unknown output values consist of training phase and testing phase. Training the prediction models, observe validation of accuracy on training dataset, and evaluation of pre-trained model for the test dataset. It be used the collection data of feeding controlling, biomass estimation and water quality. These training and testing dataset was created using function in Python, which allows to randomly partition 80% for training and 20% for testing or may it will be 70% for training and 30% for testing. It can be various percentages for training and testing. Even below 60%, it is not valid, and then the process is repeating until the learning stabilizes and stops improving. The best pre-trained model was evaluated for the test dataset during the testing phase. By doing this arrangement the whole program code will be less complex since it will divide into two as well. Figure 3.1 shows the overall software architectural design is represented. In the ensuing sub-topics, each functional component will be briefly explained.

In Figure 3.2 is a model diagram which shows a block diagram of machine learning regression. It represents the training and testing for machine learning methods. By using a dataset, the algorithm will split into training dan testing of dataset. Training dataset, the model is developed, and the test dataset is used to determine it accuracy

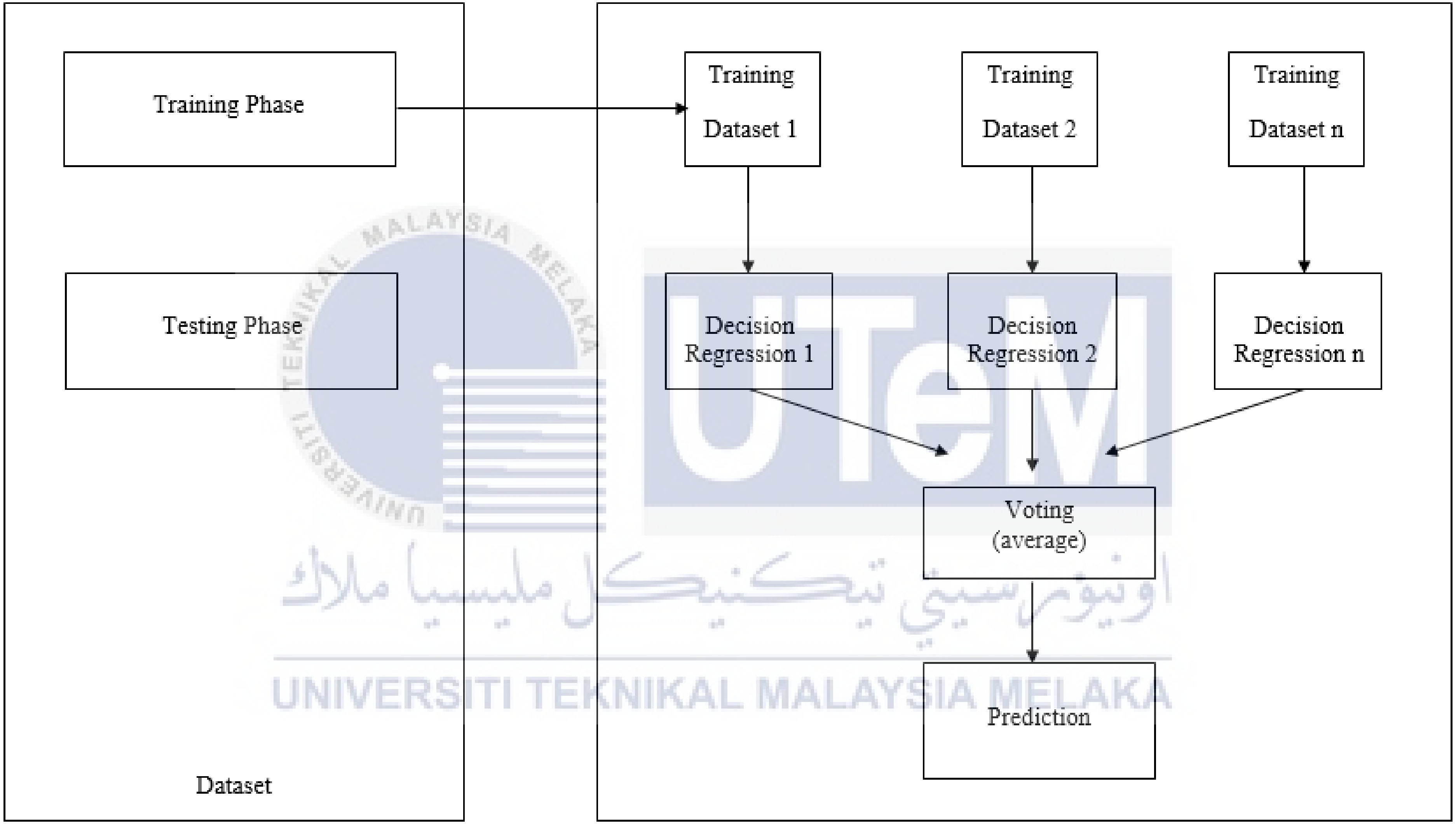


Figure 3.1 Software Architectural Design of Machine Learning Regression

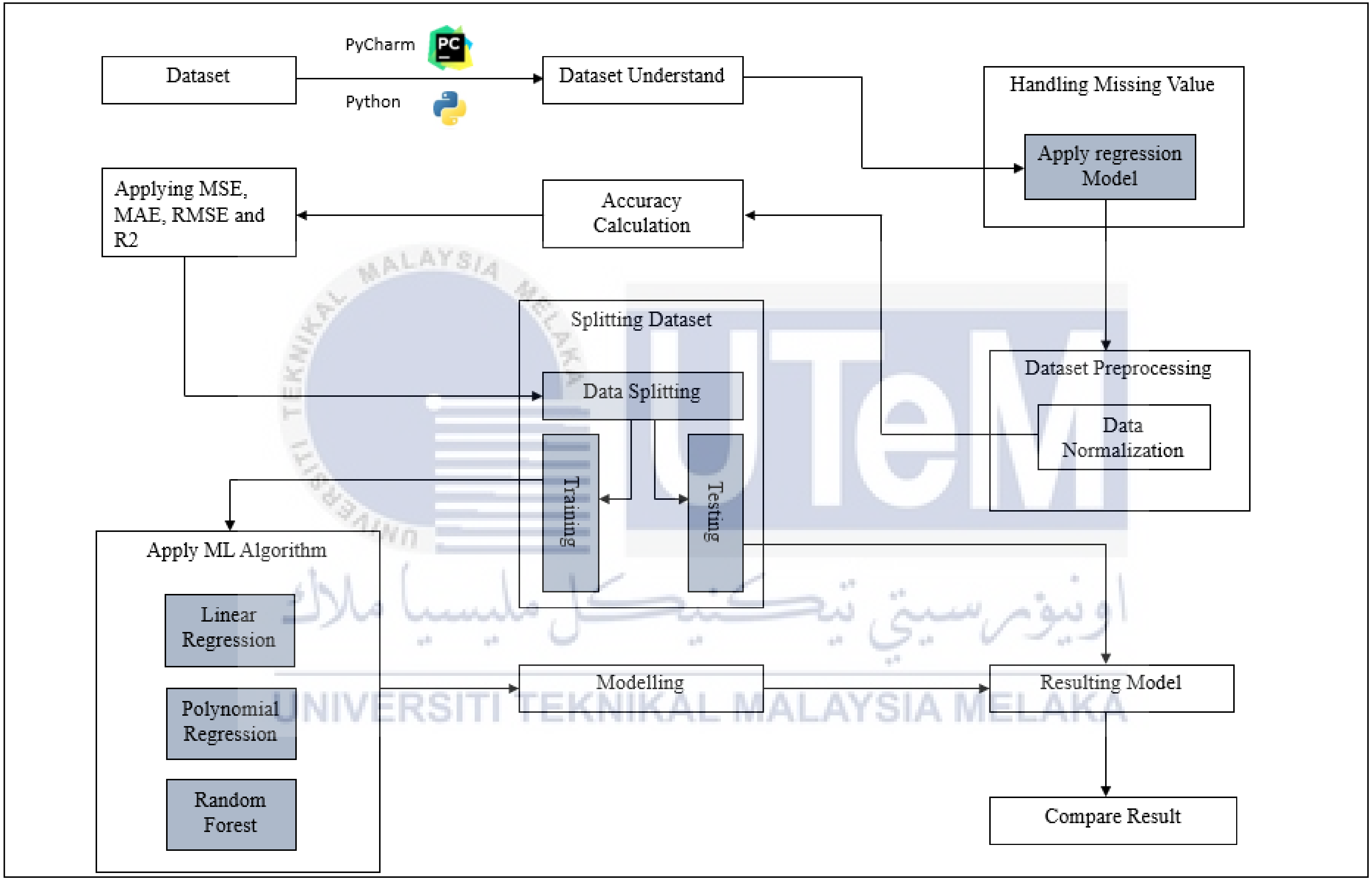


Figure 3.2 Model Diagram of Machine Learning Regression

Figure 3.3 illustrates the model diagram of Graphical User Interface (GUI) designed for prediction fish growth rate using machine learning regression. This interface allows users to select specific pages, such as the fish weight prediction page or the weight prediction page, that can estimate the weight of fish. Users can enter features by specifying length and width or selecting a certain month. The interface shows the expected and actual weight of the fish.

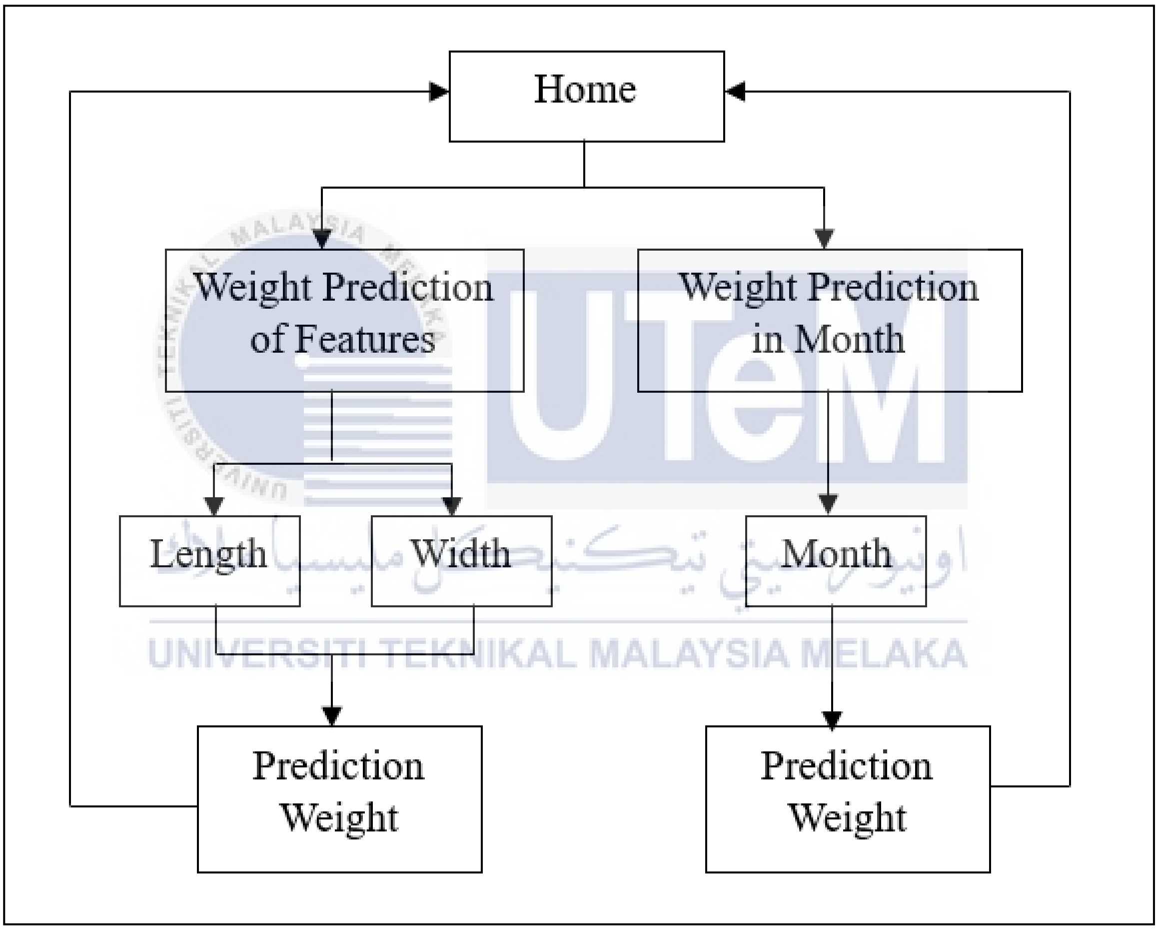


Figure 3.3 Model Diagram GUI

3.4 PyCharm Integrated Development Environment (IDE)

The software will be mainly used to build the software application for this project. Consequently, it is essential to the outcome of this project. Software like Python, PyCharm, and to produce Python GUI Development were used to complete this project.

Figure 3.4, PyCharm is an integrated development environment (IDE) is to create specifically for phyton programming. It offers various of features and tools to improve the efficiency of Python. PyCharm is to create a convenient environment for productive Python. PyCharm also assists developers in creating Python plugins using several available APIs. The IDE allows it to work directly with several database without integrating it with other tools. PyCharm provides methods for upgrading, installing, and uninstalling for Python. This suggests that each project has its collection of packages, which is considered a best practice for managing Python.

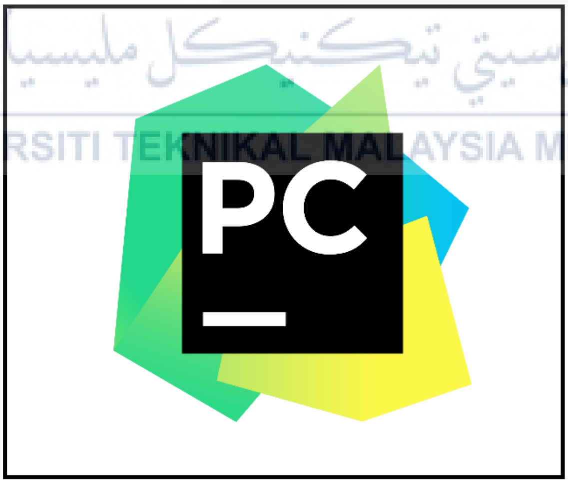


Figure 3.4 PyCharm Integrated Development Environment (IDE)

Source: PyCharm 2013

3.4.1 Python

Figure 3.5, A complete environment for Python development is the best IDE is provided by PyCharm, which offers a host of capabilities to aid with coding, debugging, testing, and others tasks. With working with Python with PyCharm. Python applications are created using PyCharm. Libraries of Pycharm that support Python is scientific, such as Matplotlib, Numpy, and Anaconda. It helps in building projects of machine learning. The standard library for Python is rather large and is at disposal. Contrary to other programming languages, this implies that programmers do not need develop code for every possible scenario. Libraries are available for various features, including expression, unit testing, databases, and image manipulation. There are thousands of additional components available in the Python Package Index in addition to the standard library, which is growing collection of them.

Figure 3.5 Python Programming Language

Source: Python 2023

3.4.2 Python GUI development using Python 3

In Figure 3.6, is an interface drawn on the screen for the user to interact with is known as a graphical user interface (GUI). Several tools and frameworks are available in Python that make it easy to develop graphical user interfaces (GUI). There are several GUI libraries to create user interface in python such as Tkinter, PyGTK, PyQT, wxPython, and Kivy. There are different types of packages for GUI development. The GUI library provides a wide range of functionality and complexity, and design options. The library chosen will depend on the specific needs, level of familiarity with the toolkit, and the platform targeting. In this project will be able chosen Tkinter.

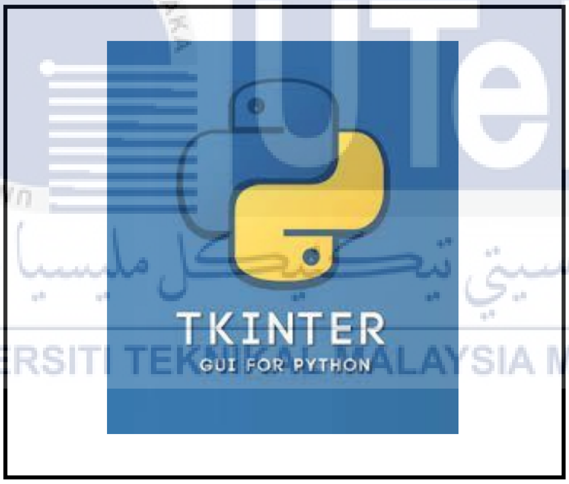


Figure 3.6 Tkinter of GUI

Source: Tkinter 2020

3.5 Software Flow Chart

Project flow is determined by data; therefore, research and data collection are essential to its production. Any data received can be changes or added depending on the latest needs. A prototype design will be made before the system is ready to verify that it works properly. This will be done after researching and defining the system requirements. Software prototype may need to be changed or sent back for further research and data collection if requests, such as new information or needed revisions arise. By doing so, it will be guaranteed that the system is always up to date.

Figure 3.7 shows the complete operational workflow for prediction model for fish growth rates and regression model evaluation such as linear regression, polynomial regression, and random forest regression that explain how the program work from Start until the last step End. The best regression is chosen based on the three regression models to predict the weight of fish.

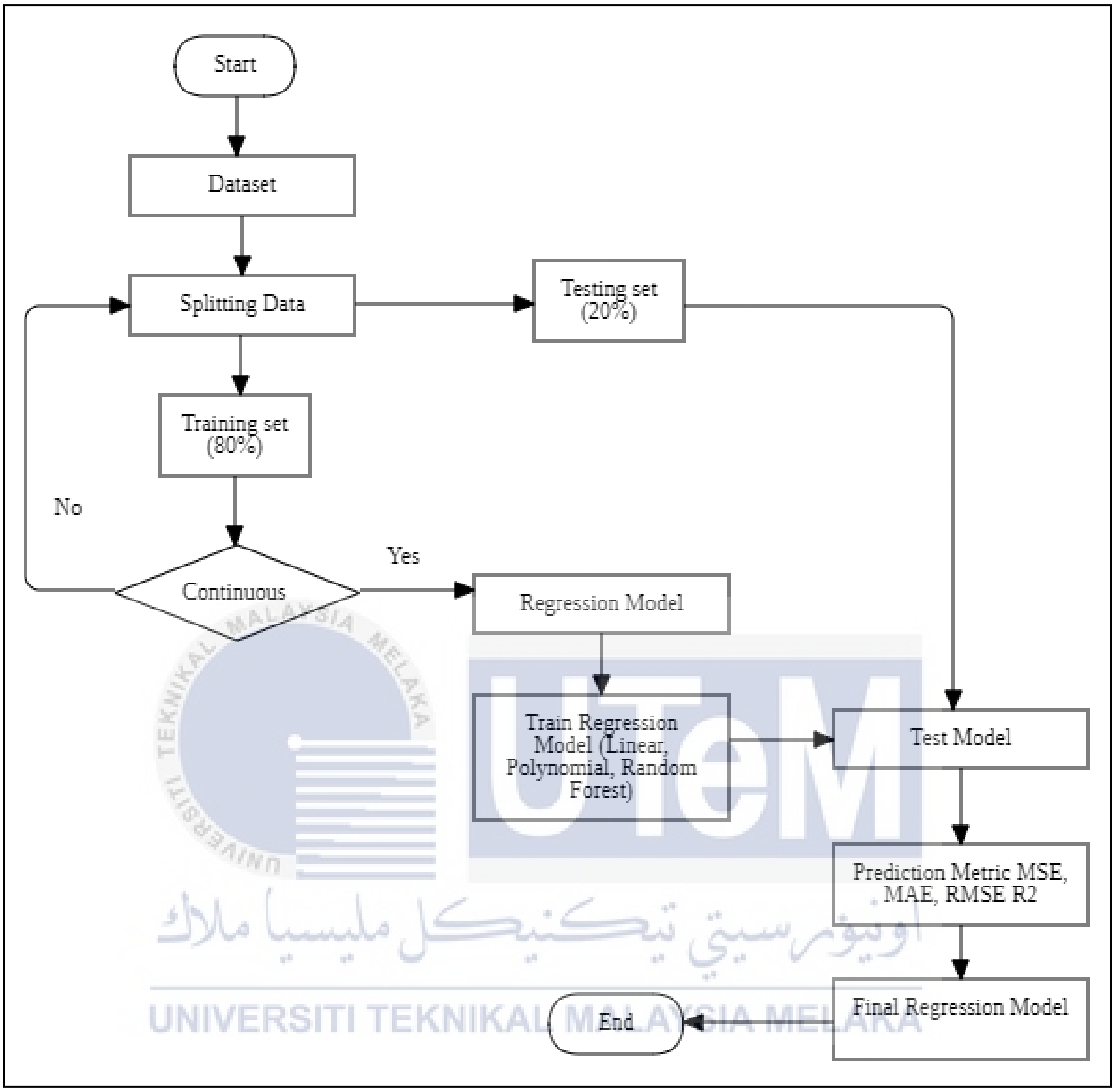
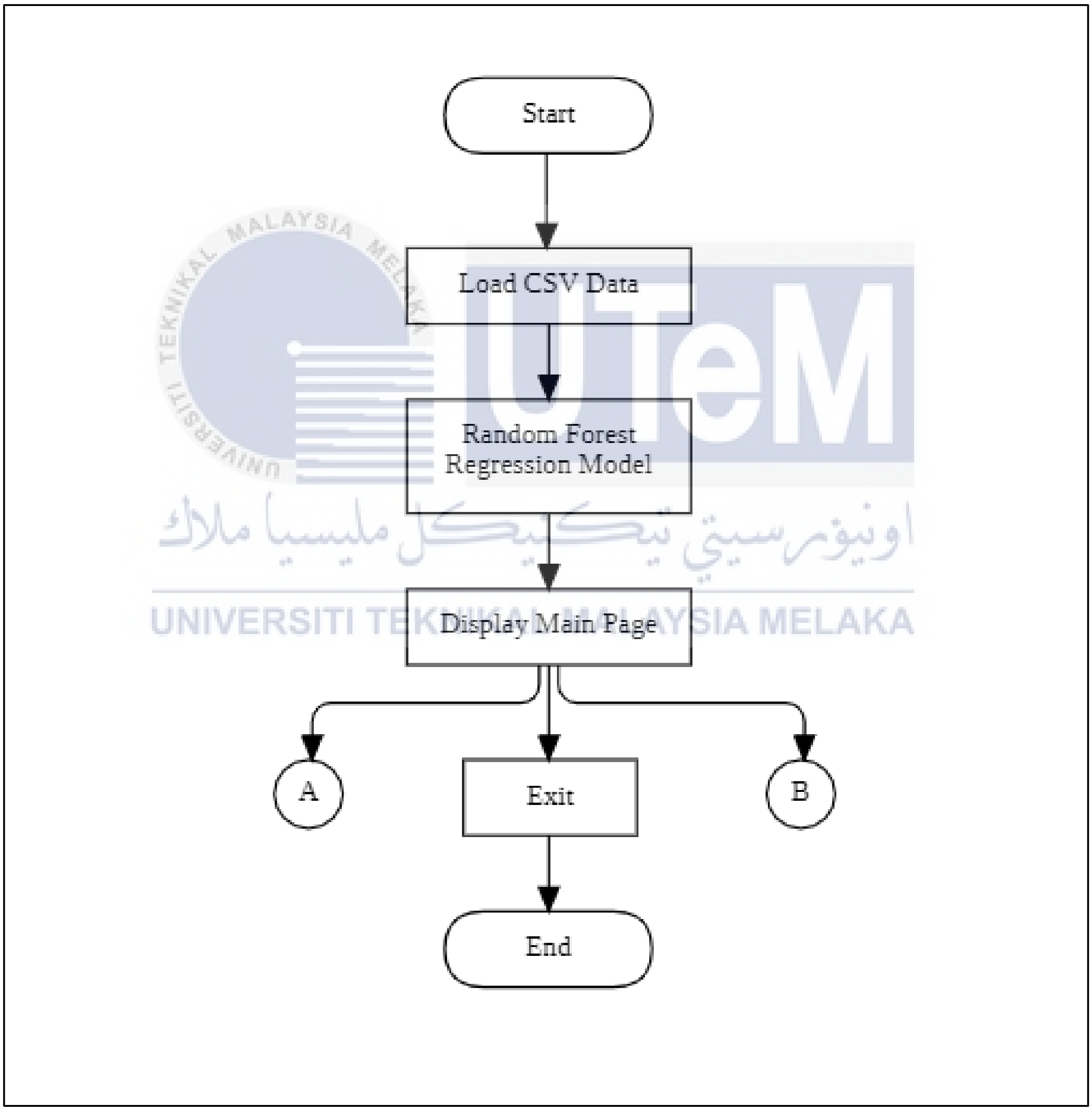


Figure 3.7 Flowchart of System Prediction Model for Fish Growth Rate

Figure 3.8 shows flowchart of graphical user interface (GUI) developed for predicting fish weight. This project implements a trained machine learning model which is Random Forest Regression to predicted weight in grams of the fish. Fish's width, length, and months are as an input. It allows to select from month 1 until month 12. The prediction weight of the fish based on fish's width and length will predict the weight in grams. Then for month based on selected month will predict weight in grams.



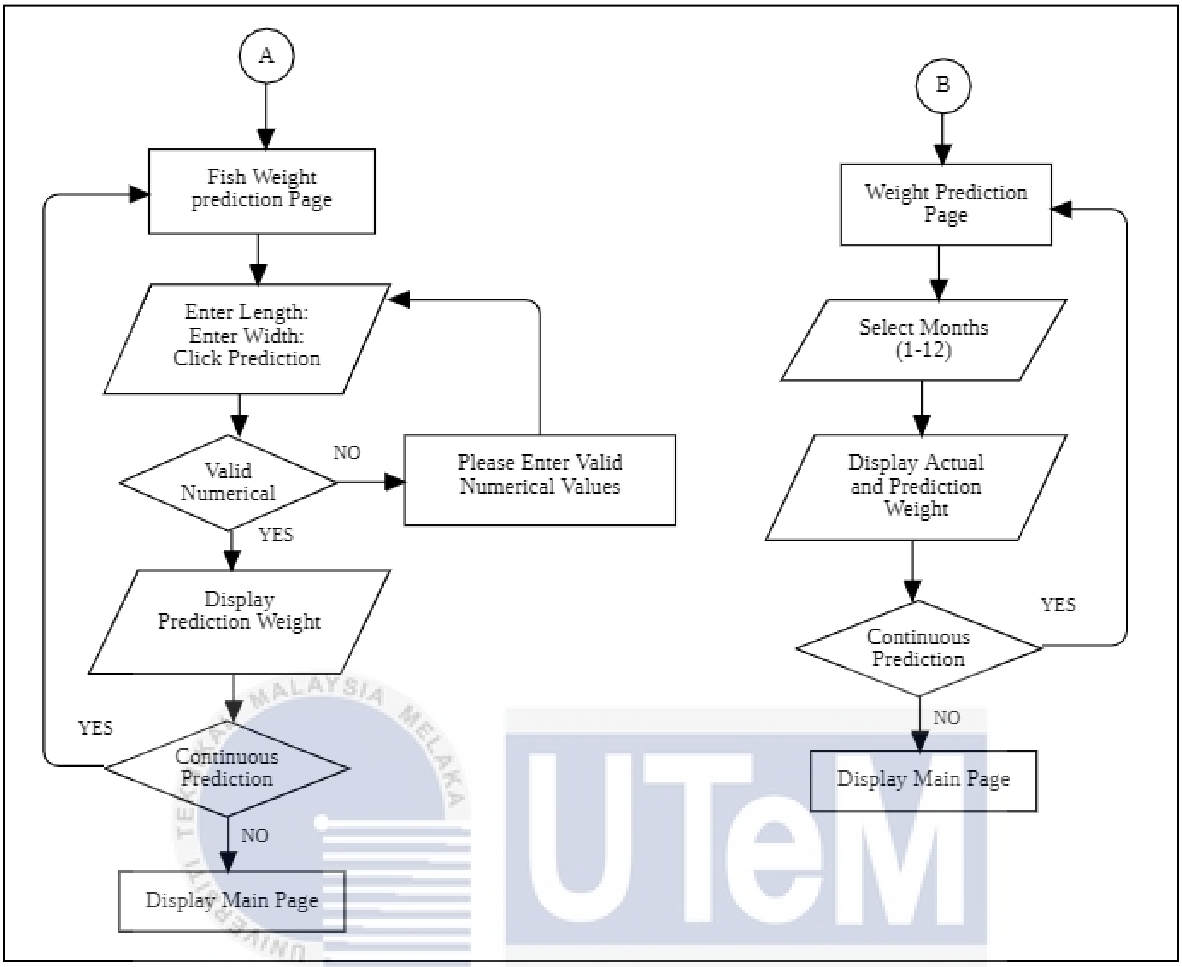


Figure 3.8 Flowchart of Prediction Weight in GUI

3.6 Equipment

Below is list of equipment in Table 3.1 shows the software application will be used for this project. Each of the software has their own function to make this project successfully running.

Table 3.1 Software Application

Software	Description	Version
Windows 11 Home Single Language	Operating system	Version 22H2
Python	Computer programming language are often used to create software and website, automated process and analyze data.	Version 3.11
PyCharm	Dedicated Python IDE providing a wide range of essential tools for Python developers.	Version 2023.1.2
StarUML	Drawing tools	Version 5.1.0
Microsoft Office	Report writing	Office 365 2021

3.7 Summary

This chapter covers project planning and system requirements in the framework of the system development process. The methodology of the entire project comprises software architectural design, software development. in this project, using PyCharm IDE, python, to develop project that includes a graphical user interface (GUI) for machine learning algorithms. The GUI enables interaction with the machine learning model, data input, results viewing, and configuration customization. When developing and testing the project’s code, managing dependencies, and integrating version control for collaborative development, PyCharm offers a convenient environment.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Introduction

This chapter presents the results and analysis of the prediction model for fish growth rates using machine learning in aquaculture technology. Basically, the main result is predicting fish weight to helps estimated fish population distribution and size, which is important for managing and monitoring fish. Extensive design and analysis features are the main emphasis of the design when complex intricacies are considered. The design phase when the design and analysis are carried out in more details as mentioned in the analysis section, the system is developed according to modules such as insert, update, save and delete data. The interface design will describe the system's input and output.

4.2 Result and Analysis

As a result, this system will build from one stage to the next using the work planning technique and methodology. Result in prototype that provides an overview of the software development process and identifies the pros and cons of prototype that needs to be performed before the entire system is implemented through deployment. A variable for prediction fish growth rate data has been observed. The most accurate model for prediction fish growth rates may be basic of this project. This project uses a dataset that is divided into two part which is training and testing. This model will be tested and trained with three regression models to find the accuracy of fish weight by using new technology.

4.3 Datasets Setup and Preparation

Datasets setup for this project are properly selected and after considering several factors, datasets the Kaggle website are chosen. This dataset consists of 159 biometric measurements across several common species of grouper fish which are *Bluelined*, *Leopard*, *Greasy*, *Cloudy*, *Duskytail*, and *Sixbar*. Grouper fishes' Biometric measurements are comprising of weight, height, width, vertical length (VLength), horizontal length (HLength) and diagonal length (DLength). Total count for corresponding groupers' species and biometric measurements are depicted as in Figure 4.1 shows Total Count among Grouper's Species. Correlation analysis against all biometrics measurement's variables as in Table 4.1 shows the result is close to 1 and can be concluded that this dataset is valid.

```
Species
Duskytail    56
Bluelined    35
Leopard      20
Spotty       17
Sixbar       14
Cloudy       11
Greasy        6
Name: count, dtype: int64
```

```
RangeIndex: 159 entries, 0 to 158
Data columns (total 7 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   Species  159 non-null    object
 1   Weight   159 non-null    float64
 2   VLength  159 non-null    float64
 3   DLength  159 non-null    float64
 4   HLength  159 non-null    float64
 5   Height   159 non-null    float64
 6   Width    159 non-null    float64
dtypes: float64(6), object(1)
memory usage: 8.8+ KB
None
```

Figure 4.1 Total Count among Grouper's Species

Table 4.1 Correlation Result Against all Biometric Measurement's variables.

	Weight	**VLength**	**DLength**	**HLength**	**Height**	**Width**
Weight	1.000000	0.916010	0.918891	0.923356	0.724561	0.886956
VLength	0.916010	1.000000	0.999517	0.992031	0.625378	0.867050
DLength	0.918891	0.999517	1.000000	0.994103	0.640441	0.873547
HLength	0.923356	0.992031	0.994103	1.000000	0.703409	0.878520
Height	0.724561	0.625378	0.640441	0.703409	1.000000	0.792881
Width	0.886956	0.867050	0.873547	0.878520	0.792881	1.000000

4.4 Regression Model Evaluation and Selection

In Table 4.2 the result of regression score of model evaluation and selection. This regression occurs because random forest is capable non-linear patterns and interactions between variables by splitting the feature space into areas and different model which is linear regression and polynomial regression. Random Forest Regression could identify complex interaction between variables effectively. Out of three machine learning regression model that have been evaluated, the Random Forest Regression was determined to be the most predictive capability that achieve a high accuracy. This model shown the lowest value of MSE, MAE, and RMSE which has been determined to be 6490.34, 46.20, 80.56 and 0.97 respectively.

Table 4.2 Regression Score of Model Evaluation

Model Name	MSE	MAE	RMSE	R^2
Linear Regression	16818.02	86.63	129.68	0.91
Polynomial Regression	182217.34	149.87	426.87	0.03
Random Forest Regression	6490.34	46.20	80.56	0.97

Hence, the Random Forest Regression model outperformed the Linear Regression and Polynomial Regression models, especially when those values result in fish weight prediction accuracy of several common species of grouper fish. Consequently, build a performance machine learning model with parameter fish which is duration, width and length will be requirement for the next phase of this project. This model will go through training and testing, for performance model for reaching high accuracy in predicting fish weight.

4.5 Performance Model Machine Learning Evaluation

Fish's width and length dataset are fed as an input into Machine Learning model to predict their weight as a target variable accordingly. These input datasets were loaded for preprocessing and manipulation which later will further split into two categories, training, and test datasets respectively. The Standard Scaling method is used to normalize the data to ensure equal scaling across features, as well as target variables.

Training and test process involving three selected Machine Learning models which are Linear Regression, Polynomial Regression with degree 2 and Random Forest Regression. Underlying patterns for all those models are revealed through training dataset execution while the outcome of the prediction performance right after the training process are evaluated by executing test dataset. The performance measures are indicated by the statistical parameter values which are R-squared Values of R^2, Mean Squared Error of MSE, a Mean Absolute Error of MAE and Root Mean Squared Error of RMSE.

4.5.1 Correlation Result

The process of identifying relevant variable for prediction model can be simplified with correlation analysis. Table 4.3 show all the variables did not have multicollinearity (R-value =|1|) which can lead to unsatisfactory predictions between correlations with variables that have been evaluated. The ability of variables to decide at a predictable rate can determined using correlation.

The project represents the parameters of fish, and shown the correlation coefficient demonstrate between the parameters. However, when the correlation indicates to positive, it indicates that one variable is highly predictable when compared to the other. So, based on that all the parameters which is width, length and weight were fitted with the Tree-Based Regression and Linear Regression models.

Table 4.3 Correlation Between Variables

	Width	**Length**	**Weight**
Width	1.000000	0.989781	0.916928
Length	0.989781	1.000000	0.934092
Weight	0.916928	0.934092	1.000000

4.5.2 Model Performance Result

Based on this project of machine learning models, the performance using various metrics such as Mean squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R-Squared (R^2). It is explicit from the comprehensive evaluation that Random Forest Regression models generally perform more accurately for predicted fish weight than Linear Regression and Polynomial Regression.

Compared to other models, the Random Forest Regression models showed lower MAE and RMSE values, indicating that its prediction was more accurate. Although the Mean Square Error (MSE) of Linear Regression model is lower than other models, it is important to interpret the results in relation to certain characteristics of the data as well as modelling assumptions. A lower MSE indicates that the linear model may perform better in reducing the sum of squared difference between the predicted and actual values. MSE is a statistic punishes larger errors more severely.

In Table 4.4 the result of performance score for prediction model. Out of three performance machine learning model that have been evaluated, the Random Forest Regression was determined to be the most predictive capability that achieve a high accuracy.

This model shown the lowest value of MSE, MAE, and RMSE which has been determined to be 6540986.24, 1468.15 and 2557.54 respectively.

Table 4.4 Performance Score for Machine Learning Model

Model Name	MSE	MAE	RMSE	R^2
Linear Regression	8158007.71	2029.82	2856.22	0.84
Polynomial Regression	8801912.38	2169.81	2966.80	0.83
Random Forest Regression	6540986.24	1468.15	2557.52	0.87

Hence, the Random Forest Regression model outperformed the Linear Regression and Polynomial Regression models, especially when those values result in fish weight prediction accuracy. Consequently, build a regression model will be requirement for the next phase of this project. This model will go through training and testing, such as implementing the random forest regression technique, with the purpose of reaching high accuracy in predicting fish weight.

4.5.3 Regression Analysis and Model Evaluation

Figure 4.2 represented graph is separate into three subplots, each shows the accuracy of various regression models which are Linear Regression, Polynomial Regression and Random Forest Regression in predicting fish weight depending on parameters such as length and width from test dataset. In this subplot, the y-axis shows the weight predicted by the fitted model, while the x-axis shows the actual weight of the fish.

The blue dots in the first of subplot, which are consistent with the Linear Regression, show the expected weight compared to the actual. In the same way the second subplot uses green dots to show the performance of Polynomial Regression. The red dots in the third subplot represent the Random Forest Regression findings.

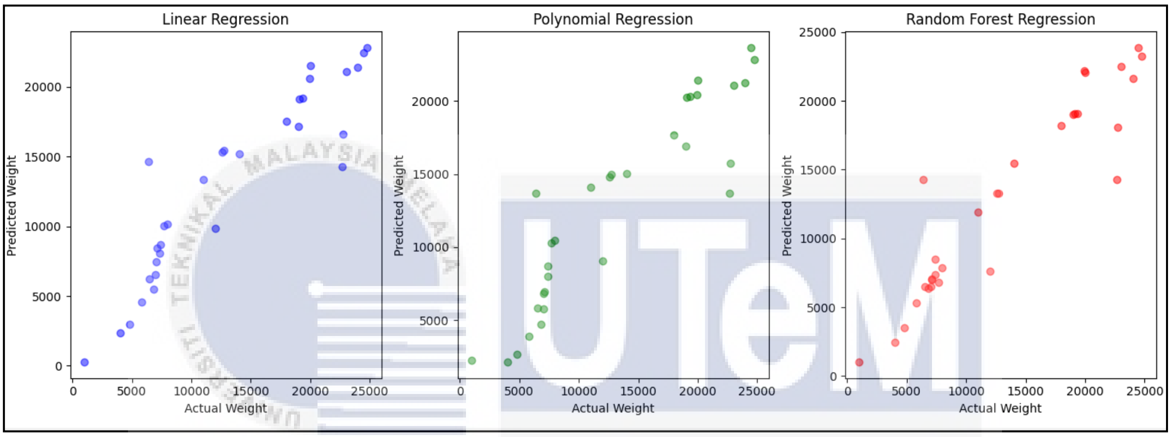


Figure 4.2 Scatter Comparison of Actual vs. Predicted Weight

Hence, these subplots show the Random Forest Regression aligns with the actual weights. Therefore, an accurate prediction occurs by well-performing Random Forest Regression that display a distribution of points that follows a diagonal line.

4.6 Graphical User Interface

A graphical user interface (GUI) for the prediction fish weight. A GUI built to reduce the traditional or manual method of estimating fish weight to determine the growth rate of the fish. Therefore, this project is built new technology based on solutions to improve the performance and growth rate of fish are prioritized. Random Forest Regression were implemented to predict the fish weight expected in the coming days. The input was in .CSV file which consists three features which is width, length, weight, and months. For each month in the dataset is trained by Random Forest Regression model, and the prediction are made on a test set. The application of GUI has three main pages that each have their own role.

To demonstrate the estimation weight of fish using a GUI interface. Random Forest Regression is an analysis method that applied to this prediction. Figure 4.3 is main page interface. Each display has their own function that can user choose to select on such as Fish weight by length and width or fish weight by month. The main pages will provide options to navigate to various features. It also has exit button to close the main pages.

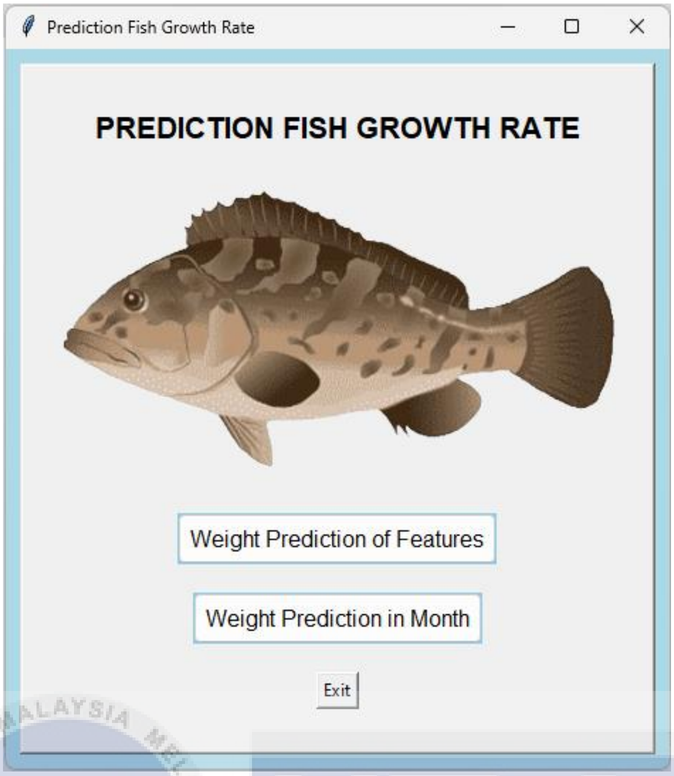


Figure 4.3 Main Page of GUI

Figure 4.4 is interface for prediction fish wight with input fish's width and length that can display estimated weight of fish. When insert length of fish in cm and width of fish in cm, it will display the estimated weight of the fish. Figure 4.5 shows the estimated weight of the fish displayed is approximately the actual weight of the fish. Then it has a home button that go back to main page.

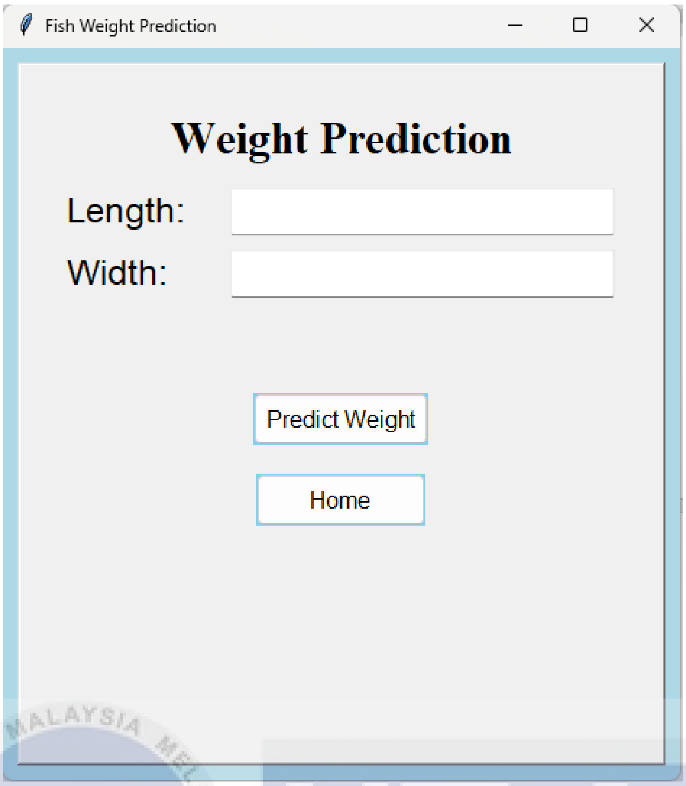


Figure 4.4 Fish Weight Prediction Page

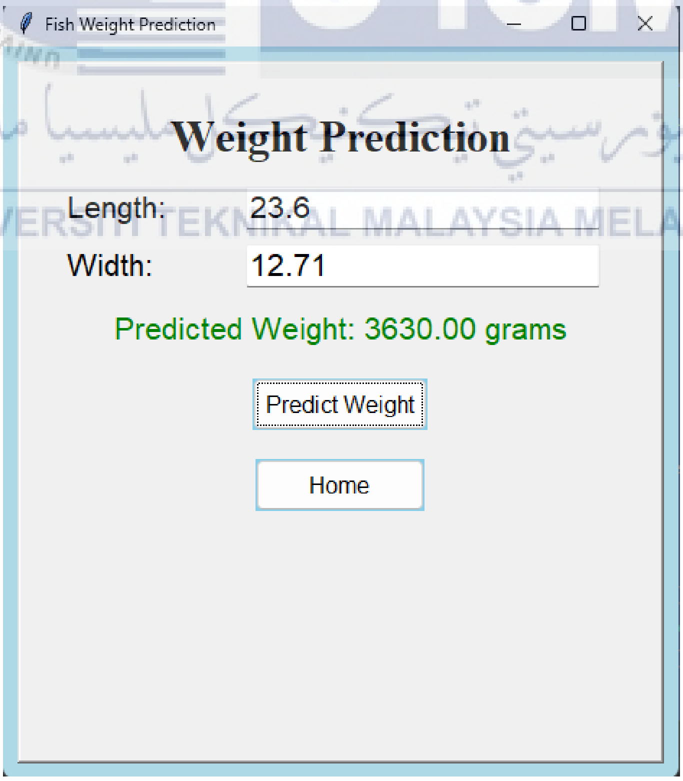


Figure 4.5 Prediction Weight of the Fish

Figure 4.6 shown the interface for prediction weight with input months that can display estimated of fish. To make it easier for user, when inserting the fish of month interface. User only need to select the month in which the fish is farmed, and the estimated weight of fish in that month will be displayed. Figure 4.7 show the prediction weight in months. When the user selects the month, it displays the actual and predicted weight in that month. This page also has home button where when click the button it will return to main page.

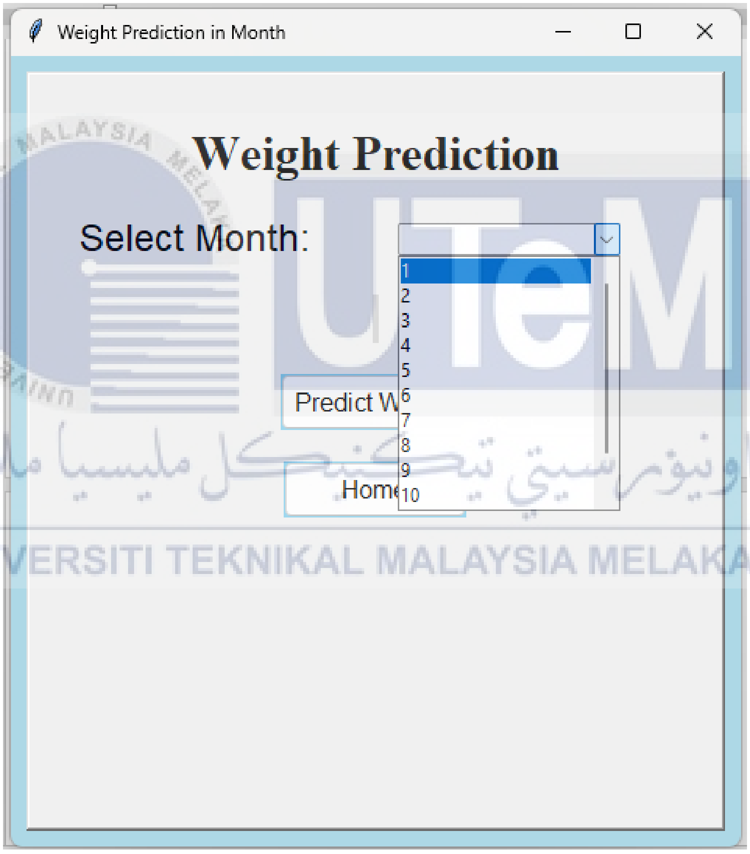


Figure 4.6 Fish Weight Prediction in Month Page

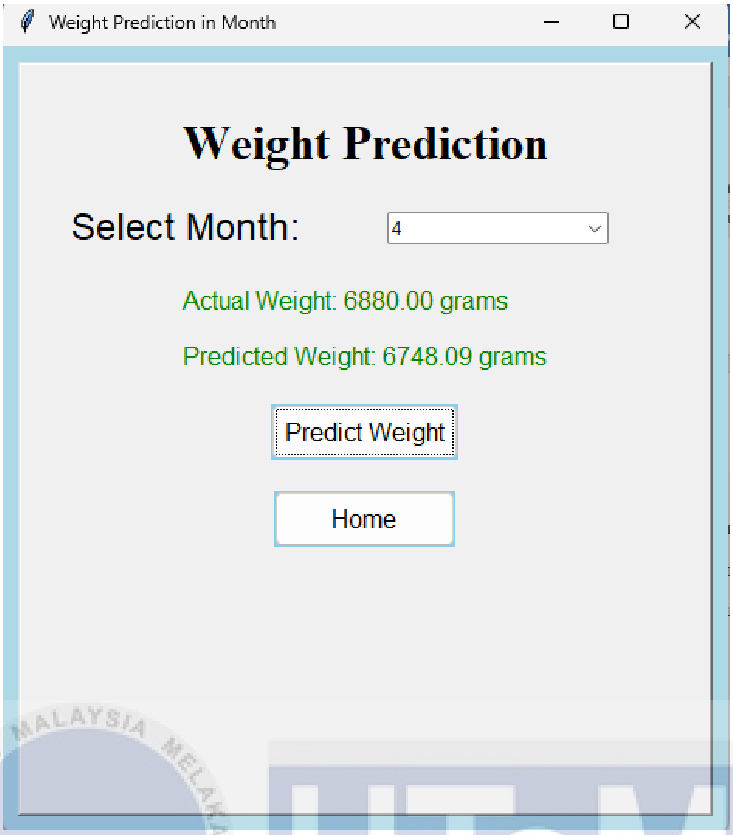


Figure 4.7 Prediction Weight in Months

4.7 Summary

This chapter presented the implementation of datasets and algorithms was the foundation for the machine learning development models in aquaculture. Data and algorithms are two main components of artificial intelligence and machine learning. This strategy that all developments systems will run on time and smoothly. That ensure the system can be completed within the specified timeframe, avoiding disruption of the system's development operations.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This final chapter outline the main findings from conducted and verification process, thus will conclude the entire project successful measures. Basically, this project is use of machine learning regressions algorithms for predicting fish growth rate in aquaculture technology. Through the experience of this project, several areas have been defined as potential for improved in future work and studies related to the aquaculture technology.

5.2 Conclusion

This thesis presents a method prediction model for fish growth rates using machine learning regression method in aquaculture technology. The proposed predictions model provides valuable information on different fish species dynamic and growth patterns, which helps aquaculture operators make better decisions. The methodology applying machine learning regression concepts specifically to prediction fish growth rate is the focus of this project. For the purposed of develop predictions, this thesis uses two different types of regression which are linear regression (also known as linear, and polynomial) and tree-based regression (also known as random forest). However, compared to linear regression, the tree-based Random Forest Regression model yields more accurate prediction results. Fish weight was reliably predicted by the ML proposal, which had high R^2 values, low MSE and MAE values. With the highest accuracy value among the three proposed ML regressions, Random Forest is the most effective prediction model, the lowest values for MSE, MAE, and RMSE score. Systematically the outline involves developing predictive models including data collection, model training and testing using existing methods to verify effectiveness.

5.3 Potential for Commercialization

Grouper fish is an essential priority for Malaysian fish farmer. Hence it is raised in specific raising facilities for subsequent marketplace sale. Despite the general incidence of traditional fish farming methods, this project aims to demonstrate the beneficial effects of new technologies in filed, especially by implementing the integration of machine learning. Specially, the goal is to develop a vision and prediction system that is capable of monitoring the growth rate and health of grouper fish spawn. Furthermore, the project intentions to develop an automatic inspection method for mature grouper fish seeds, deploying the abilities of machine learning techniques. The effective implementation of these technologies is predicted to significantly improving grouper fish farmers' production and income in Malaysia.

5.4 Future Works

In the future, machine learning (ML) techniques and algorithms may contribute to more accurate predictive models. Exploring advanced machine learning algorithms, classification, regression, and element of accuracy for fish growth rate such as feeding controlling, biomass estimation and water quality may yield improved the growth rate and other techniques for improved predictive accuracy and model interpretability such as image processing and segmentation. Further investigation, into assemble learning and deep learning method, or combining machine learning algorithms with other methods such as image processing and segmentation, may lead to more accurate predictions, interpretable models, and generalization. It can help the sustainable and conversation use of fish populations and their habitats by integrating weight prediction into aquaculture research and management methods.

REFERENCES

[1] B. C. Small, R. W. Hardy, and C. S. Tucker, "Enhancing fish performance in aquaculture," *Animal Frontiers*, vol. 6, no. 4, pp. 42–49, Oct. 2016, doi: 10.2527/af.2016-0043.

[2] T. Burzykowski, A.-J. Rousseau, M. Geubbelmans, and D. Valkenborg, "Introduction to machine learning," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 163, no. 5, pp. 732–734, May 2023, doi: 10.1016/j.ajodo.2023.02.005.

[3] H. Abbass, "Editorial: What is Artificial Intelligence?," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 2. Institute of Electrical and Electronics Engineers Inc., pp. 94–95, Apr. 01, 2021. doi: 10.1109/TAI.2021.3096243.

[4] P. Xu, "Review on Studies of Machine Learning Algorithms," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, May 2019. doi: 10.1088/1742-6596/1187/5/052103.

[5] Q. Liu and Y. Wu, "Supervised Learning," *Encyclopedia of the Sciences of Learning*, pp. 3243–3245, 2012.

[6] S. Dridi, "UNSUPERVISED LEARNING-A SYSTEMATIC LITERATURE REVIEW A PREPRINT," 2021.

[7] A. Mishra and A. Dasgupta, "Supervised and Unsupervised Machine Learning Algorithms for Forecasting the Fracture Location in Dissimilar Friction-Stir-Welded Joints," *Forecasting*, vol. 4, no. 4, pp. 787–797, Dec. 2022, doi: 10.3390/forecast4040043.

[8] X. Zhu and A. B. Goldberg, "Introduction to Semi-Supervised Learning," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 3, no. 1, pp. 1–130, May 2009.

[9] P. L. Bokonda, K. Ouazzani-Touhami, and N. Souissi, "Predictive analysis using machine learning: Review of trends and methods," *2020 International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*, Nov. 2020.

[10] S. Benzer, F. H. Garabaghi, R. Benzer, and H. D. Mehr, "Investigation of some machine learning algorithms in fish age classification," *Fish Res*, vol. 245, p. 106151, May 2022.

[11] "A Complete guide to Understand Classification in Machine Learning."

[12] S. Kumar and V. Bhatnagar, "A Review of Regression Models in Machine Learning," *Journal of Intelligent Systems and Computing*, vol. 2, no. 2, pp. 40–47, Sep. 2021.

[13] "Understanding Polynomial Regression Model."

[14] D. Castillo, "Machine Learning Regression Explained - Seldon." Oct. 2021.

[15] "polynomial." Aug. 2022.

[16] T. T. E. Vo, H. Ko, J. H. Huh, and Y. Kim, "Overview of smart aquaculture system: Focusing on applications of machine learning and computer vision," *Electronics (Switzerland)*, vol. 10, no. 22. MDPI, Nov. 01, 2021. doi: 10.3390/electronics10222882.

[17] L. F. Rahman, M. Marufuzzaman, L. Alam, M. A. Bari, U. R. Sumaila, and L. M. Sidek, "Developing an ensembled machine learning prediction model for marine fish and aquaculture production," *Sustainability (Switzerland)*, vol. 13, no. 16, Aug. 2021, doi: 10.3390/su13169124.

[18] M. C. Chiu, W. M. Yan, S. A. Bhat, and N. F. Huang, "Development of smart aquaculture farm management system using IoT and AI-based surrogate models," *J Agric Food Res*, vol. 9, Sep. 2022, doi: 10.1016/j.jafr.2022.100357.
[19] S. Zhao *et al.*, "Application of machine learning in intelligent fish aquaculture: A review," *Aquaculture*, vol. 540, p. 736724, Jul. 2021, doi: 10.1016/J.AQUACULTURE.2021.736724.
[20] "How data collection & data preprocessing assist machine learning."
[21] S. Benzer, F. H. Garabaghi, R. Benzer, and H. D. Mehr, "Investigation of some machine learning algorithms in fish age classification," *Fish Res*, vol. 245, p. 106151, Jan. 2022, doi: 10.1016/J.FISHRES.2021.106151.
[22] L. Zhang, J. Wang, and Q. Duan, "Estimation for fish mass using image analysis and neural network," *Comput Electron Agric*, vol. 173, p. 105439, Jun. 2020, doi: 10.1016/J.COMPAG.2020.105439.
[23] H. S. AlZubi, W. Al-Nuaimy, J. Buckley, and I. Young, "An intelligent behavior-based fish feeding system," in *2016 13th International Multi-Conference on Systems, Signals & Devices (SSD)*, 2016, pp. 22–29. doi: 10.1109/SSD.2016.7473754.
[24] L. I. Zhen-bo, N. I. U. Bing-shan, P. Fang, and L. I. Guang-yao, "Estimation Method of Fry Body Length Based on Visible Spectrum," *Spectroscopy and Spectral Analysis*, vol. 40, no. 11, p. 1243, Jun. 2020, [Online]. Available: https://www.researching.cn/articles/OJc28c68bbfeeddf2b
[25] C. Zhou *et al.*, "Near infrared computer vision and neuro-fuzzy model-based feeding decision system for fish in aquaculture," *Comput Electron Agric*, vol. 146, pp. 114–124, Mar. 2018, doi: 10.1016/J.COMPAG.2018.02.006.
[26] C. Zhou *et al.*, "Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision," *Aquaculture*, vol. 507, pp. 457–465, May 2019, doi: 10.1016/J.AQUACULTURE.2019.04.056.
[27] D. J. White, C. Svellingen, and N. J. C. Strachan, "Automated measurement of species and length of fish by computer vision," *Fish Res*, vol. 80, no. 2–3, pp. 203–210, Sep. 2006, doi: 10.1016/J.FISHRES.2006.04.009.
[28] M. Zounemat-Kermani *et al.*, "Can decomposition approaches always enhance soft computing models? Predicting the dissolved oxygen concentration in the St. Johns River, Florida," *Applied Sciences (Switzerland)*, vol. 9, no. 12, Jun. 2019, doi: 10.3390/app9122534.
[29] H. Lu and X. Ma, "Hybrid decision tree-based machine learning models for short-term water quality prediction," *Chemosphere*, vol. 249, p. 126169, Jun. 2020, doi: 10.1016/J.CHEMOSPHERE.2020.126169.
[30] F. Akhter, H. R. Siddiquei, E. E. Alahi, S. C. Mukhopadhyay, M. E. E. ; Alahi, and S. C. Mukhopadhyay, "computers Recent Advancement of the Sensors for Monitoring the Water Quality Parameters in Smart Fisheries Farming Recent Advancement of the Sensors for Monitoring the Water Quality Parameters in Smart Fisheries," 2021, doi: 10.3390/computers.
[31] R. Koch, "Machine Learning Datasets - Definition, Applications, Resources." Jul. 2022.
[32] "Datasets for Machine Learning."
[33] A. I. Arafat, T. Akter, M. F. Ahammed, M. Y. Ali, and A. Al Nahid, "A dataset for internet of things based fish farm monitoring and notification system," *Data Brief*, vol. 33, p. 106457, Dec. 2020, doi: 10.1016/J.DIB.2020.106457.
[34] M. M. Islam, M. A. Kashem, and J. Uddin, "Fish survival prediction in an aquatic environment using random forest model," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 3, p. 614, Sep. 2021.

[35] M. I. Ashraf, "Advanced Techniques for Artificial Intelligence with Python," *Software Engineering*, vol. 2023, no. 1, pp. 7–10, doi: 10.5923/j.se.20231001.02.
[36] S. Raschka, J. Patterson, and C. Nolet, "Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence," *Information (Switzerland)*, vol. 11, no. 4. MDPI AG, Apr. 01, 2020. doi: 10.3390/info11040193.
[37] H. Saleh, J. A. Layous, and S. A. Republic, "Machine Learning-Regression Control of a 3DOF hover system View project Design and Implementation of a Reaction Wheel-Actuated Cube View project Machine Learning-Regression", doi: 10.13140/RG.2.2.35768.67842.
[38] S. Zhao *et al.*, "Application of machine learning in intelligent fish aquaculture: A review," *Aquaculture*, vol. 540, p. 736724, Jul. 2021, doi: 10.1016/J.AQUACULTURE.2021.736724.
[39] J. Gladju, B. S. Kamalam, and A. Kanagaraj, "Applications of data mining and machine learning framework in aquaculture and fisheries: A review," *Smart Agricultural Technology*, vol. 2. Elsevier B.V., Dec. 01, 2022. doi: 10.1016/j.atech.2022.100061.
[40] G. G. Monkman, K. Hyder, M. J. Kaiser, and F. P. Vidal, "Using machine vision to estimate fish length from images using regional convolutional neural networks," *Methods Ecol Evol*, vol. 10, no. 12, pp. 2045–2056, Dec. 2019, doi: 10.1111/2041-210X.13282.
[41] H. Tan, "Machine Learning Algorithm for Classification," *J Phys Conf Ser*, vol. 1994, no. 1, p. 012016, 2021, doi: 10.1088/1742-6596/1994/1/012016.
[42] G. Sanchez, A. Mario, and C. Arroyo, "Automatic measurement of fish weight and size by processing underwater hatchery images Supervised learning approaches for spatial-domain multi-focus image fusion View project." [Online]. Available: https://www.researchgate.net/publication/329250230