American International University-Bangladesh (AIUB)

# Modeling the Growth Dynamics and Price Prediction of Mud Crab (Scylla Olivacea) through Machine Learning and Multivariate Analysis

Islam, Shahriar Mushficul (19-39497-1)
Tanim, Sharia Arfin (20-42096-1)
Ayon, Shahriar Siddique (20-42265-1)
Khan, Mursalin (20-43035-1)

***Under the supervision of***

**Prof. Dr. Kamruddin Nur**
Professor, Department of Computer Science

*A Thesis submitted for the degree of Bachelor of Science (BSc) in Computer Science and Engineering (CSE) under the faculty of Science and Technology (FST) at*

**AMERICAN INTERNATIONAL UNIVERSITY-BANGLADESH (AIUB)**

February, 2024

## Abstract

Aquaculture, especially Mud Crab cultivation, is vital for meeting seafood demand; analyzing growth dynamics and predicting prices is essential for sustainability and economic viability. The primary objective of this research paper is to analyze the utilization of machine learning algorithms to predict prices and to model the growth dynamics of Mud Crab (Scylla Olivacea). We employ identical models to independently predict both crab growth and pricing. Our thorough evaluation of seven regression models reveals that three outperform the rest, demonstrating superior performance. Further scrutiny based on mean cross-validation RMSE values highlights the exceptional performance of our proposed Ensemble Model combining the Linear Regression and Bayesian Ridge Regression models. The proposed ensemble model excels with perfect fitting (0.00 MAE, MSE, RMSE, RMSLE, MAPE, and R2 of 1.00). To enable real-time predictions for crab growth and price, a web interface is also implemented which utilizes our proposed model.

## Declaration

This thesis is composed of our original work, and contains no material previously published or written by another person except where due reference has been made in the text. We have clearly stated the contribution of others to our thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, financial support and any other original research work used or reported in our thesis. The content of our thesis is the result of work we have carried out since the commencement of Thesis.


. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

**Islam, Shahriar Mushficul**
19-39497-1
Department of Computer Science & Engineering
American International University-Bangladesh

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

**Tanim, Sharia Arfin**
20-42096-1
Department of Computer Science & Engineering
American International University-Bangladesh

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

**Ayon, Shahriar Siddique**
20-42265-1
Department of Computer Science & Engineering
American International University-Bangladesh

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

**Khan, Mursalin**
20-43035-1
Department of Computer Science & Engineering
American International University-Bangladesh

**Approval**

This thesis titled **"Modeling the Growth Dynamics and Price Prediction of Mud Crab (Scylla Olivacea) through Machine Learning and Multivariate Analysis"** has been submitted to the following respected member of the board of examiners of the Faculty of Science and Technology impartial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering on dated and has been accepted satisfactory.

.........................................

**Prof. Dr. Kamruddin Nur**
*Professor* (Supervisor)
Department of Computer Science
American International University-Bangladesh

.........................................

**Prof. Dr. Md. Asraf Ali**
*Professor* (External)
Department of Computer Science
American International University-Bangladesh

.........................................

**Dr. Akinul Islam Jony**
*Associate Professor & Head of Undergraduate*
Faculty of Science & Information Technology
American International University-Bangladesh

.........................................

**Prof. Dr. Dip Nandi**
*Professor & Associate Dean*
Faculty of Science and Technology
American International University-Bangladesh

.........................................

**Mashiour Rahman**
*Sr. Associate Professor & Dean-in-charge*
Faculty of Science and Technology
American International University-Bangladesh

# Contents

# List of Figures

## Contributions by authors to the thesis

| | Shahriar Mushficul Islam | Sharia Arfin Tanim | Shahriar Siddique Ayon | Mursalin Khan | Contribution (%) |
|---|---|---|---|---|---|
| | *19-39497-1* | *20-42096-1* | *20-42265-1* | *20-43035-1* | |
| Conceptualisation | 5% | 30% | 40% | 25% | 100(%) |
| Data curation | 10% | 20% | 40% | 30% | 100(%) |
| Formal analysis | 10% | 30% | 30% | 30% | 100(%) |
| Investigation | 25% | 25% | 25% | 25% | 100(%) |
| Methodology | 5% | 25% | 45% | 25% | 100(%) |
| Validation | 4% | 32% | 32% | 32% | 100(%) |
| Theoretical derivations | 25% | 25% | 25% | 25% | 100(%) |
| Preparation of figures | 0% | 50% | 25% | 25% | 100(%) |
| Writing – original draft | 25% | 30% | 30% | 15% | 100(%) |
| Writing – review & editing | 0% | 100% | 0% | 0% | 100(%) |

**Acknowledgments**

I would like to express my heartfelt gratitude to Prof. Dr. Kamruddin Nur, Professor in the Department of Computer Science at the American International University-Bangladesh (AIUB). His guidance, unwavering support, and wealth of knowledge have been instrumental in the successful completion of this thesis.

I extend my sincere appreciation to my dedicated teammates, Shahriar Mushficul (19-39497-1), Sharia Arfin Tanim (20-42096-1), Shahriar Siddique Ayon (20-42265-1), and Mursalin Khan (20-43035-1). Our collaborative efforts and diverse skills have significantly enriched the quality of this research project.

Special thanks to the faculty and staff at AIUB, especially those in the Department of Computer Science and Engineering, for providing an excellent academic environment and resources necessary for the successful execution of our thesis.

I would like to acknowledge the invaluable contributions of all individuals involved in the data collection, analysis, and other aspects of this project. Your dedication has been crucial to the achievement of our research goals.

Lastly, I want to express my gratitude to my family and friends for their unwavering support and encouragement throughout this academic journey. Your belief in my abilities has been a driving force behind the successful completion of my Bachelor of Science in Computer Science and Engineering.

This thesis project would not have been possible without the collective efforts and support of everyone mentioned above. Thank you all for being an integral part of this academic endeavor.

## Keywords

Mud crab, Machine learning, Analysis, Prediction

# Chapter 1

# Introduction

Aquaculture, which is essential for supplying the world's seafood need. Among other seafood and fishes, mud crab cultivation is also very popular and profitable [1]. Mud crab is an ecologically important species naturally found in Bangladesh, especially in the tidal rivers of Khulna, Bagherhat, Satkhira, Cox's Bazar and Chattogram [2]. Mud crab farming thrives in Bangladesh, India, and China, thanks to mangroves and brackish waters. In FY22-23, mud crab earnings are 5.27million(EP Bdata), showingadeclinefrom11.82 million in the previous July to Dec FY21-22, indicating the influence of global recession on export trends [3]. The world's largest mangrove forest, shared by Bangladesh and India, is an ideal habitat for mud crabs. Approximately 7,770 km2 of estuaries and backwaters in this region contribute to an estimated 13,209 tonnes of potential crab resources, part of the total potential of 43,816 tonnes in India [13] [4]. China's Crab Market is projected to be USD 11.21 billion in 2023, with an anticipated growth to USD 14.81 billion by 2028, at a CAGR of 5.72% during the forecast period [14] [5]. Bangladesh's government supports mud crab cultivation through initiatives and policies targeting aquaculture development. With an extensive coastline along the Bay of Bengal, the country has favorable environments in brackish water bodies and estuaries for mud crab farming [6]. The tropical climate ensures consistent temperatures, promoting year-round growth. By integrating biological and economic perspectives, our research enhances the holistic understanding of Mud Crab farming. Our research demonstrates how technology may adapt to major problems, advancing Mud Crab cultivation and the discussion on tech-driven sustainability in aquaculture. Mud crab price was frequently set by local or regional market forces, with few regulation and sustainability measures.

The analysis fills a significant vacuum in the literature on the cultivation of mud crabs (Scylla Olivacea), as most of the previous research has concentrated on biological issues while ignoring important economic factors that are necessary for sustainable operations [7] [8]. Our research aims to bridge this disparity, providing a nuanced understanding essential for informed decisions by industry stakeholders. By focusing on this critical gap, our study contributes to a holistic approach, enhancing knowledge and offering practical implications for sustainable Mud Crab cultivation. In October to November 2018, a batch of one hundred female mud crabs was sourced from local fishermen in Raoping county, Guangdong province, China (geographic coordinates: N23°2620, E116°5447) for research and business purposes. Selected for robust health and intact appendages, these live crabs were sent to Shantou University's Marine Biology Institute for thorough examination [9]. Aquaponics integrates water-based plant cultivation and fish rearing, optimizing aquaculture and hydroponics [10]. Utilizing a DS18B20 temperature sensor submerged in the tank ensures fish-friendly temperatures (20°C to 30°C) [11]. This system thrives on temperature control synergy, embodying the regenerative nature of aquaponics. This study bridges gaps in Mud Crab literature, analyzing both biological and economic aspects. Mud crab growth displays size-weight patterns from juvenile to adult stages, with varying values in different developmental phases [12] [13] [14]. These fluctuations are vital indicators of the species' maturation and overall health, providing valuable insights for effective aquaculture management and sustainable practices. Profitability is ensured through optimized fish growth, where the study of mud crab moult increments (mean 1.67 ± 0.48 cm) aids in estimating moulting proportions and assessing trap selectivity [15].

This study fills a gap by using advanced machine learning and multivariate analysis to model Mud Crab growth and predict pricing trends. Our approach integrates innovative tools to analyze the ecological and economic aspects of Mud Crab (Scylla Olivacea) agriculture. Six models, including an ensemble model, were used to predict growth and price, and evaluated using six metrics in the regression problem. Smallscale, unregulated fishing was used to gather mud crabs. The mud crab industry faces several environmental challenges. These issues have raised concerns about the availability of suitable habitats for crab populations and how these factors might influence growth. Mud crab populations often exhibit genetic variability, which can result in variations in growth rates and size [16]. Understanding the genetic diversity of mud crabs can help in selective breeding programs to produce faster-growing and more resilient individuals [17] [18]. Unanswered questions about predation variations, impactful predators, and the impact of competition on crab growth are crucial for conservation and aquaculture. The study offers useful economic models to maximize mud crab farming, resolving farmers' worries about profitability and expansion.

## 1.1 Mudcrab

Scientifically referred to as "Scylla serrata," the mudcrab is an exceptional decapod crustacean that flourishes in a variety of coastal ecosystems across the globe. These organisms, which are distinguished by their robust exoskeletons and unique body structures, are common in estuaries, intertidal zones, mangroves, and littoral waters. Having the ability to tolerate different levels of salinity, their adaptability renders them exceptionally versatile and suitable for a wide array of environments.Mudcrabs are distinguished by their formidable exoskeleton, which resembles armor and provides robust protection against predators and environmental hazards. In general, the carapaces of these organisms exhibit a camouflaged or variegated design, facilitating their seamless integration with the muddy or sandy substrates in which they inhabit. Their streamlined, broad bodies facilitate rapid locomotion on the ocean floor, which is advantageous in their pursuit of sustenance and refuge.Mudcrabs exhibit a diverse dietary pattern, being omnivorous scavengers that consume detritus, algae, mollusks, small fish, and various invertebrates.

As scavengers and bottom-dwellers, they contribute to nutrient cycling and ecosystem equilibrium by aiding in the decomposition of organic matter, thereby playing a vital role in the maintenance of coastal ecosystem health.The life cycle of these crustaceans is noteworthy, as it consists of multiple molting stages during which they remove their exoskeletons to make room for development. Mudcrabs are more susceptible to predation and gentler during the molting phase, until their new exoskeleton has fully hardened.In numerous coastal areas, mudcrabs possess cultural and economic significance in addition to their ecological value. Their succulent and flavorful flesh renders them highly desirable as an ingredient in a wide range of cuisines across the globe. Due to their widespread appeal in seafood markets, mudcrabs have become the subject of commercial aquaculture and fishery endeavors, which have provided sustenance for numerous coastal communities.Mudcrab populations are frequently the focus of conservation initiatives owing to their economic and ecological significance. By striking a balance between harvesting and preserving their habitats and populations, sustainable management practices guarantee the species' continued existence in coastal ecosystems.The mudcrab is an essentially resilient and adaptable creature that contributes to the culinary heritage of numerous cultures and maintains the ecological equilibrium of coastal environments. The organism's remarkable adaptability to various habitats and substantial contributions to ecosystems and human sustenance underscore its importance within the complex web of marine life.

## 1.2 International Demand

The international demand for mud crabs has surged in recent years, fueled by their delectable taste and high market value. These crustaceans have become sought-after commodities in the global seafood trade, driving a significant demand that spans various continents.Primarily valued for their succulent meat, mud crabs are a culinary delicacy in numerous cuisines, particularly in Asian countries like China, Singapore, Malaysia, Thailand, and the Philippines. The sweet, tender flesh of these crabs is highly prized in gastronomy and is often featured in diverse dishes, from stir-fries to soups and crab cakes.The rising demand for mud crabs is not confined to Asia alone. Their popularity has extended to international markets in Europe, North America, and beyond, where they are considered a gourmet item. This growing global appetite has led to an increase in both import and export activities, with countries like Australia, Indonesia, Sri Lanka, and others actively involved in the mud crab trade.However, meeting this international demand poses challenges. Wild mud crab populations face pressure from overfishing and habitat degradation, leading to fluctuations in natural stocks. Consequently, to satisfy the demand, aquaculture has emerged as a significant contributor to the mud crab supply chain.Aquaculture practices have gained prominence as a means to meet the rising market demand while alleviating pressure on wild populations. Countries like Indonesia, the Philippines, and Vietnam have established mud crab farming operations, utilizing ponds, pens, or mangrove areas to cultivate these crustaceans. Controlled environments allow for the rearing of mud crabs from larvae to market size, ensuring a consistent supply to meet international demands.Efforts to improve farming techniques, such as selective breeding, optimizing feed formulations, and disease management, aim to enhance production efficiency and meet stringent quality standards required by international markets.The economic implications of the international mud

Figure 1.1: A captivating array of mud crab species showcasing their unique colors, sizes, and distinctive features in their natural habitat.

crab trade are substantial. For many coastal communities, mud crab aquaculture and harvesting provide livelihoods and income opportunities, contributing to local economies.However, sustainability remains a critical concern. Balancing the booming demand with responsible harvesting and aquaculture practices is crucial to preserve wild populations and safeguard the health of coastal ecosystems. Regulations, monitoring of fishing practices, and promoting sustainable aquaculture methods are essential steps toward ensuring the long-term viability of mud crab populations and meeting international demand without compromising their ecological significance.the international demand for mud crabs continues to rise, driven by their gastronomic appeal and market value. While aquaculture presents a promising solution to meet this demand, responsible practices and conservation efforts are imperative to safeguard the species and their habitats for future generations.

## 1.3 Machine learning

Machine learning has begun to play a crucial role in various aspects of mud crab research, particularly in areas related to population dynamics, aquaculture, and conservation efforts. These advanced computational techniques offer innovative solutions to analyze complex data sets, predict trends, and optimize management strategies.One significant application of machine learning in mud crab research involves population monitoring and assessment. Scientists utilize machine learning algorithms to analyze vast amounts of ecological data, including habitat characteristics, environmental parameters, and population demographics. By processing these data, machine learning models can aid in predicting population trends, understanding habitat preferences, and identifying potential threats to mud crab populations. These insights are invaluable for designing effective conservation measures and sustainable management practices.In aquaculture, machine learning techniques are instrumental in optimizing farming practices and improving production efficiency. By leveraging machine learning algorithms, aquaculturists can analyze diverse data sources, such as water quality parameters, feed compositions, growth rates, and disease patterns. These analyses help in developing predictive models to optimize feeding regimes, monitor health conditions, and enhance overall productivity in mud crab farms. By implementing machine learning-driven solutions, aquaculture practices can become more efficient, reducing resource waste and ensuring higher yields.Additionally, machine learning aids in disease detection and management in mud crab aquaculture. These algorithms can process patterns in disease outbreaks, identify potential risk factors, and predict susceptibility to infections. By doing so, preventive measures can be implemented proactively, minimizing the impact of diseases on mud crab populations and the aquaculture industry.Furthermore, machine learning contributes to enhancing genetic selection and breeding programs in mud crab aquaculture. Algorithms analyze genetic data to identify desirable traits for growth, disease resistance, and other economically valuable characteristics. By understanding the genetic makeup of mud crabs through machine learning-driven analyses, breeders can develop improved strains and accelerate the selective breeding process.However, the successful application of machine learning in mud crab research relies on the availability of high-quality, comprehensive datasets. Collaboration between researchers, aquaculturists, and data scientists is essential to collect, curate, and share data that can fuel these machine learning algorithms effectively.machine learning holds immense potential in revolutionizing mud crab research, particularly in population monitoring, aquaculture optimization, disease management, and genetic enhancement. Its application not only contributes to the sustainable management of mud crab populations but also enhances the productivity and resilience of mud crab aquaculture, paving the way for a more efficient and sustainable industry.

### 1.3.1 Mudcrab Price Prediction using Machine Learning

Predicting mud crab prices using machine learning involves leveraging historical pricing data, market trends, and various factors affecting the market to forecast future price movements. Machine learning models can analyze these complex datasets and make predictions based on patterns and relationships found in the data.To create a mud crab price prediction model, the first step involves collecting extensive historical data on mud crab prices. This dataset would include information on factors like seasonality, regional variations, supply chain dynamics, market demand, economic indicators, and any other relevant variables that might influence prices.Machine learning algorithms, such as regression models, time series analysis, or even more complex algorithms like neural networks, can be trained using this dataset. These models learn from historical patterns and relationships within the data to predict future mud crab

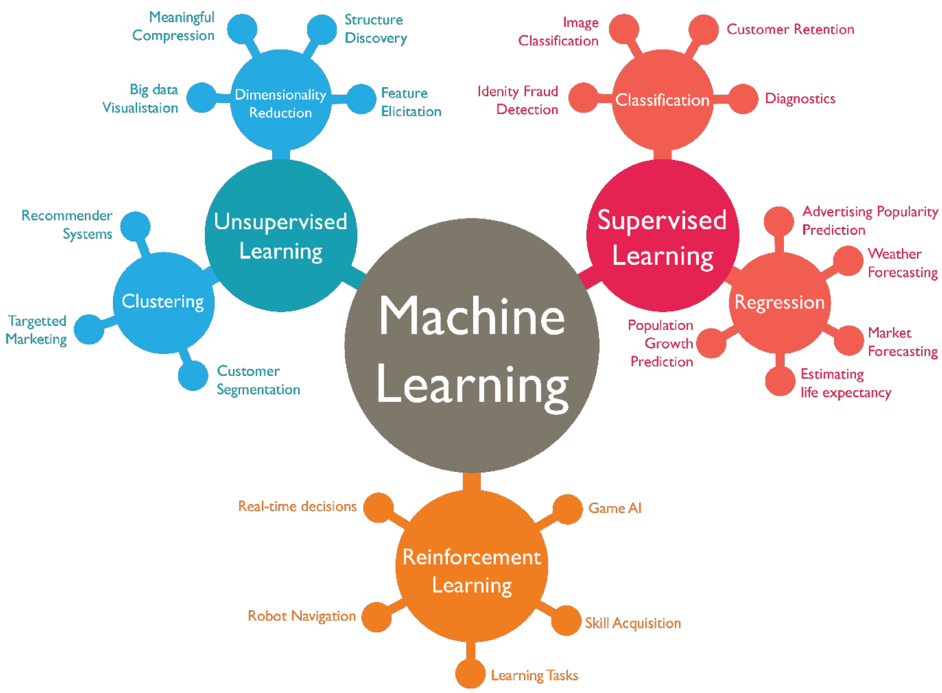



Figure 1.2: Classification of Machine Learning Paradigms.

prices.However, it's important to note that predicting market prices involves inherent uncertainties due to various unpredictable factors. Changes in consumer preferences, unexpected environmental events, regulatory changes, or geopolitical factors can significantly impact market dynamics, making price prediction challenging.Nevertheless, machine learning-based price prediction models offer valuable insights for market participants, traders, aquaculturists, and policymakers. They can assist in making informed decisions regarding production, sales strategies, inventory management, and risk mitigation.The successful deployment of machine learning for mud crab price prediction requires collaboration between data scientists, domain experts, market analysts, and stakeholders involved in the mud crab industry to ensure accurate data collection, model development, and practical application of predictions in real-world market scenarios.

## 1.4 Chapter Summary

In this chapter we discussed mud crab's ecological, cultural, and economic importance. The Mudcrab section explains Scylla serrata's flexibility and critical function in coastal ecosystems, illuminating their fascinating life cycle and significant impact on local economies and global cuisines. The narrative emphasizes conservation and the difficult balance between economic activity and habitat preservation for mud crabs to coexist with their habitats and human populations. In International Demand, the chapter discusses the rising global popularity of mud crabs as a delicacy and the issues of overfishing and habitat degradation. Aquaculture is studied as a way to fulfill rising international demand, with a heavy emphasis on responsible techniques to protect ecological balance and coastal community livelihoods. Machine learning's revolutionary significance in mud crab research is examined in the chapter's final section,

”Machine Learning.” Population monitoring, disease control, and genetic enhancement require large datasets and collaboration between researchers, aquaculturists, and data scientists. The next sub-section, ”Mudcrab Price Prediction using Machine Learning,” discusses how machine learning can predict mud crab prices, acknowledging its inherent complexities but highlighting its value to industry stakeholders for informed decision-making. Overall, the chapter weaves a narrative that spans ecological understanding, commercial dynamics, and technology breakthroughs, particularly machine learning. A balanced strategy, incorporating scientific knowledge, sustainable practices, and technological advances, is emphasized to preserve mud crab species’ survival and coexistence with humans.

# Chapter 2

# Literature Review

The production of mud crab stands out due to the economic significance of the industry and the great demand for it in the market. The production of mud crabs is one of the most renowned examples of aquaculture, which has arisen as an important option to supply the growing demand for seafood all over the world. For the purpose of ensuring a consistent supply of seafood over an extended period of time, it is of utmost significance to employ aquaculture practices that are environmentally responsible. The significance of approaching problems in unique ways is brought into focus by this. This research provides a contribution to the ever-changing environment of aquaculture research by employing machine learning (ML) to evaluate complex data and improve decision-making processes in the rearing of mud crabs. This research was recently published in the journal Aquaculture Research. It was discovered that three of the seven alternative regression models were superior performers after a comprehensive investigation of all of them was carried out. It was via this that I gained an understanding of effective strategies for forecasting Mud Crab's growth and pricing. As an additional point of interest, the novel integration of ensemble models, in particular the proposed combination of Linear Regression and Bayesian Ridge Regression models, demonstrates exceptional performance across a variety of evaluation metrics, thereby showcasing a promising avenue in the field of aquaculture literature. This is an additional point of interest. The greater trend of adopting technology to facilitate decision-making in aquaculture is reflected in the creation of a web interface for real-time forecasts, which is in keeping with this trend. This adds to an enhancement in the practical applicability of the models that have been constructed after being applied to situations that occur in the actual world. When all of these findings are taken into consideration, new and interesting insights into the convergence of aquaculture, machine learning, and decision support systems are provided. Not only do these findings have implications for the management of seafood supplies, but they could also have implications for the production of mud crab in a manner that is environmentally responsible.

Ikhwanuddin et al. [19] two captivating datasets were unearthed after extensive research and meticulous data collection, shedding light on the growth patterns of mud crabs. The first delves into the enigmatic Growth Band Counts (GBC) of wild mud crabs, scrutinizing 31 female and 45 male Scylla

Olivacea specimens from Setiu Wetlands, Malaysia. These crabs were meticulously handled, identified, sexed, tagged, and examined to understand their GBC. Their gastrointestinal mills were carefully extracted, revealing mesocardiac and zygocardiac ossicles – key elements for GBC analysis. After four days, meticulous transverse sections of the gastric mills were prepared, paving the way for deciphering these crabs' growth histories. Meanwhile, the second dataset focuses on the dynamic duo of body weight (BW) and carapace width (CW) increments in juvenile female S. Olivacea. This study, encompassing 135 crabs from the same Malaysian wetlands, meticulously tracked their growth after molting. Each crab's initial BW and CW were documented, followed by a fascinating experiment. Researchers induced leg autotomy – the self-detachment of limbs – to trigger molting. Then, they diligently raised these crabs until their shells had fully hardened seven days after molting. This meticulous approach allowed them to analyze the increment sizes of both BW and CW, as well as the intriguing correlation and regression between these two crucial metrics. Table 2 eloquently summarizes the mean, standard deviation, and the extremes (highest and lowest) of CW and BW increases observed during molting. The data paints a vivid picture of the growth potential within these mud crabs. Further solidifying this intricate relationship, Table 3 reveals a statistically significant (p ¡ 0.001) association between CW and BW increases. This implies that as a crab's carapace expands, its overall body weight tends to follow suit. Finally, Table 4 seals the deal with a regression analysis (p = 0.002) and a noteworthy mean square value of 68.665, conclusively confirming the pivotal role of CW in driving BW increases in S. Olivacea. In essence, these datasets unveil a captivating dance between BW and CW in S. Olivacea, where growth in one metric intricately leads to growth in the other. This intertwined relationship sheds light on the fascinating biology of these mud crabs, offering valuable insights for researchers and conservationists alike.

Bhuiyan et al. [20] even though mud crabs are a highly expensive and well-established industry, Bangladesh does not have any institutional marketing facilities for them. An analysis was conducted to determine the natural mud crab fattening, growing, and harvesting practices of Khulna, as well as the livelihoods of marketing channel actors, stakeholder interests, and crab harvester interests. Participatory rural evaluation methods were used to collect qualitative and quantitative data from the Paikgacha Upazila in Khulna, which is located in the southwestern region of Bangladesh. The data collection process took place between July and December 2019, and involved a total of 200 individuals, 16 stakeholder focus group discussions, and 24 KII professionals. Excel presented qualitative data. SPSS numbers were also provided. Presented are the results. Intermediates belonging to crab, wild, and farm exporters were discovered by the researchers. While marine sources are responsible for 90 percent of the exportable crab, aquaculture only generates 10 percent of it. There were 53% of crab harvesters who made between 188 and 295 dollars per month, 54% of them utilized traditional bait and hooks, and 90% of them gathered during the months of August, September, and October. There were six male crabs and six female crabs in each live weight class. It costs 1100-1650 BDT per kilogram for FF1 females, 750-1100 BDT for Ks1 females, and 1000-1450 BDT for XXL males. The accuracy of the mud crab value chain was 67.3 percent from the point of collection to the point of sale to the foreign client, including the exporter and the aratder. There are a lot of successful corporations that disregard the impact they have on less successful players. Reduce regulations and encourage alternative growing methods in order to retain crab producers.

Sujan et al. [21] we investigated the economic viability and resource efficiency of fattening mud crabs as a source of revenue for disadvantaged coastal Bangladeshis who are confronting the challenges of salt intrusion and climate change. A total of one hundred fifty Bagerhat and Satkhira mud crab farmers were questioned regarding their productivity, returns, and expenses during the months of February and March 2018.The Cobb-Douglas production function was utilized by the researchers in order to evaluate resource efficiency and productivity. The results of their analysis revealed that the costs of fattening mud crabs per acre were a variable of 6104$andaconstantof$4293. In a manner similar to that of coastal shrimp farming, mud crab farming offered a gross return of $10,522, a gross margin of $6229, a net return of $4418 per hectare, and a benefit-cost ratio of 1.72, without any discounting. Feast, bamboo barriers, and crablets all contributed to a 68 percent difference in the fattening of mud crabs.An increase of ten percent in the use of crablets, feed, and bamboo fence might potentially boost the output of mud crabs by five and a half percent, two and a half percent, and one and a half percent, respectively. Disease, crablet quality, budgetary limits, and lack of expertise are some of the challenges that mud crab producers face, which ultimately leads to a breeding failure rate of 57.33 percent. Farmers like mud crab fattening because of its high productivity and resilience to illness, despite the challenges it presents. Profits are increased through training.

Jahan et al. [22] the objective of this study was to conduct an investigation on the implications of live crab harvesting and fattening from an economic standpoint in the southwest region of Bangladesh. This information was gathered from thirty fatteners and twenty collectors who were placed in three different villages.A rough estimate of the costs that are related with accounting the business. While there are only three live gross weight classifications for female crabs (F1, F2, and F3), there are five live gross weight classifications for male crabs (XXL, XL, L, M, and SM). Female crabs are classified as F1, F2, and F3. Consider, for instance, a crab farm that spans an area of one acre.The crab harvests that took place between September and February brought in a total of BDT 416,706 in revenue, whereas the costs associated with fattening the crabs during that time period amounted to BDT 729,619 in expenditures. Between the months of November and February, each acre yields amounted to a total of BDT 864,314 in total. The benefit-to-cost ratio for mud crab that was profitable was found to be 1.5, according to the findings. As a result of the fact that cultivation brought in BDT 13,495 and BDT 179,254 per acre, fattening brought in BDT 134,950, and harvesting brought in BDT 139,510 per season, this was the result. A collaboration between the government and non-governmental organizations (NGOs) is something that the sector is striving for in order to achieve its goals of lowering prices, improving technology, raising awareness, and growing regional markets.

A.A. Laith et al. [23] invertebrates that live in maritime environments have a substantial quantity of bioactive chemicals, which allow for the production of pharmaceuticals that are used in medicinal applications. There are a number of bioactive compounds that are found in crabs, and among these are metabolites that have antiviral, antibacterial, and antifungal properties. Crabs are famous for this content. A wide variety of organs and tissues contain these metabolites in various concentrations. It is worth noting that these organisms possess the capability to act as facultative pathogens, which can have substantial consequences for animal populations that are exposed to stressors such as inadequate circumstances and overpopulation [24].Seventy-five different metabolites were identified by the use of LC/MS-QTOF in the research investigation. Subsequently, these metabolites were filtered down to 19 that satisfied severe requirements, which included a P-Corr value of less than 0.01 and a change of at least two times since the initial analysis.The focus of this literature review is a groundbreaking study that examined the antibacterial activity of entire extracts from two different species of crabs, namely the marine blue swimmer crab (Portunus pelagicus) and the mud crab (Scylla tranquebarica), against fish pathogenic bacteria. The study was conducted in order to determine which crab species had the highest antibacterial activity.

N. Azra et al. [25] research conducted on the maturation diet of broodstock has the impact of improving the quality of the berried females. This, in turn, has an effect on the output of hatcheries and the quantity of larvae produced by species that are farmed. A review of research that has been conducted on mud crab broodstock using specific diets provides light on the influence that maturation diets have on the reproductive performance of the broodstock as well as the quality of the larvae. The research was conducted on mud crabs. The fact that natural, artificial, and mixed diets all have different effects on the nutritional status of the animals makes it difficult to optimize broodstock diets for hatchery productivity. This is because natural diets are more beneficial than artificial diets. It is due to the fact that every single one of them has a different effect on the circumstance. Additionally, the completion of additional research has the potential to improve the diets of crustacean broodstock and to boost hatchery production by filling in knowledge gaps. Both of these outcomes are possible. Both of these outcomes are important in their own right.

Somboonna et al. [26] there is a possibility that mud crabs are resistant to the virus that causes white spot syndrome, according to Scylla serrata, which has made this suggestion. To determine whether or not this assumption is accurate, we investigated the vulnerability of mud crabs that belong to the species Scylla Olivacea and Scylla paramamosa. In order to ascertain the identities of these species, the mitochondrial 16 S ribosomal gene analysis was brought into play. Both S. Olivacea and S. paramamosain were put to WSSV challenges that consisted of a single dosage as well as a series of doses throughout the course of the trial.The difficulties that are caused by WSSV have a different impact on S. Olivacea and S. Paramamosain than they do on other species. When compared to S. Paramamosain, S. Olivacea is more sensitive to the white spot illness than it is to White Spot illness. It is possible that the susceptibility of S. Serrata to white spot disease in the Scylla genus varies from species to species. This is something that needs verification. Based on both single and serial challenges, as well as a recent study that demonstrated little vulnerability, this conclusion has been reached.

Table 2.1: Summary of Related Work's

| Reference | Year | Method | Dataset | Performance | Limitation |
|---|---|---|---|---|---|
| [27] | 2018 | Linear, GRNN, RBF, 3-MLP, 4-MLP, OLM, PLS, RR, EN, QRL, KNN, ELM, ANN-1HL, ANN-a1HL, ANN-2HL, ANN-3HL, RFCIT | AMBIT2 (source-forge.net) | RMSE: 4 layer MLP (0.524), 3 layer MLP (0.538), RBF (0.689), GRNN (0.893), Linear (1.052), RF-CART (0.771) | Silico models are useful for mechanistic understanding, OECD QSAR validation guidelines for machine learners are inexplicit and should be expanded. |
| [28] | 2021 | SVM, NBM, KNN, BPNN, AdaBoost, SVM, CNN, RNN, ANN, RNN, LSTM, DBN, CNN | Online public datasets | Fish Species Classification: CNN, SVM, NBM, KNN, YOLOv3, BPNN, AdaBoost all 90%, Fish Gender Identification: CNN, DT 95%, Fish Age Prediction: SVM: 96.65, Fish Feeding Behavior Analysis: SVM 94.5%, ANN 100%, Fish Group Behavior Analysis: CNN 82.5%, Abnormal Fish Behavior Detection: RNN: 89.89% | Challenges in species recognition and classification in specific environments. |
| [29] | 2018 | CNN | 90 well labelled crab images | Training dataset Classification Accuracy: 98.82, Validation dataset Classification Accuracy: 99.43 | Full cost of machine not explained, time required not specified. |
| [30] | 2020 | KVM, SVM | GitHub PAML Dataset | Only 2 machine learning algorithms used. | |
| [31] | 2021 | SVM, LB | Two 2bRAD data sets: PRJNA338774, PRJNA416847 | SVM = (ASL18, PAL 16) = (0.76, 0.59), LB = (ASL18, PAL16) = (0.77, 0.60) | Neither Bayesian nor ML methods predicted across datasets. |
| [32] | 2021 | (KHVD), (QTL), (SNPs) | simulating phenotypic and their corresponding genotypic datasets | an accuracy of approximately 0.95% | SVM and RF predicted competitively. All evaluated datasets yielded low forecasts, hence DT alone is not recommended. |
| [33] | 2021 | (ARIMA),(ML), (VAR) | time-series datasets | R2 value became 0.75, which is higher inaccuracy | Few climatic variables affecting fish catchment may be identified. |
| [34] | 2020 | NETSO, LOLP | Time-series datasets | A list of NetImbVol values, to which a decomposition of quantiles (5%, 80%, 90% and 9%) is performed. | Few models predict market prices. Its likelihood and adoption would affect market outcome rendering the model meaningless. Deterministic or quasi deterministic market-affecting variables. |

## 2.1 Chapter Summary

In this chapter we discussed about the reviewed literature presents a comprehensive examination of various facets of mud crab cultivation, encompassing economic, environmental, and biochemical perspectives. Bhuiyan et al. [20] reveal the absence of institutional marketing facilities for mud crabs in Bangladesh, offering insights into natural mud crab fattening practices and stakeholder dynamics. Sujan et al. [21] contribute valuable information on the economic viability and resource efficiency of mud crab farming, particularly in the context of challenges faced by coastal communities. Jahan et al. [22] emphasize the profitability of live crab harvesting and fattening, pointing to the need for collaboration between the government and NGOs to achieve sector goals. A.A. Laith et al. [23] delve into the metabolomic content of mud crabs, highlighting their bioactive compounds and potential pharmaceutical applications. Azra et al. [25] review maturation diets for mud crab broodstock, offering insights into optimizing diets for hatchery productivity. Lastly, Somboonna et al. [26] investigate mud crab resistance to the white spot syndrome virus, revealing species-specific susceptibility. Collectively, these studies contribute diverse perspectives that enhance our understanding of mud crab cultivation, providing valuable insights for sustainable practices and informing various stakeholders in the industry.

# Chapter 3

# Methodology

## 3.1 Methods and Materials

This section elucidates the methodology employed in this study, comprising several pivotal stages: data collection, data preprocessing, feature engineering and machine learning model selection. For a visual representation of this methodology, please refer to Figure1. A graphic representation of the suggested technique is shown in Figure1.

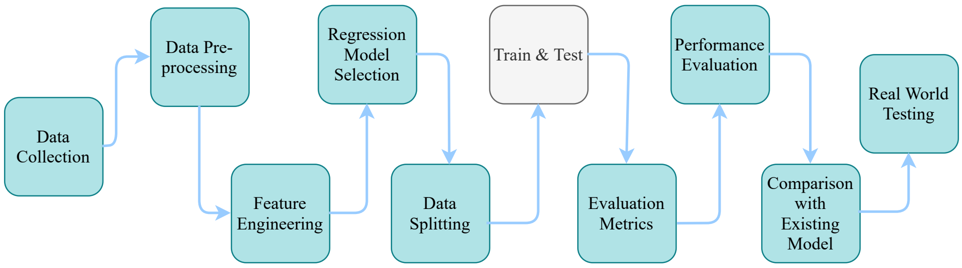



Figure 3.1: Proposed methodology for mud crab growth and price prediction.

### 3.1.1 Data Collection and Description

In our research, we utilized the dataset [19], which was derived from a publication titled ”Dataset on body weight, carapace width increment, and growth band count of mud crabs, Scylla Olivacea.” Our objective was to forecast the size of crabs as they grow and the price of crabs. It is a thorough collection of 136 data points that were painstakingly obtained for the purpose of the study and conservation of the Mud Crab (Scylla Olivacea), which is an important crustacean species that can be found in coastal locations across Southeast Asia. There are essential insights into the species' population dynamic growth and price prediction that could be greatly benefited by applying machine learning approaches in a variety of ways in the future, such as genomic analysis, early warning systems, conservation planning, and so on. This dataset contains essential information such as size, weight, gender distribution, and size of increments, which provide essential insights into the species' population.

Here are some significant features that played a pivotal role in the prediction process, and we thoroughly deliberated upon them.

### Carapace Width of mud crabs

When referring to mud crabs, the term ”carapace width” refers to the measurement of the breadth of the hard upper shell, also known as the carapace, of a mud crab. A portion of the exoskeleton of the crab is known as the carapace, and it is responsible for covering and protecting the front portion of the crab's body, which includes the thorax and the base of its legs. [35].

Research and fisheries management frequently employ the carapace width as a metric for estimating the size and maturation of mud crabs. This is a customary measurement, typically in millimeters, measured from the widest point of the carapace [36]. It is a crucial metric because it can be used to ascertain the crab's age and size, both of which are valuable information for regulating and managing crab populations.

### Growth size of mud crabs

The growth size or increment size in mud crabs refers to the increase in size or dimensions of various body parts, such as carapace width, over a specific period of time or under specific conditions [37]. This measurement is often used to track the growth and development of mud crabs, and it can be a crucial parameter in understanding their life cycle and assessing their health and maturity. It is typically expressed in units of measurement, such as grams, millimeters, centimeters, or inches, depending on the context and the species of mud crab being studied [38].

Important factors that are considered when measuring growth size or increment size in mud crabs include:

- Carapace width is a common measure, reflecting overall size and development.
- Crab growth often correlates with age, vital for tracking development.
- Crab weight offers insights into growth and nutrition.
- In some mud crab species, claw growth is significant and a key metric.
- Water conditions (temperature, salinity), and food availability affect mud crab growth.
- The frequency of crab molting is a crucial growth factor.
- Differences in growth between male and female crabs and the progression from juvenile to mature stages are important considerations.

### Crab Price

The price of mud crabs, often referred to as "Crab price," represents the cost or market value of mud crabs, a type of crab that is highly sought after for its delicious meat. The price of mud crabs can vary depending on factors such as the crab's size, species, geographic origin, seasonality, supply and demand, processing, quality, and the region or market in which they are sold [39]. Prices are typically quoted per unit, such as per kilogram or per pound, and can fluctuate over time due to various market forces and economic conditions. The specific price of mud crabs can vary widely by location and may be influenced by cultural preferences, local regulations, and economic factors.

Considering the cost of production and market dynamics from Selina Wamucii [40], a modest retail price can be expected. To compare this price prediction with body weight increments, studying the correlation between price changes and body weight increments would be an interesting aspect. During October in the Malaysian market, the typical retail price for a kilogram averages at 4.58 USD [41].

1 gram average retail price: $\frac{4.58}{1000}\,\text{USD} = 0.00458\,\text{USD}$

Next, we applied this value as a multiplier to each increment in crab size (g) to derive a new column representing crab prices for our predictive analysis...

## 3.1.2 Data Preprocessing

When it comes to the field of regression machine learning methodologies, data preprocessing is a vital component. This is because it plays a crucial role in ensuring the quality and dependability of prediction models that are being produced. For the purpose of ensuring that the original dataset that we utilized for our research was of a high quality and integrity, we made sure to choose it with a great deal of care. We began the preprocessing stage by determining the rate at which the crabs were developing. This was the first thing that we performed. Due to the fact that there were no missing values or categorical data that needed to be handled, we decided to delete the "No. of crab" column as a consequence of this operation. As a result of this, our dataset is now ready to accommodate the next procedures that will be incorporated into our research.

In the area of estimating crab prices, we introduced a significant component that we called "Crab price," and it was utilized in the context of the projection of crab prices. This characteristic was obtained by multiplying the "Body Weight After (g)" column by the conversion factor 0.00458, which is the retail price that is applicable in Malaysia. This was done in order to achieve this characteristic. This augmentation was carried out at the preprocessing step in order to align our dataset with the specific market conditions that were present at the time. The precision of our price projections has increased as a direct consequence of this.

### 3.1.3 Feature Engineering

In the field of machine learning, feature engineering is a process that is both essential and revolutionary. It has the potential to improve the performance of models and to derive useful insights from data. In order to maximize the potential of a machine learning algorithm to make accurate predictions, it is necessary to construct, modify, or choose features (variables) from raw data in a manner that is both creative and meaningful. By following this process, noise is reduced, dimensionality is decreased, interpretability is enhanced, and the model's ability to recognize detailed patterns and correlations within the data is maximized. It is possible to discover patterns and dependencies with the assistance of a correlation heatmap, which is a visual representation of the strength and direction of interactions between variables in a dataset. In the course of our investigation, we utilized a method known as visual exploration in order to get a full comprehension of the interrelationships that exist between the characteristics of the dataset.

Figure 3.2 provides a visual representation in the form of a correlation heatmap, which vividly illustrates the complex web of connections among the variables. This heatmap offers valuable insights into the strength and direction of correlations between features, enabling us to identify patterns and dependencies critical to our study.

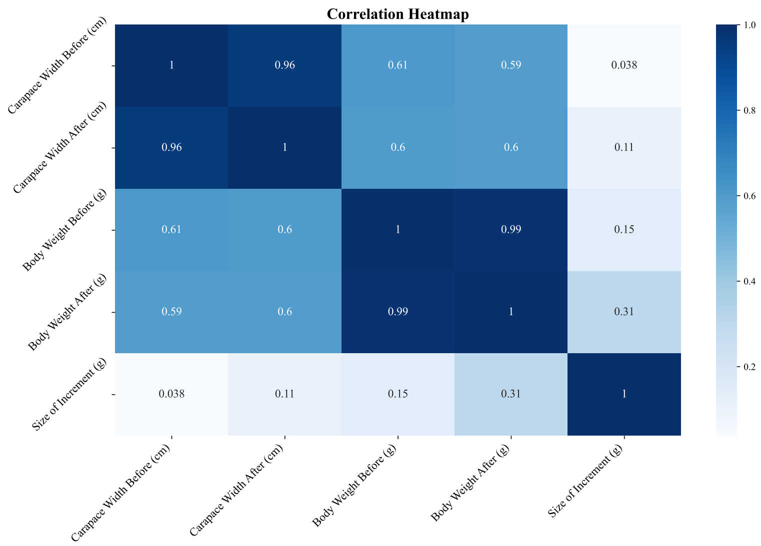



Figure 3.2: Correlation heatmap for all features.

Feature importance quantifies the individual contributions of variables to a machine learning model's predictive performance, facilitating the identification of critical drivers and insights within the dataset.

In Figure 3.3, we present a visual representation of feature importance for crab growth obtained from a Random Forest regression model.

Additionally, Figure 3.4 illustrates the feature importance of crab price, also derived from a Random Forest regression model, shedding light on the factors influencing price predictions. This analysis

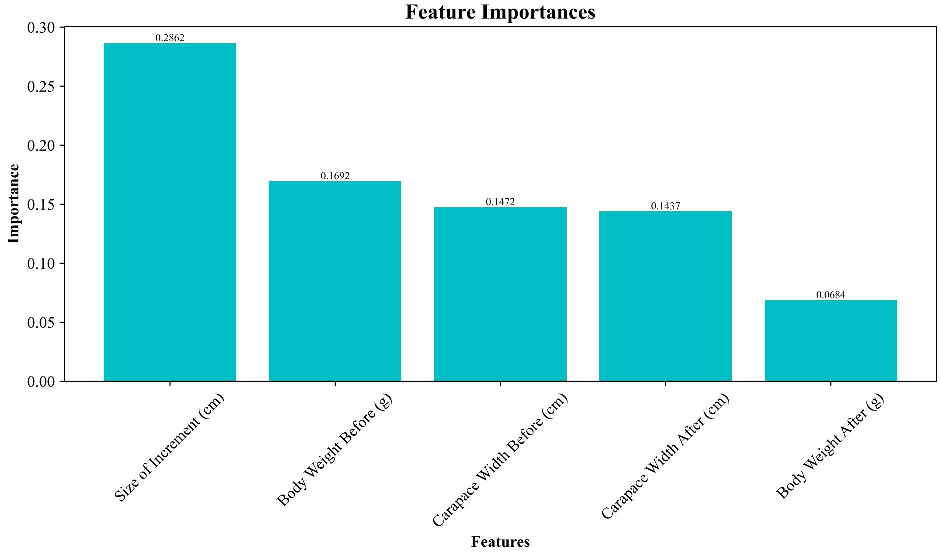



Figure 3.3: List of important features of crab growth identified by Random Forest Regression.

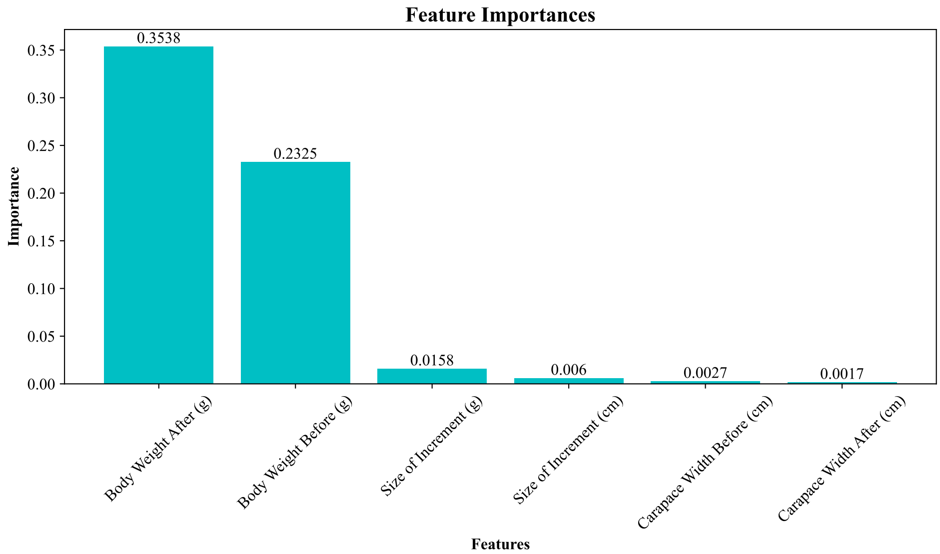



Figure 3.4: List of important features of crab price identified by Random Forest Regression.

provides valuable insights into the relative significance of each feature.

### 3.1.4 Machine Learning Models

After completing the feature engineering phase, the subsequent step involved the careful selection of a machine learning model to predict the growth and price of mud crabs. Our dataset comprises 136 individual mud crab samples, each characterized by its unique growth rate. The price of these crabs is inherently tied to their growth rates. To enhance the accuracy and overall performance of our predictions, we employed various machine learning algorithms, meticulously fine-tuning our models to deliver more precise and reliable results within this comprehensive dataset.

### Random Forest Regression

By applying Random Forest regression to our dataset, we can effectively predict both crab growth and price. The Random Forest regression algorithm is a powerful ensemble learning technique that leverages the strength of multiple decision trees to make accurate predictions. It combines the predictions from each tree, reducing bias and variance.

The Random Forest model [42] is represented as:

$$\hat{Y} = \frac{1}{n}\sum_{i=1}^{n} f_i(X) \tag{3.1}$$

Where:

- $\hat{Y}$ is the predicted crab growth or price.
- $n$ is the number of decision trees in the forest.
- $f_i(X)$ is the prediction made by the $i$-th decision tree for the input features $X$.

### Linear Regression

Using a linear equation to analyze the data that was observed, Linear Regression models aim to model the relationship between a dependent variable, typically represented as $Y$, and one or more independent variables, denoted as $X_1, X_2, \ldots, X_P$, where $P$ is the number of predictors. The linear relationship [43] is represented as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_P X_P + \varepsilon \tag{3.2}$$

Where:

- $Y$ is the predicted crab growth or price.
- $\beta_0$ is the intercept.
- $\beta_1, \beta_2, \ldots, \beta_P$ are the coefficients of the predictors.
- $\varepsilon$ is the error term.

In the context of Linear Regression, the Mean Squared Error (MSE) is used to measure the goodness of fit, which is given by:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{3.3}$$

Where:

- $n$ is the number of observations.
- $Y_i$ is the actual observed value.
- $\hat{Y}_i$ is the predicted value.

**Lasso Regression**

Our experiments demonstrate the effectiveness of Lasso Regression models in predicting crab growth and prices. Lasso Regression not only provides accurate predictions but also highlights the importance of specific predictors, making it a valuable tool for industry stakeholders. Lasso Regression is a linear regression technique that introduces an L1 regularization term to the linear regression equation. The Lasso model [44] aims to minimize the following objective function:

$$\text{minimize}\left(\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2+\lambda\sum_{j=1}^{P}|\beta_j|\right) \tag{3.4}$$

Where:

- $Y_i$ is the actual observed value.
- $\hat{Y}_i$ is the predicted value.
- $n$ is the number of observations.
- $P$ is the number of predictors.
- $\beta_j$ represents the coefficients of the predictors.
- $\lambda$ is the regularization parameter that controls the strength of the L1 regularization term.

The L1 regularization term $\lambda\sum_{j=1}^{P}|\beta_j|$ encourages sparsity in the coefficient estimates, effectively selecting a subset of the most relevant predictors. The first term in the objective function, $\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2$, is the Mean Squared Error (MSE), which measures the goodness of fit of the model. The second term, $\lambda\sum_{j=1}^{P}|\beta_j|$, is the L1 regularization term, which encourages the model to have a sparse set of predictors by penalizing the absolute values of the coefficients $\beta_j$, effectively selecting a subset of the most important predictors.

### Ridge Regression

Ridge Regression proves to be a robust and interpretable tool for predicting crab growth and prices. Ridge Regression is a linear regression technique that introduces an L2 regularization term into the linear regression equation. The Ridge model [44] aims to minimize the following objective function:

$$\text{minimize}\left(\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2+\lambda\sum_{j=1}^{p}\beta_j^2\right) \tag{3.5}$$

Where:

- $Y_i$ is the observed crab growth or price.
- $\hat{Y}_i$ is the predicted value.
- $p$ is the number of predictors.
- $\beta_j$ represents the coefficients for the predictors.
- $\lambda$ is the regularization parameter.

The first term, $\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2$, is the Mean Squared Error (MSE), which measures the goodness of fit of the model. The second term, $\lambda\sum_{j=1}^{p}\beta_j^2$, is the L2 (Ridge) regularization term that encourages stability in the coefficient estimates and helps prevent overfitting.

### Bayesian Ridge Regression

Bayesian Ridge Regression (BRR) is a powerful machine learning algorithm that can be used for both regression and classification tasks. It is a regularized version of linear regression that uses a Bayesian approach to estimate the model parameters. This makes it more robust to overfitting and allows for the incorporation of prior knowledge about the model parameters. The Bayesian Ridge model [45] assumes that the coefficients are random variables following a Gaussian distribution:

$$B\sim\mathcal{N}(0,a^2) \tag{3.6}$$

This formulation embraces model uncertainty, allowing the coefficients to vary probabilistically. The model aims to minimize the following objective function:

$$\text{minimize}\left(\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2+\lambda\sum_{j=1}^{p}\beta_j^2\right) \tag{3.7}$$

Where:

- $Y$ represents observed crab growth or price.

- $\hat{Y}$ is the predicted value.
- $p$ stands for the number of predictors.
- $\beta$ symbolizes the coefficients for the predictors.
- $\lambda$ denotes the regularization parameter.

This Bayesian approach not only provides accurate predictions but also quantifies uncertainty, yielding prediction intervals.

**KNN Regression**

KNN Regression is a non-parametric approach that estimates the target variable $Y$ for a given data point by averaging the target values of its $k$ nearest neighbors [46]. The prediction equation is as follows:

$$\hat{Y}_x = \frac{1}{k}\sum_{i=1}^{k} Y_i \tag{3.8}$$

Where:

- $\hat{Y}$ is the predicted value of the target variable.
- $k$ is the number of nearest neighbors considered.
- $Y_i$ represents the target values of the $k$ nearest neighbors.

This non-parametric approach allows the model to capture intricate relationships within the data without assuming a specific functional form.

## 3.2 Chapter Summary

In this chapter, we have presented the knowledge that is already available in the Automotive Industry, which will give a realistic sense of the possibilities of automotive IOT. This chapter discusses the lessons that were learned regarding the benefits of utilizing the Internet of Things in the manufacturing of automobiles, its impact on the overall car concept, and the applications of IOT in the development of automotive software and connected cars, which can improve overall traffic safety by enhancing communication. After that, it discusses an important factor that should be taken into consideration during the development of automobile software: Vehicle to Vehicle (V2V), Vehicle to Infrastructure (V2I), Vehicle to Pedestrians (V2P), Vehicle to Network (V2N), Car-Device (V2D), Vehicle to Cloud White paper (V2C), and Vehicle-to-Grid (V2G) . The chapter that preceded after this Internet of Things-based predictive maintenance, autonomous vehicles, shared mobility, fleet management and telematics in vehicles, information and entertainment systems, and driver-less cars.

# Chapter 4

# Model Development

subsectionProposed Ensemble Regression Ensemble regression, a formidable technique in machine learning, serves as a cornerstone for enhancing predictive modeling accuracy by amalgamating outputs from multiple individual regression models. At the forefront of this approach is the ”Voting Regressor,” an ensemble learning technique that intricately combines predictions from diverse regression models, thereby culminating in a robust final prediction. In this specialized application, the ensemble integrates two distinctive individual regression models: Linear Regression and Bayesian Ridge Regression. The ensemble is meticulously constructed utilizing the ’VotingRegressor’ class from scikit-learn, where each model contributes its prediction for the target variable based on input features. The Voting Regressor, with its flexible architecture, computes a weighted average of predictions from individual models for the final prediction. These weights are customizable, allowing for the assignment of varying importance to each model. The resultant Voting Regressor combines predictions from its constituent regressors, Linear Regression and Bayesian Ridge Regression [47], through a weighted average as expressed in Equation 4.1:

$$\hat{Y}_{\text{Voting}}(x) = w_{\text{Linear}} \cdot \hat{Y}_{\text{Linear}}(x) + w_{\text{Bayesian}} \cdot \hat{Y}_{\text{Bayesian}}(x) \tag{4.1}$$

Here,

- $\hat{Y}_{\text{Voting}}(x)$ is the predicted crab growth or price by the Voting Regressor.
- $\hat{Y}_{\text{Linear}}(x)$ is the prediction by Linear Regression.
- $\hat{Y}_{\text{Bayesian}}(x)$ is the prediction by Bayesian Ridge Regression.
- $w_{\text{Linear}}$ is the weight for Linear Regression.
- $w_{\text{Bayesian}}$ is the weight for Bayesian Ridge Regression.

This ensemble model strategically leverages the strengths of both constituent regressors, effectively mitigating their individual weaknesses.

### 4.0.1 Model Implementation

The implementation of the proposed ensemble regression model involves a meticulous step-by-step process:

- **Data Preprocessing:** The dataset, encompassing features related to mud crab cultivation, growth, and pricing, undergoes thorough cleaning and preprocessing to address missing values and ensure data uniformity.

- **Feature Selection:** Relevant features influencing crab growth and pricing are thoughtfully selected based on domain knowledge and insights gleaned from exploratory data analysis.

- **Model Training:** Individual regression models, including Linear Regression and Bayesian Ridge Regression, undergo training using the preprocessed data.

- **Ensemble Construction:** The powerful 'VotingRegressor' from scikit-learn is enlisted to craft the ensemble, harmonizing predictions from the adeptly trained Linear Regression and Bayesian Ridge Regression models.

- **Model Evaluation:** Rigorous evaluation of the ensemble model ensues, employing metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and others to gauge its performance.

- **Hyperparameter Tuning:** Fine-tuning of weights assigned to individual models within the ensemble takes place, ensuring optimization of predictive accuracy.

- **Web Interface Development:** In a stride towards practicality, a user-friendly web interface materializes to facilitate real-time predictions for crab growth and price, seamlessly integrating the capabilities of the trained ensemble model.

The resulting ensemble regression model, harmoniously blending the strengths of Linear Regression and Bayesian Ridge Regression, exhibits promising outcomes in the prediction of mud crab growth and pricing. The thoroughness of the model development process underscores its robustness and applicability in real-world scenarios.

## 4.1 Chapter Summary

The culminating chapter on model development unfolds the significance of ensemble regression, particularly the prowess embedded in the Voting Regressor, in elevating predictive accuracy. The amalgamation of two potent regression models, Linear Regression and Bayesian Ridge Regression, fashions a resilient ensemble that adeptly harnesses the strengths of each constituent. The meticulous steps involved in data preprocessing, feature selection, model training, and hyperparameter tuning validate the efficacy of the ensemble in predicting mud crab growth and pricing. The practical implementation of a web interface further amplifies the model's utility, empowering stakeholders in the mud crab industry with real-time predictive insights. This chapter serves as a testament to the transformative potential of machine learning techniques, addressing challenges and fostering sustainability in the realm of aquaculture on a profound scale.

# Chapter 5

# Result

In this part of the article, we will provide an explanation of the experimental study, which includes providing specifics on the environment in which the models were deployed, as well as the evaluation criteria that were applied to establish the effectiveness of the models. After that, we carry out a full review of the data of the experiment, with the purpose of offering vital insights into the success of the suggested strategy, the influence of a range of scenarios on its performance, and the implications that it has in the real world.

### 5.0.1 Experimental Setup

For the purpose of this research study, the cloud-based platform known as Google Colab was utilized as the primary instrument for the development, installation, and training of our machine learning models. This technology was utilized for the purpose of this research study. This particular research study made use of this technology in order to accomplish its goals. Specifically, this research study utilized this technology in order to achieve the objectives that it set out to achieve. In order to guarantee that the project would be effective, it was essential to carry out this activity in this particular place. Our ability to carry out a wide variety of tests without the need for a substantial number of significant processing resources was made possible by the fact that the free edition of Google Colab provided us with access to a powerful and widely accessible computational ecosystem. Due to this, we were able to carry out a wide range of different tests. Consequently, we were able to carry out a significant number of tests of varying kinds as a result of this. We were able to conduct out a significant number of tests as a consequence of this, which allowed us to achieve a result that was beneficial.

We were able to improve the overall efficiency of our research attempts and raise the degree of communication that occurs among the members of our team due to the fact that we took advantage of the collaborative capabilities that Google Colab provides. Because of the exploitation of these qualities, which made it possible for them to do so, the members of our research team were able to interchange code and datasets in a manner that was fully seamless. This was made possible by the fact that they were able to do so. The entirety of the software code was meticulously constructed in Python3, and

Scikit-Learn was the major machine learning tool that was utilized in our research and findings over the course of this project. Python3, the programming language that was utilized throughout the entire procedure, was employed to program the software's instructions.

### 5.0.2 Evaluation Matrices

Predictions were generated for two significant variables associated with the mud crab dataset utilizing these models. This action was taken subsequent to the model being trained utilizing 80% of the dataset. The dataset indicates that these variables were identified. Specifically, the "size of increment (g)" and the "price of crab" were the variables that were considered in relation to this specific circumstance. A subset of twenty percent of the remaining dataset was selected for the purpose of testing, or putting these models to the test. The degree of inaccuracy present in the predictions generated by a regression model can be utilized as a metric to assess the model's performance or capabilities. This is a feasible course of action. This is a potentiality that may be implemented. To assess and convey the performance of an extensive range of models, the subsequent metrics were employed, with detailed explanations of each one provided subsequently:

**Mean Absolute Error (MAE)**

Mean Absolute Error (MAE) is a commonly used metric to assess the accuracy of a regression model's predictions. It offers a clear-cut and comprehensible indicator of the predicted accuracy of the model. It quantifies the average absolute difference between the model's predictions and the actual values. Smaller MAE values indicate a more accurate model. Mean Absolute Error (MAE) is calculated using the following equation:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (10)$$

- MAE represents the Mean Absolute Error. - $n$ is the total number of data points. - $y_i$ represents the actual values. - $\hat{y}_i$ represents the predicted values.

Where: MAE is the Mean Absolute Error, $n$ is the total number of data points, $y_i$ represents the actual target value for the $i$-th data point, $\hat{y}_i$ represents the predicted value for the $i$-th data point.

The histogram depicted in Figure 5(a) showcases the Mean Absolute Error (MAE) concerning the 'Size of Increment (g)' as computed by various Machine Learning models. The outcomes reveal that the linear regression, Bayesian Ridge regression, and ensemble model algorithms demonstrate notably superior performance. In Figure 6(a), the histogram presents the Mean Absolute Error (MAE) in relation to the 'Crab price,' as assessed across different Machine Learning models. The findings highlight exceptional performance from all six model algorithms, with the exception of the lasso regression, which exhibits a comparatively higher error of approximately 0.05.

**Mean Squared Error (MSE)**

Mean Squared Error (MSE) is another fundamental metric for evaluating the performance of regression models. It quantifies the average squared differences between predicted values and actual values. MSE is advantageous because it penalizes larger errors more than smaller ones due to the squaring operation. Smaller MSE values indicate a more accurate model, and it's particularly useful when you want to give more weight to larger errors. Mean Squared Error (MSE) is computed using the following equation:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (11)$$

- MSE represents the Mean Squared Error. - $n$ is the total number of data points. - $y_i$ represents the actual target value for the $i$-th data point. - $\hat{y}_i$ represents the predicted value for the $i$-th data point.

Where: MSE is the Mean Squared Error, $n$ is the total number of data points, $y_i$ represents the actual target value for the $i$-th data point, $\hat{y}_i$ represents the predicted value for the $i$-th data point.

In Figure 5(b), the histogram visually represents the Mean Squared Error (MSE) with regard to the 'Size of Increment (g)' calculated across various Machine Learning models. The results underscore the remarkable performance of the linear regression, Ridge regression, Bayesian Ridge regression, and ensemble model algorithms. Moving to Figure 6(b), a parallel histogram offers insights into the Mean Squared Error (MSE) concerning the 'Crab price,' as evaluated among different Machine Learning models. These findings affirm that all seven model algorithms consistently demonstrate superior performance in this context.

**Root Mean Square Error (RMSE)**

Root Mean Square Error (RMSE) is a widely used metric for assessing the accuracy of regression models, similar to Mean Squared Error (MSE). RMSE represents the square root of the average squared differences between predicted and actual values. Smaller RMSE values indicate a more accurate model, and it is particularly useful when you want to quantify the typical size of prediction errors. Root Mean Square Error (RMSE) is calculated using the following equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (12)$$

- RMSE represents the Root Mean Square Error. - $n$ is the total number of data points. - $y_i$ represents the actual target value for the $i$-th data point. - $\hat{y}_i$ represents the predicted value for the $i$-th data point.

In Figure 5(c), the histogram provides a visual representation of the Root Mean Squared Error (RMSE) concerning the 'Size of Increment (g)' across a range of Machine Learning models. The results underscore the exceptional performance of the linear regression, Bayesian Ridge regression, and ensemble model algorithms in this context. Transitioning to Figure 6(c), the histogram sheds light on the Root Mean Squared Error (RMSE) in relation to the 'Crab price,' as evaluated across diverse Machine

Learning models. The findings also point to the noteworthy superior performance of the linear regression, Ridge regression, Bayesian Ridge regression, and ensemble model algorithms.

**R-squared (R2 )**

R-squared ($R^2$) is a statistical measure that represents the proportion of the variance in the dependent variable (target) that is explained by the independent variables (features) in a regression model. R-squared calculates the degree of model fit and generally falls within the interval of 0 to 1. A higher $R^2$ value indicates that a larger proportion of the variance in the target variable is explained by the model. An $R^2$ value of 1 indicates a perfect fit, where the model predicts the target variable without error, while an $R^2$ value of 0 means that the model does not explain any of the variance in the target variable. $R^2$ is calculated using the following equation:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (13)$$

- $R^2$ represents the R-squared value. - $n$ is the total number of data points. - $y_i$ represents the actual target value for the $i$-th data point. - $\hat{y}_i$ represents the predicted value for the $i$-th data point. - $\bar{y}$ represents the mean of the actual target values.

Within Figure 5(d), the histogram visually presents the $R^2$ values for the 'Size of Increment (g)' across a spectrum of Machine Learning models. The findings highlight the exceptional performance of the linear regression, Ridge regression, Bayesian Ridge regression, and ensemble model algorithms, achieving a perfect fit with $R^2$ values of 1. In contrast, the $R^2$ results for random forest, lasso regression, and KNN regression are notably lower, approximately 0.29, 0.93, and 0.58, respectively. Turning our attention to Figure 6(d), the histogram provides a glimpse into the $R^2$ concerning the 'Crab price,' as assessed across various Machine Learning models. The results affirm that five of the model algorithms exhibit superior performance, with KNN regression standing out as particularly impressive with a value of 0.99. Conversely, lasso regression underperforms, yielding a less favorable result of approximately 0.28.

**Root Mean Squared Logarithmic Error (RMSLE)**

Root Mean Squared Logarithmic Error (RMSLE) is a metric often used to evaluate the accuracy of regression models, especially when dealing with data that exhibits a wide range of values. RMSLE measures the average logarithmic difference between predicted and actual values. A model with a lower RMSLE value is considered more accurate. It's particularly useful for applications where the relative errors between predicted and actual values matter more than absolute errors. RMSLE is calculated using the following equation:

$$\text{RMSLE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(1+y_i) - \log(1+\hat{y}_i))^2} \quad (14)$$

- RMSLE represents the Root Mean Squared Logarithmic Error. - $n$ is the total number of data points. - $y_i$ represents the actual target value for the $i$-th data point. - $\hat{y}_i$ represents the predicted value for the $i$-th data point.

Figure 5(e) presents a histogram that visually illustrates the RMSLE in the context of the 'Size of Increment (g)' across a variety of Machine Learning models. The results underscore the excellent performance of five model algorithms, with the exception of random forest and KNN regression, which exhibit comparatively less favorable results. Transitioning to Figure 6(e), the histogram provides valuable insights into the RMSLE regarding the 'Crab price,' evaluated across a spectrum of different Machine Learning models. The findings reveal that all seven model algorithms consistently demonstrate superior performance in this context.

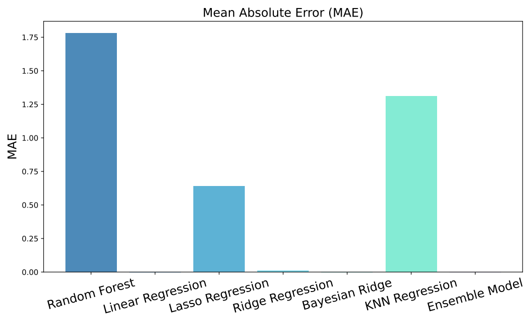
**Mean Absolute Percentage Error (MAPE)**

Mean Absolute Percentage Error (MAPE) is a widely used metric for assessing the accuracy of forecasts or predictions, particularly in cases where relative errors are more important than absolute errors. MAPE is expressed as a percentage, making it easy to interpret. It measures the average magnitude of errors as a percentage of the actual values. Smaller MAPE values indicate a more accurate model. MAPE is calculated using the following equation:

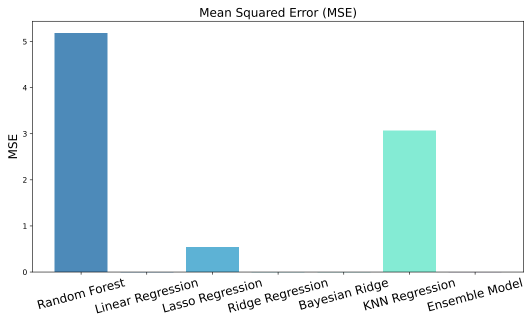
$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|y_i - \hat{y}_i|}{|y_i|}\right) \times 100\% \quad (14)$$

- MAPE represents the Mean Absolute Percentage Error. - $n$ is the total number of data points. - $y_i$ represents the actual target value for the $i$-th data point. - $\hat{y}_i$ represents the predicted value for the $i$-th data point.

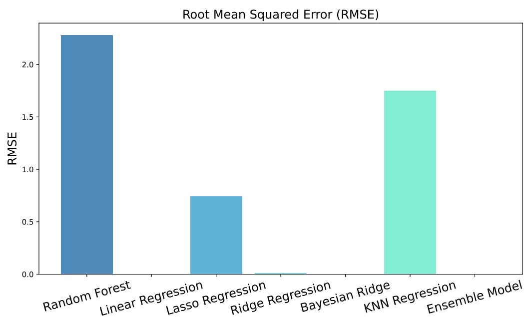
Within Figure 5(f), the histogram visually portrays the Mean Absolute Percentage Error (MAPE) with regard to the 'Size of Increment (g)' for a variety of Machine Learning models. The results highlight the notably superior performance of the linear regression, Bayesian Ridge regression, and ensemble model algorithms in this context. Transitioning to Figure 6(f), the histogram sheds light on the Mean Absolute Percentage Error (MAPE) in relation to the 'Crab price,' as evaluated across a variety of Machine Learning models. The results underscore the notably superior performance of the linear regression, Ridge regression, Bayesian Ridge regression, and ensemble model algorithms, with all achieving exceptional accuracy as reflected in the low MAPE values of 0.00.



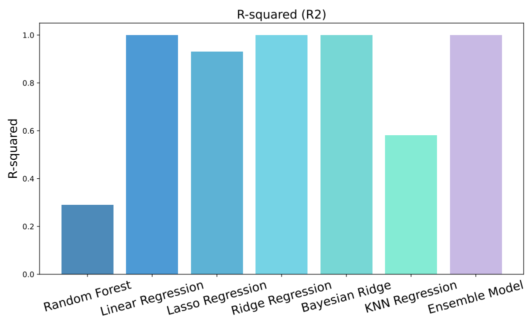

(a) MAE



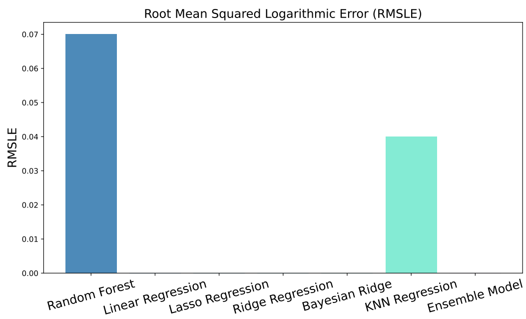

(b) MSE



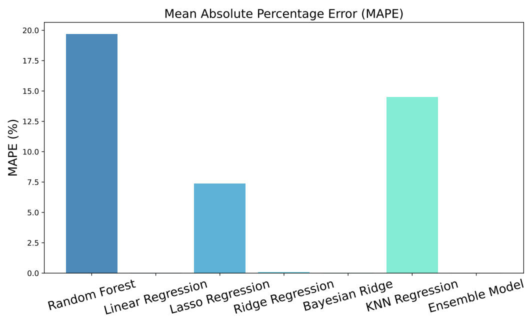

(c) RMSE




(d) R-squared

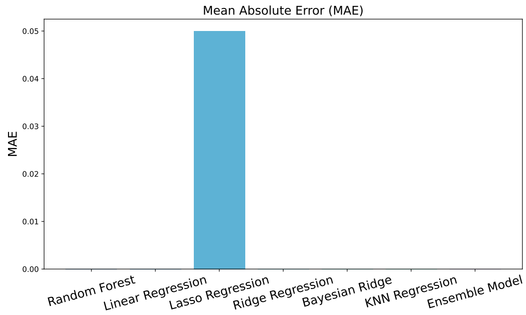



(e) RMSLE

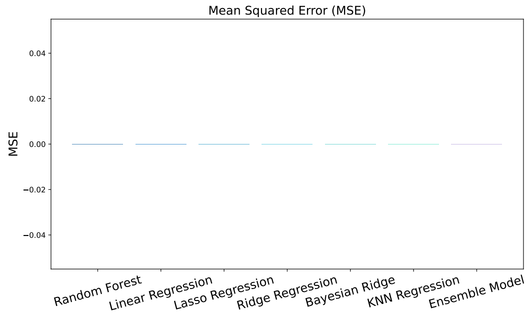



(f) MAPE

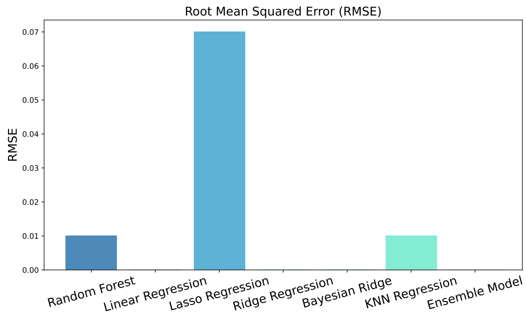
Figure 5.1: Assessment of the performance (in terms of the 'Size of Increment (g)') of diverse Machine Learning regression models through the utilization of multiple evaluation metrics, including MAE, MSE, RMSE, R-SQUARED, RMSLE, and MAPE.



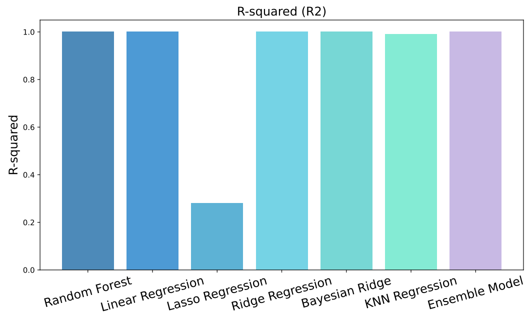

(a) MAE



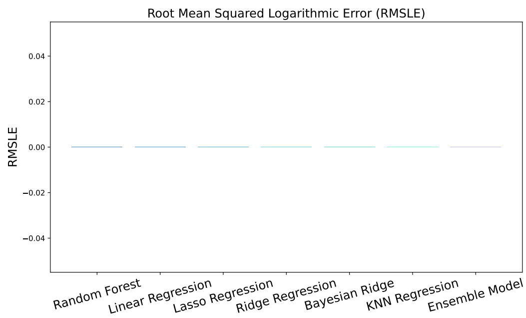

(b) MSE



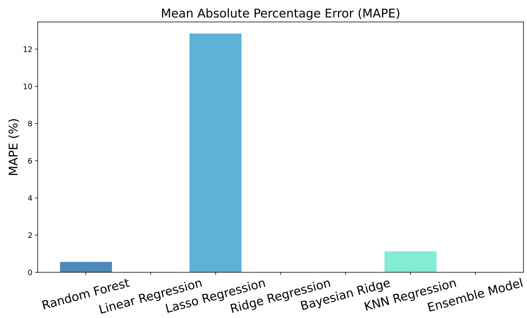

(c) RMSE




(d) R-squared




(e) RMSLE




(f) MAPE

Figure 5.2: Assessment of the performance (in terms of the 'Crab price') of diverse Machine Learning regression models through the utilization of multiple evaluation metrics, including MAE, MSE, RMSE, R-SQUARED, RMSLE, and MAPE.

### 5.0.3 Hyper-Parameters Tuning

Following the completion of the process of selecting the model and the establishment of the experimental setup, more hyperparameter refining was carried out. As a result of this improvement, the accuracy was significantly increased in comparison to the initial default parameter settings.

Table 5.1: Hyperparameter Settings of Different Machine Learning Models

| **Model Name** | **Hyperparameter Settings** |
|---|---|
| Random Forest Regression | `n_estimators`: [10, 50, 100, 150, 200]<br>`max_depth`: [None, 10, 20, 30, 40]<br>`min_samples_split`: [2, 5, 10]<br>`min_samples_leaf`: [1, 2, 4]<br>`random_state = 42`<br>`cv = 5` |
| Linear Regression | `'fit_intercept' (default=True)`<br>`'normalize' (default=False)`<br>`cv = 5` |
| Lasso Regression | `'alpha' (default=1.0)` |
| Ridge Regression | `'alpha' (default=1.0)` |
| Bayesian Ridge Regression | `'n_iter' (default=300)`<br>`'tol' (default=1e-3)`<br>`'alpha 1' (default=1e-6)`<br>`'alpha 2' (default=1e-6)`<br>`'lambda 1' (default=1e-6)`<br>`'lambda 2' (default=1e-6)`<br>`'compute_score' (default=False)`<br>`'fit_intercept' (default=True)` |
| KNN Regression | `n_neighbors = 5`<br>`cv = 5` |
| Ensemble Model | `estimators = [ ]` |

The Random Forest Regression hyperparameters determine the behavior of the model. `n_estimators` sets the number of decision trees in the ensemble, with common choices like 10, 50, 100, 150, or 200. `max_depth` controls the maximum depth of each tree (e.g., `None`, 10, 20, 30, or 40). `min_samples_split` specifies the minimum samples required to split a node (e.g., 2, 5, or 10), while `min_samples_leaf` sets the minimum samples for a leaf node (e.g., 1, 2, or 4). `random_state = 42` ensures reproducibility, and `cv = 5` is the number of folds in cross-validation. These hyperparameters allow customization to find the right balance between model complexity and generalization for a given regression task.

Linear Regression hyperparameters include `fit_intercept` (default: `True`), determining whether to calculate the intercept; `normalize` (default: `False`), indicating if features should be normalized; and `cv = 5`, specifying the cross-validation fold count. These settings help customize the model's behavior, handling intercept, feature scaling, and evaluation strategy.

Lasso and Ridge Regression hyperparameters include `alpha` (default: 1.0), controlling the regularization strength. A higher alpha value increases regularization, reducing model complexity.

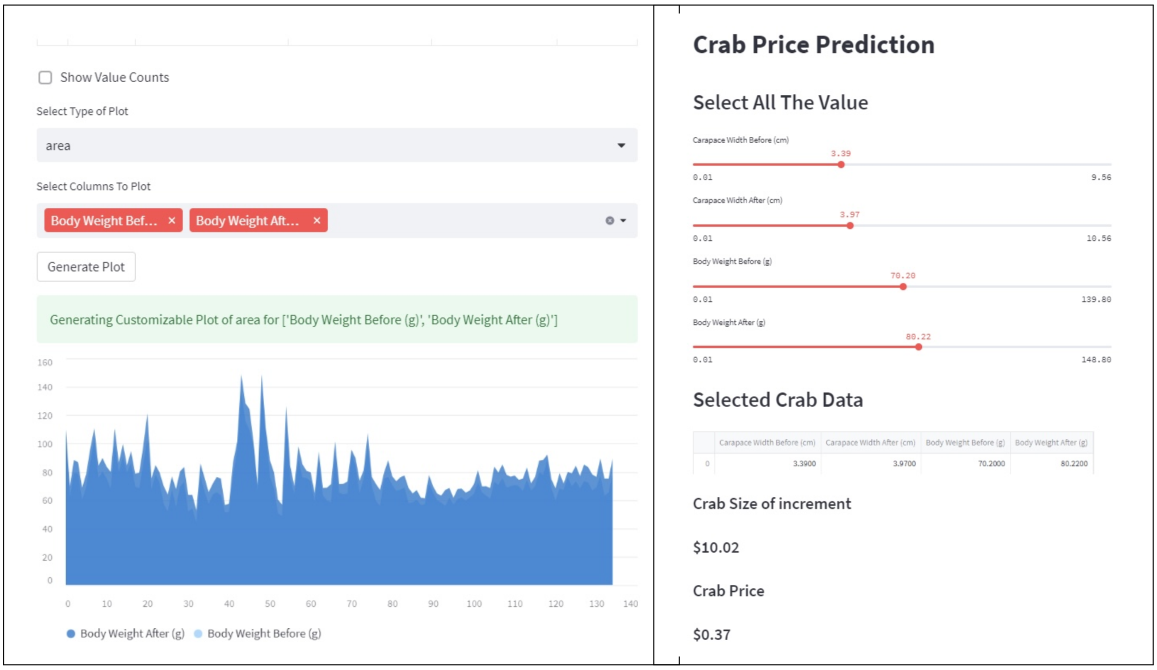
Bayesian Ridge Regression offers a set of hyperparameters to fine-tune the model. It involves `n_iter` (default: 300) for controlling the number of optimization iterations, `tol` (default: $1 \times 10^{-3}$) to define the convergence tolerance. There are also hyperparameters related to regularization, including `alpha_1` and `alpha_2` (both default: $1 \times 10^{-6}$), influencing the precision of model weights and noise. Additionally, `lambda_1` and `lambda_2` (default: $1 \times 10^{-6}$) are hyperparameters governing the regularization strength. Lastly, `fit_intercept` (default: `True`) decides whether the model should fit an intercept term during training. These hyperparameters provide flexibility in adapting Bayesian Ridge Regression to various data scenarios and regularization requirements.

K-Nearest Neighbors (KNN) Regression relies on `n_neighbors` (default: 5), setting the number of neighbors for predictions and influencing model sensitivity. `cv = 5` is used for cross-validation folds, enabling the trade-off between prediction accuracy and model flexibility.

In an ensemble model, the `estimators` hyperparameter in `VotingRegressor`, with a default value of an empty list, allows users to specify a list of model names and instances for the ensemble. By default, the ensemble combines the predictions of models like Linear Regression and Bayesian Ridge using equal, uniform weights for each model, resulting in an equally weighted ensemble prediction.

### 5.0.4 Web Interface Implementation

We created this website for global reach of crab prize. We also created this website for information sharing, online presence, business promotion, credibility and at last brand building etc. This website helps many people to get information about crab price. Figure8 (a)In picture (a) we have selected value for our crab model. Parameter name as Carapace Width Before (cm), Carapace Width After (cm), Body Weight Before (g), Body Weight After (g). This value is shown as Selected Crab Data. After selecting our value our two model is showing crab price and increment size.8(b)In this section people can chose EDA that shows shape, columns, summary, selected columns, value counts, correlation and pie plot. Here is just a brief view about our EDA process that shows summary.8(c)On this section we can see no of crab and crab price visualization. we can clearly see that its shows positive linear regression on price set. This section is called plots section and with this section anyone can visualize to understand dataset better. This (d) customized feature allows us to draw and find patterns between datasets. We can see that the body weight after is more than the body weight before. Customers can easily detect the relation between features of the dataset and work as necessary.




(a) Visualization

(b) Selecting value to predict price and growth

Figure 5.3: Streamlit crab price and growth prediction.

### 5.0.5 Results

In Figures 6 and 7, a comprehensive examination of diverse evaluation metrics has been undertaken to discern the most efficacious models for predicting crab growth and crab prices. In the context of crab growth prediction Figure 6, three models emerge as optimal performers: Linear Regression, Bayesian Ridge Regression, and the Ensemble Model. Transitioning to crab price prediction Figure 7, a quartet of models, namely Linear Regression, Ridge Regression, Bayesian Ridge Regression, and the Ensemble Model, demonstrate superior predictive capabilities. These findings encapsulate the meticulous assessment of multiple metrics, highlighting the nuanced strengths of each model in the distinct realms of crab growth and pricing prognostication. Table III In the realm of predicting crab growth rates, three top-performing models have been identified.

Table 5.2: Comparison of Mean Cross-Validation RMSE Values for Top 3 Models in Crab Growth Rate Prediction.

| **Model Name** | **Mean Cross-Validation RMSE** |
|---|---|
| Linear Regression | $1.61 \times 10^{-14}$ |
| Bayesian Ridge Regression | $4.90 \times 10^{-10}$ |
| Ensemble Model | $3.45 \times 10^{-10}$ |

To discern the most suitable model, an evaluation based on the mean cross-validation RMSE scores is conducted. The table below presents a comparative analysis of these models to determine their relative efficiency: Through this evaluation, we aim to identify the model that demonstrates the highest level of performance and efficiency in predicting crab growth rates. Table IV, In the domain of predicting crab prices, four standout models have been identified as the most promising. To ascertain the most suitable model, we turn our attention to an evaluation grounded in the mean crossvalidation RMSE scores. Below is a tabulated comparison of these models, shedding light on their relative efficacy: This assessment aims to unveil the model that exhibits the highest level of performance and suitability for accurately predicting crab prices.

Through this evaluation, we aim to identify the model that demonstrates the highest level of performance and efficiency in predicting crab growth rates.

Table II,In the domain of predicting crab prices, four standout models have been identified as the most promising. To ascertain the most suitable model, we turn our attention to an evaluation grounded in the mean cross-validation RMSE scores. Below is a tabulated comparison of these models, shedding light on their relative efficacy:

Table 5.3: Comparison of Mean Cross-Validation RMSE Values for Top 3 Models in Crab Price Prediction

| **Model Name** | **Mean Cross-Validation RMSE** |
|---|---|
| Linear Regression | $1.0619720555400591 \times 10^{-16}$ |
| Ridge Regression | $7.4189946604826 \times 10^{-6}$ |
| Bayesian Ridge Regression | $4.5677155706423115 \times 10^{-8}$ |
| Ensemble Model | $2.283857786059634 \times 10^{-8}$ |

This assessment aims to unveil the model that exhibits the highest level of performance and suitability for accurately predicting crab prices.

## 5.1 Discussion

There is multiple regressor model we run for better result. Among those algorithms best accuracy is shown by Linear regression and Bayesian ridge regression also ensemble model. Ensemble learning combines the outputs of the individual models to obtain a final prediction. Hence this approach has been shown to improve overall accuracy while reducing the computational cost compared to traditional techniques. This model demonstrates superior predictive capabilities. We also find out Mean Cross-Validation RMSE Values for Top 3 Models in Crab Growth Rate Prediction and Crab Price Prediction. We used multivariate analysis and we predicted similar result. We also created a website with 4 parameter that takes value and gave it to linear regression model and our model gives price and growth estimate. We can also do EDA and visualize our dataset to get further idea.

Despite the promising results achieved in this study, there is still room for further improvements in the proposed method. For example, we can use deep learning method. ANN model can give a lot of precise accuracy. There also Transfer Learning and many other deep learning formulas we can apply and check if this gives better performance or not. While deep learning models have shown impressive results in many applications, their black-box nature limits their interpretability. Therefore, exploring methods to extract meaningful information from the features learned by the proposed architecture can help to build trust and understanding of the model's decisionmaking process. Beside on the website we can used sign up and login method to save further information. Website customization option could be great choice for customer. we can use API to Automatically change website price and it should be changed according to each day price. People can also have a reminder option to buy crab at low cost if they select certain option. We can Rapid advertisement this website in sea area. We can also make premium subscription to send crab price rate to our premium user. We can also make e commerce website to sell crab directly to our customer, also can make user friendly navigation system. Attractive and consistent design also another option. High quality live image, website customization, CTA, Social Media Integration also be a good choice for customization. We can use the CIA Triad for enhance security and keep regular update website and a feedback and testing for general type of people and analytics section for each of people also be a finest choice for our website.

## 5.2 Chapter Summary

In this chapter, we discussed the results of our research on smart car interior designs. The experimental analysis included details about the environment in which the machine learning models were implemented and the assessment criteria used to evaluate their effectiveness. The experimental setup utilized the Google Colab platform for development, installation, and training of machine learning models. The collaborative features of Google Colab facilitated efficient teamwork and code/data exchange among team members. Python3 and Scikit-Learn were the primary tools employed for coding. The evaluation matrices used for assessing the models included Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), R-squared ($R^2$), Root Mean Squared Logarithmic Error (RMSLE), and Mean Absolute Percentage Error (MAPE). The hyperparameters tuning process was conducted to refine the models' performance, enhancing accuracy compared to default settings. A table provided details of hyperparameter settings for different machine learning models, including Random Forest Regression, Linear Regression, Lasso Regression, Ridge Regression, Bayesian Ridge Regression, KNN Regression, and Ensemble Model. The web interface implementation involved creating a website for global reach, information sharing, online presence, business promotion, credibility, and brand building. The website included a section for selecting values to predict crab prices and growth, an exploratory data analysis (EDA) section for visualizing dataset characteristics, and plots for visualizing the number of crabs and crab prices. Customized features allowed users to draw and find patterns between datasets. The results section presented a comprehensive examination of evaluation metrics for diverse machine learning regression models. Figures and tables showcased the performance of models in predicting crab growth rates and prices. The top-performing models were identified based on mean cross-validation RMSE scores. The chapter concluded with a discussion on potential improvements, such as exploring deep learning methods like Artificial Neural Networks (ANN), Transfer Learning, and other deep learning formulas. Suggestions for website enhancements, security measures, user-friendly features, and business strategies were also provided.

# Chapter 6

# Conclusion

Concerns about the environment and the market have been growing, which has been the impetus for the transition from conventional methods to sustainable methods in the aquaculture of mud crabs. According to the conclusion, the absence of early measurements and norms that are pertinent to sustainability has brought to light the importance of competent management, which has therefore promoted research that draws from a number of fields. These challenges have been successfully addressed by the fields of data science, economics, aquaculture management, and biology. This has enabled these fields to successfully face the challenges that are presented by restricted growth models and market instability. For the purpose of achieving resilience and sustainability, the growth of mud crabs in aquaculture demands the participation of various individuals working together. This is necessary in order to achieve the desired results. Over the past few years, researchers have been able to make more precise predictions regarding the prices of mud crabs as a result of technical advancements that have made it possible for them to do so. In conclusion, the aquaculture business conducts an analysis of growth patterns, environmental considerations, and economic ramifications in order to maximize profitability and ensure environmental conservation when it comes to aquaculture. In an effort to encourage environmentally responsible behaviors, this is done. It has been shown that the exploitation of mud crab farming has a good influence on both the economy and the safety of food supplies in regions that are not located on the shore. This observation was made in locations that are not located on the coast. There is evidence that mud crab farming is both economical and sustainable, as demonstrated by the examination of the value chain, economic models, and resource efficiency. To sum everything up, this is the situation. Research, collaboration, and innovation are three essential components that must be incorporated into the process in order to effectively solve challenges and safeguard the mud crab industry. As a consequence of this, it is possible to establish a healthy sector that has the potential to be of assistance to later generations.

# References


[1] M. SUJAN, “Profitability and efficiency of mud crab fattening in bangladesh,” Ph.D. dissertation, DEPARTMENT OF DEVELOPMENT AND POVERTY STUDIES, SHER-E-BANGLA AGRICULTURAL . . . , 2018.

[2] M. Molla, M. Islam, S. Islam, and M. Salam, “Socio-economic status of crab collectors and fatteners in the southwest region of bangladesh,” *Journal of the Bangladesh Agricultural University*, vol. 7, no. 452-2016-35781, 2009.

[3] M. S. Islam, “Status of mud crab aquaculture in bangladesh,” in *Proceedings of the International Seminar-Workshop on Mud Crab Aquaculture and Fisheries Management, 10-12 April 2013, Tamil Nadu, India*. Rajiv Gandhi Centre for Aquaculture (MPEDA), 2015, pp. 1–6.

[4] Y. C. Thampi Sam Raj, A. Mandal, G. Kumaran, A. Sethuramalingam, P. Srinivasan, and J. Kumar, “Status of mud crab industry in india,” in *Proceedings of the International Seminar-Workshop on Mud Crab Aquaculture and Fisheries Management, 10-12 April 2013, Tamil Nadu, India*. Rajiv Gandhi Centre for Aquaculture (MPEDA), 2015, pp. 17–26.

[5] M. Intelligence. (No date) Crab market size & share analysis - growth trends & forecasts (2023 - 2028). Accessed on 9-28-23. [Online]. Available: https://www.mordorintelligence.com/industry-reports/crab-market

[6] D. T. K. Paul, “Cost-benefit analysis of crab farming in sundarbans areas of west bengal - a study,” *Journal of Emerging Technologies and Innovative Research (JETIR)*, vol. 9, no. 5, p. j563, May 2022, © 2022 JETIR. [Online]. Available: http://www.jetir.org

[7] Asriani, M. Y. Karim, and H. Y. Azis, “Study of mud crab (scylla olivacea) growth which cultivated in silvofishery system in various types of mangrove vegetation,” *International Journal of Scientific and Research Publications (IJSRP)*, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:92176490

[8] M. H. Sakib, S. Ahmmed, Y. Mahmud, and M. L. Islam, “Nursery rearing of orange mud crab, *Scylla olivacea* (herbst, 1896): Optimizing pond habitat and stocking density,” *Journal of*

*Bangladesh Agricultural University*, vol. 20, no. 1, pp. 97–102, 2022, research Article. [Online]. Available: http://baures.bau.edu.bd/jbau

[9] Q. Wu, K. Waiho, Z. Huang, S. Li, H. Zheng, Y. Zhang, M. Ikhwanuddin, F. Lin, and H. Ma, "Growth performance and biochemical composition dynamics of ovary, hepatopancreas and muscle tissues at different ovarian maturation stages of female mud crab, scylla paramamosain," *Aquaculture*, vol. 515, p. 734560, 2020.

[10] L. C. Naegel, "Combined production of fish and plants in recirculating water," *Aquaculture*, vol. 10, no. 1, pp. 17–24, 1977.

[11] M. Saaid, N. Fadhil, M. M. Ali, and M. Noor, "Automated indoor aquaponic cultivation technique," in *2013 IEEE 3rd international conference on system engineering and technology*. IEEE, 2013, pp. 285–289.

[12] H. Motoh, D. de la Peña, M. Dimaano, and E. Tampos, "Preliminary report on the first production of the crab stages of alimasag, *Portunus pelagicus* (linnaeus)," Aquaculture Department, Southeast Asian Fisheries Development Center, SEAFDEC/AQD Institutional Report 2(3), 1978, downloaded from http://repository.seafdec.org.ph, SEAFDEC/AQD's Institutional Repository. [Online]. Available: http://hdl.handle.net/10862/2326

[13] F. S. Ghory, *Developmental Stages of Crabs Collected from Northern Arabian Sea*. LAP LAMBERT Academic Publishing, November 2018.

[14] C. Shelley and A. Lovatelli, "Mud crab aquaculture: a practical manual," *FAO Fisheries and aquaculture technical paper*, no. 567, p. I, 2011.

[15] S. Moser, D. Macintosh, S. Pripanapong, and N. Tongdee, "Estimated growth of the mud crab scylla olivacea in the ranong mangrove ecosystem, thailand, based on a tagging and recapture study," *Marine and Freshwater Research*, vol. 53, no. 7, pp. 1083–1089, 2002.

[16] B. K. Chakraborty and S. A. Azad, "Present status of mud crab population in bangladesh," *International Journal of Aquaculture and Fishery Sciences*, March 2018.

[17] C. Rumisha, "Genetic diversity and gene flow among the giant mud crabs (*Scylla serrata*) in anthropogenic-polluted mangroves of mainland tanzania: Implications for conservation," *Fisheries Research*, vol. 205, April 2018.

[18] W. Wang, C. Ma, W. Chen, Z. Jin, M. Zhao, F. Zhang, Z. Liu, and L. Ma, "Population genetic diversity of mud crab (scylla paramamosain) from southeast coastal regions of china based on mitochondrial coi gene sequence," *Gene*, vol. 751, p. 144763, 2020.

[19] M. Ikhwanuddin, A. Amin-Safwan, N. Hasyima-Ismail, and M. N. Azra, "Dataset on body weight, carapace width increment and growth band count of mud crabs, scylla olivacea," *Data in brief*, vol. 26, p. 104477, 2019.

[20] M. S. Bhuiyan, M. M. Shamsuzzaman, M. M. Hossain, S. J. Mitu, and M. M. H. Mozumder, "Mud crab (scylla serrata forsskal 1775) value chain analysis in the khulna region of bangladesh," *Aquaculture and Fisheries*, vol. 6, no. 3, pp. 330–336, 2021.

[21] M. H. K. Sujan, M. M. H. Kazal, M. S. Ali, and M. S. Rahman, "Cost-benefit analysis of mud crab fattening in coastal areas of bangladesh," *Aquaculture Reports*, vol. 19, p. 100612, 2021.

[22] H. Jahan, M. S. Islam, and S. Islam, "Economic performance of live crab (scylla serrata) business in the southwest coastal region of bangladesh," *International Journal of Fisheries and Aquatic Studies*, vol. 4, no. 1, pp. 453–457, 2016.

[23] A. Laith, M. Ambak, A. Abol-Munafi, W. Nurhafizah, and M. Najiah, "Metabolomic analysis of marine and mud crabs based on antibacterial activity," *Aquaculture Reports*, vol. 7, pp. 7–15, 2017.

[24] C. Sindermann, "Disease in marine aquaculture," *Helgoländer Meeresuntersuchungen*, vol. 37, pp. 505–530, 1984.

[25] M. N. Azra and M. Ikhwanuddin, "A review of maturation diets for mud crab genus scylla broodstock: Present research, problems and future perspective," *Saudi Journal of Biological Sciences*, vol. 23, no. 2, pp. 257–267, 2016.

[26] N. Somboonna, S. Mangkalanan, A. Udompetcharaporn, C. Krittanai, K. Sritunyalucksana, and T. Flegel, "Mud crab susceptibility to disease from white spot syndrome virus is species-dependent," *BMC research notes*, vol. 3, pp. 1–12, 2010.

[27] T. H. Miller, M. D. Gallidabino, J. I. MacRae, S. F. Owen, N. R. Bury, and L. P. Barron, "Prediction of bioconcentration factors in fish and invertebrates using machine learning," *Science of The Total Environment*, vol. 648, pp. 80–89, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969718330869

[28] S. Zhao, S. Zhang, J. Liu, H. Wang, J. Zhu, D. Li, and R. Zhao, "Application of machine learning in intelligent fish aquaculture: A review," *Aquaculture*, vol. 540, p. 736724, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0044848621003860

[29] D. Wang, M. Holmes, R. Vinson, G. Seibel, and Y. Tao, "Machine vision guided robotics for blue crab disassembly: Deep learning based crab morphology segmentation," in *2018 ASABE Annual International Meeting*. St. Joseph, Michigan: American Society of Agricultural and Biological Engineers, 2018. [Online]. Available: https://elib.mnakeabibtexrary.asabe.org/abstract.asp?JID=5AID=49279CID=det2018T=1

[30] C. Yang, X. Su, D. Liu, Z. Guo, F. Wang, and Y. Lu, "A new method of aquatic animal personality analysis based on machine learning (paml): Taking swimming crab portunus trituberculatus as an example," *Frontiers in Marine Science*, vol. 7, p. 32, 2020.

[31] L. Bargelloni, O. Tassiello, M. Babbucci, S. Ferraresso, R. Franch, L. Montanucci, and P. Carnier, "Data imputation and machine learning improve association analysis and genomic prediction for resistance to fish photobacteriosis in the gilthead sea bream," *Aquaculture Reports*, vol. 20, p. 100661, 2021.

[32] C. Palaiokostas, "Predicting for disease resistance in aquaculture species using machine learning models," *Aquaculture Reports*, vol. 20, p. 100660, 2021.

[33] 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*, vol. 13, no. 16, p. 9124, 2021.

[34] A. Lucas, K. Pegios, E. Kotsakis, and D. Clarke, "Price forecasting for the balancing energy market using machine-learning regression," *Energies*, vol. 13, no. 20, p. 5420, 2020.

[35] A. Mohapatra, R. K. Mohanty, S. K. Mohanty, and S. Dey, "Carapace width and weight relationships, condition factor, relative condition factor and gonado-somatic index (gsi) of mud crabs (scylla spp.) from chilika lagoon, india." 2010.

[36] F. Megat, S. Kamaruddin, M. Ikhwanuddin, M. Ambak *et al.*, "Study on carapace width growth band counts relationship of orange mud crab, scylla olivacea (herbst, 1796) from terengganu coastal waters, malaysia." *Pakistan journal of biological sciences: PJBS*, vol. 20, no. 3, pp. 140–146, 2017.

[37] T. Dahril, Y. Harjoyudanto *et al.*, "Study of growth and survival of mud crab (scylla serrata, forskal) with different salinity levels in culture media," in *IOP Conference Series: Earth and Environmental Science*, vol. 934, no. 1. IOP Publishing, 2021, p. 012015.

[38] P. P. GABITO, F. C. SENTINA, and N. D. LUNA, "The growth capacity of mud crab (scylla serrata) in cages under mangove and at open campocpoc bay."

[39] F. Zannatul, X. Zhang, M. R. Hasan *et al.*, "Mud crab (scylla sp.) marketing system in bangladesh." *Asian Journal of Food and Agro-Industry*, vol. 3, no. 2, pp. 248–265, 2010.

[40] S. Wamucii. (2023) Crab prices in malaysia. Accessed: November 9, 2023. [Online]. Available: https://www.selinawamucii.com/insights/prices/malaysia/crabs/#retail-prices

[41] ——, "Crab prices in malaysia," https://www.selinawamucii.com/insights/prices/malaysia/crabs/retailprices, 2023, accessed: November 9, 2023.

[42] A. Saini, "An introduction to random forest algorithm for beginners," *Data Science Blogathon*, August 2022, published as part of the Data Science Blogathon. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/10/an-introduction-to-random-forest-algorithm-for-beginners/

[43] A. Name, "Linear regression using mllib," *Analytics Vidhya*, June 2022. [Online]. Available: https://www.analyticsvidhya.com/blog/2022/06/linear-regression-using-mlib/

[44] ——, "Ridge and lasso regression: A complete tutorial with python," *Analytics Vidhya*, January 2016. [Online]. Available: https://www.analyticsvidhya.com/blog/2016/01/ridge-lasso-regression-python-complete-tutorial/

[45] ——, "Different types of regression models," *Analytics Vidhya*, January 2022. [Online]. Available: https://www.analyticsvidhya.com/blog/2022/01/different-types-of-regression-models/

[46] ——, "Introduction to k-nearest neighbors for regression in python," *Analytics Vidhya*, August 2018. [Online]. Available: https://www.analyticsvidhya.com/blog/2018/08/k-nearest-neighbor-introduction-regression-python/

[47] V. Kotu and B. D. PhD, *Predictive Analytics and Data Mining*. ScineceDirect, 2015.