Dissertation on

AI for Aquaculture in India

Submitted in partial fulfilment of the requirements for the award of degree of

Bachelor of Technology
in
Computer Science & Engineering

UE17CS490B – Capstone Project Phase - 2

Submitted by:

Sujay Gad **PES1201700177**
Mayur R B **PES1201700714**
Priyadith T P **PES1201701324**

Under the guidance of

Prof.Uma D
Professor
PES University

January - May 2021

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF ENGINEERING
PES UNIVERSITY
(Established under Karnataka Act No. 16 of 2013)
100ft Ring Road, Bengaluru – 560 085, Karnataka, India

PES UNIVERSITY

(Established under Karnataka Act No. 16 of 2013)
100ft Ring Road, Bengaluru – 560 085, Karnataka, India

FACULTY OF ENGINEERING

CERTIFICATE

This is to certify that the dissertation entitled

AI for Aquaculture in India

is a bonafide work carried out by

Sujay Gad	**PES1201700177**
Mayur R B	**PES1201700714**
Priyadith T P	**PES1201701324**

in partial fulfilment for the completion of seventh semester Capstone Project Phase - 2 (UE17CS490B) in the Program of Study - Bachelor of Technology in Computer Science and Engineering under rules and regulations of PES University, Bengaluru during the period Jan. 2021 – May. 2021.It is certified that all corrections / suggestions indicated for internal assessment have been incorporated in the report. The dissertation has been approved as it satisfies the 8th semester academic requirements in respect of project work.

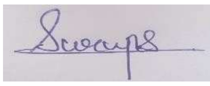
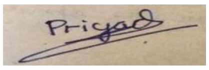
Signature	Signature	Signature
Dr. Uma D	**Dr. Shylaja S S**	**Dr. B K Keshavan**
Professor	Chairperson	Dean of Faculty

External Viva

Name of the Examiners	**Signature with Date**
1._______________________	_______________________
2._______________________	_______________________

DECLARATION

We hereby declare that the Capstone Project Phase - 2 entitled **AI in Aquaculture in India** has been carried out by us under the guidance of Prof. Uma D, Professor and submitted in partial fulfilment of the course requirements for the award of degree of **Bachelor of Technology** in **Computer Science and Engineering** of **PES University, Bengaluru** during the academic semester January – May 2021. The matter embodied in this report has not been submitted to any other university or institution for the award of any degree.

PES1201700177	Sujay Gad	
PES1201700714	Mayur R B	
PES1201701324	Priyadith T P	

ACKNOWLEDGEMENT

I would like to express my gratitude to Prof. Uma D, Department of Computer Science and Engineering, PES University, for her continuous guidance, assistance, and encouragement throughout the development of this UE17CS490B - Capstone Project Phase – 2.

I am grateful to the project coordinators Prof. Silviya Nancy, Prof. Sunitha for organizing, managing, and helping with the entire process.

I take this opportunity to thank Dr. Shylaja S S, Chairperson, Department of Computer Science and Engineering, PES University, for all the knowledge and support I have received from the department. I would like to thank Dr. B.K. Keshavan, Dean of Faculty, PES University for his help.

I am deeply grateful to Dr. M. R. Doreswamy, Chancellor, PES University, Prof. Jawahar Doreswamy, Pro Chancellor – PES University, Dr. Suryaprasad J, Vice- Chancellor, PES University for providing to me various opportunities and enlightenment every step of the way. Finally, this project could not have been completed without the continual support and encouragement I have received from my family and friends.

ABSTRACT

Aquaculture is one of the fastest growing industries in India in recent times. In this model, artificial intelligence is used to improve both quality and quantity of aquaculture production in India. Using AI to suggest species and category of aquaculture based on capital investment, climate, availability of resources as well as demand for that particular type. Time series analysis will be used to forecast future demand for aquaculture in global as well as local markets and predict prices of aquaculture products in India.

Supervised machine learning models are used in the project to detect presence and spread of diseases among aquatic organisms. Suggestion of feeding intake based on stage of production and category of organism. Suggestion of suitable methods and technologies is going to be done which can be used to improve aquaculture practices in India in the future.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Increase in population has meant greater demand for seafood. 85% of marine stocks are fully exploited or overfished hence the importance of aquaculture has grown in the last 30 odd years. In India from 1980 till 2018 aquaculture production has grown by 1834%. There is growing demand for aquaculture products in both local as well as global markets. All this points to the fact that aquaculture is going to play a key role in the years to come.

Quality of aquaculture production in India has to improve in order to keep up with the rising demand. Only 16% of freshwater area and 10% of brackish water area is currently being utilized for aquaculture hence there is massive scope for improvement in this regard.

Average productivity of freshwater aquaculture has been 2.2 tonnes/Ha and average productivity of brackish water has been 10 tonnes/Ha. Both these values can be increased up to 10 tonnes/Ha.

In India quantity of production can be increased by spreading awareness about aquaculture as an occupation, profits which can be made as well as providing help in terms of resources needed such as seeds, initial capital investment. Local institutes such as marketing societies can be set up to help small businesses. Transportation as well as accessible market places should be made available to the fish farmers.

Quality of production can be improved by improving quali ty of seeds, shifting towards modern techniques of aquaculture production, detecting and preventing spread of diseases among aquatic organisms. Policy support, public investment in infrastructure as well as research programs can further enhance growth of aquaculture sector in India. All this must be carried out by keeping environmental impact in mind.

Aim and objective is to help fish farmers improve their profits, provide employment opportunities to local youth in this sector and also contribute to economy by increasing aquaculture's contribution to GDP through increase in export activities

CHAPTER 2

PROBLEM STATEMENT

Aim of the project is to improve quality and quantity of aquaculture production in India. Aquaculture production quantity and prices of aquaculture products are forecasted using time series analysis in R. Various aquatic species are suggested to farmers based on factors such as weather, geographical area, weather and local demand. Spread of disease is detected using supervised machine learning algorithms such as decision tree, KNN etc. In case of modern aquaculture techniques ideal conditions such as water pH, temperature are also suggested to the fish farmers. At the end points are points with regards to ways in which aquaculture production can be improved further in India.

Aquaculture production quantity can be increased by encouraging more public to be take up Aquaculture. This is can be done by spreading awareness about increasing trend in prices of aquaculture products, export of aquaculture products and the benefits it will bring along in the long run. Detecting spread of disease in Aquaculture farms at an early stage will result in improving quality and quantity of Aquaculture production in India. Suggesting appropriate species based on local conditions as well as implementing suggestions given by experts can also be used in ordr to improve production in the long run.

CHAPTER 3

LITERATURE SURVEY

In this chapter, we present the current knowledge of the area and review substantial findings that help shape, inform and reform our study.

3.1 Break-even analysis and profitability of aquaculture in India

3.1.1 Introduction

Between ten to fifteen percent of global Fish diversity resides in India.India is ranked second among all countries of the world in annual fish production of 9.06 metric tonnes.India has a large coastline of 8129km and the coastal aquaculture is practiced in 50,000 hectares in the states of Karnataka, Andhra Pradesh, Kerala, West Bengal and Goa. Potential use of Aquatic resources using appropriate methods can improve the social and economic status of the coastal rural people.

3.1.2 Materials and Methods

Data from all possible sources had been collected and studied. The entire cost in aquaculture is that the combination of annual fixed charge and operating cost. Brake even analysis is a situation where the aquaculture system is neither making money or losing money, but all the costs have been covered. Other indicators include annual overhead cost net profit generated and annual operating cost.

3.1.3 Results and conclusions

This paper is based on scope and feasibility of an integrated approach in aquaculture in India. Author says that the aquaculture is one of the fastest food growing farming systems and has enormous scope for future development. Only a meagre 13% of the potential area is available for aquaculture is currently in use, so there is huge scope for improvement.

Shrimp culture:

This research paper states that white shrimp is worse than tiger shrimp in terms of overall profit even though there is very little difference in capital investment and infrastructure facilities required for both.

Integration of paddy and shrimp farming:

Integration farming with crop, livestock has turned out to be very rewarding with excellent production using minimum level of input. Integrated farming is currently being carried out in

Bheries of West Bengal, Pokkali fields of Kerala, Khar lands of coastal Karnataka during rainy season etc

Crab Culture:

The demand for live crabs in foreign markets has resulted in large scale growing of mud crabs in confinements. Crab culture has picked up in recent years in Kerala, Tamil Nadu and province. Minimum initial requirements mean it's efficient sort of aquaculture economically. The evaluation of technical-economical aspects of varied aquaculture practices reveals that there's ample scope for improving productivity of aquaculture in India. The ever-increasing export demands in addition to high price within the international market means benefits are endless. The paper also says that there's a requirement for legal framework for regulation of all kinds of aquaculture. Aquaculture also generates employment and must be dispensed by keeping economic sustanability within the mind.

3.2 Aquaculture development in India: an economic overview

3.2.1 Introduction

This paper talks about fisheries as an emerging sector in India, it emphasizes the role of fisheries in expanding exports and adding to the foreign exchange, in which the sector has exceeded expectations, it also talks about the growth, various sources of growth, contribution to GDP, socio-economic impacts, economic impacts. The paper concludes with an optimistic note for the development of Aquaculture with the need to streamline government policies to achieve better results in production and marketing.

3.2.2 Characteristics and Implementation

The paper contains details of inland and marine fish production from 1980 to 1999, the dataset also contains the contributions made by aquaculture to the Indian GDP, we calculate the different contributions made by different states towards aquaculture, the dataset contains the statistics on the different farms i.e. the type of farm, the land that it is built upon, total area it is built upon. It contains the different international markets for export.

3.2.3 Evaluation

Various parameters were taken into consideration, the paper highlights the various impacts like socio-economic impact, impact on the nation's GDP, economic impact on that region, environmental impact. The paper highlights the problems faced by the farmers like financial constraints, lack of support, logistics. Finally, the paper Gives the possible solutions to overcome these challenges.

3.2.4 Summary

The paper concludes with an optimistic note, aquaculture has a lot of scope for development, but there are various problems plaguing this sector, which if isn't solved it would hinder the path of progress for aquaculture, the paper gives a few solutions to solve these problems, with support from central and state governments.

3.3 Growth of aquaculture sector in India :Policy Directions For Future

3.3.1 Introduction

In this section,the author discusses about the fish production the world and in India. He calls the fishery sector in India as the sunrise sector. The paper gives the statistics about the fish production in World and India for the 50 years in terms of million tonnes and discusses about the contribution of this sector to the Indian economy and also talks about the growth of the fish industry.In the end ,the paper discusses about the changes in the fishery sector in different states of India.

3.3.2 Statistics and Challenges

The fishing sector includes marine fisheries,freshwater and brakish water aquaculture and inland fisheries using ponds and reservoirs. The share of Developed countries in the production of fish has decreased from 57% in 1973 to 23% in 1997 and the share of developing countries has increased from 43% in 1973 to 73% in 1997.The contribution of the fishery towards the country Gross domestic product (GDP) is increasing at a swift rate while that of agriculture sector is continually reducing.

3.3.3 Structure and Trend

This paper talks about the structure and trend of fishery industry in India.The fisheries sector production in India has shown eye-catching growth post independence.It jumped from a meagre 0.75 million tonnes in 1951 to 6.57 million tonnes in 2005-06.Over a period of 50 years the contribution of inland sector has dominated the marine production sector whose share fell down to 43 percent.The increasing pollution of the water resources,energy crisis and the increasing cost of fishing have induced an increasing realization of the potential and flexibility of aquaculture as a continuous and economical substitute to capture fishes.

The compound growth rate in fish production from 1980-81 to 2005-06 is as follows. In the year 1980-1990 the total compound growth rate was 4.39 percent and marine contribution was 3.80 percent and inland contribution was 5.28 percent.In the year 1990-2000 the total compound growth rate was 4.13 percent out of which marine contribution was 2.33 percent and inland contribution was 6.55 percent.In the year 2000-2006 ,the total compound growth was 2.75 percent of which marine sector contribution was negative -0.21 percent and inland contribution was 5.37 percent. During the pre-WTO period,the growth rate had a surge and this was due to the marine sector.The growth was about 4.78 percent as compared to post WTO period.In Post-WTO period the inland sector showed a better growth as compared to marine sector because of the improved public and private investment for the inland fishery sector through distinct development plans and research by the government of India.

In India, there has been a linear trend in the fish production. Aquaculture production is expected to play a crucial role in forthcoming decades in compensating for stagnant capture fisheries and in meeting increased demand for aquatic production. Some of major challenges of aquaculture sector to retain the current growth rate include:

1. Contribution made by fishery sector must be maintained for the benefit of security,employment and recreation.
2. As the sector expands, aquaculture production should be controlled from diseases.
3. Marine sector in aquaculture would face an energy plight in future.

3.3.4 Conclusion

In this paper, the author recognizes fisheries as powerful income employment generator as well as foreign exchange. The paper discusses the challenges aquaculture must overcome like disease spreading among creatures, availability of fish meat and fish oil etc to continue the growth shown in the last 30 odd years.

3.4 Comparison of Aquaculture practices in West Bengal and Andhra Pradesh

3.4.1 Introduction

Carp culture has proven to be a sustainable culture, freshwater aquaculture mainly depends on carp culture, this paper compares the socio-economic conditions and different practices followed by two leading fish producers i.e. Andhra Pradesh and west Bengal, it also highlights the different problems faced by the farmers of two states.

3.4.2 Characteristics and Implementation

An interview-based survey was taken by the researchers in seven districts of Andhra Pradesh and 13 districts of West Bengal, the survey covered 153 farms, the list of these Farms were provided by the Fishery Extension Officers of the Department of fisheries, Government of Andhra Pradesh and Government of West Bengal and also the officials of Krishi Vigyan Kendra.

3.4.3 Evaluation

The paper reveals that majority of farmers sought information on Aquaculture through other farmers, neighbours, they received little or no help from the government or non-government organisations, farmers of both states primarily follow intensive style of aqua-culture, majority of Andhra Pradesh farmers sourced fish seeds from other hatcheries, whereas West Bengal farmers sourced fish seeds from government centres, disease was the common problem faced by farmers of two states.

The importance of aquaculture is well known and has been demonstrated in various countries and states, aquaculture is bound to develop and progress ahead in the coming Years. However, assistance from the government, proper training, financial assistance and Emphasis on importance of using technology in aquaculture needs to be done quickly.

3.5 Disease control in Aquaculture and the responsible use of veterinary drugs and vaccines

3.5.1 Introduction

Aquatic diseases is that the one in every of the restrictions to the expansion and management of aquaculture worldwide. It's one in every of the fastest growing food production sector within the world. The aquatic diseases will be of two types. They're Infectious diseases and non-infectious diseases.

The infectious diseases are caused by viruses, bacteria and parasites and this diseases are capable of spreading through the movement of infected species.The non-infectious diseases are caused by ubiquitious opportunistic agents or other biotic and abiotic conditions.

3.5.2 Measures to combat diseases

The farmers have greatly enhanced their disease prevention awareness. National and international rules and regulation have been formed for the benefit of minimizing the danger associated with aquatic animals getting any disease during their movements.

3.5.3 Disease causes and treatment

To determine the cause for a disease there are various techniques and these techniques could range from gross observation to highly technical biomolecular based tools. The molecular based techniques provide quick results.They are also used for non-lethal sampling.

In the treatment of diseases,most countries still rely on Anti-Bacterial agents. To reduce the disease caused by infection agent prophylactic druga are used.Some of the methods used for prevention of disease include control of movement of animals onto the farm, destruction of clinically sick animals, emergency harvest of clinically healthy animals, sanitary measures. Major problems associated with chemotherapeutant are discussed.

3.5.4 Vaccination and future opportunities

Vaccination is one of the major ways to control spread of disease in aquatic species. While the vaccination is effective against certain diseases, it is still impracticable to vaccinate against shrimp and molluscan pathogens. Since antibiotics residuals are harmful, antibiotics requirement is reduced by the usage of vaccines. Vaccination leaves no residues within the product or environment and don't induce pathogen resistance. Some samples of vaccination also are being discussed.

3.5.5 Conclusion

Author suggests that aquatic animal health programmes can be introduced which will provide guidelines with respect to strict hygiene practices,sanitation standards as well as site selection and general layout of farm.Diseases can also be prevented by control of movement of animals onto the farm,destruction of clinically sick animals and disinfection at regular intervals.

Author suggests that aquatic animal health programmes can be introduced which can provide guidelines with relevancy strict hygiene practices, sanitation standards further as site selection and general layout of farm. Diseases may be prevented by control of movement of animals onto the farm,destruction of clinically sick animals and disinfection at regular intervals.

3.6 A novel approach to fish disease diagnostic system based on machine learning

3.6.1 Introduction

Fish disease is one of the foremost obstacle within the cultivation sector resulting in loss of over ten billion to the Indian agricultural export price. Here it speaks about the EUS, which is a serious disease in many countries including India. The disease irruption typically happens throughout the winter. The temperature may be a essential issue for fish mortality.

Modern Machine Learning technique which combines PCA-FAST-NN is used to combat EUS disease. This helps fish farmers to find out about this disease in quick time.

3.6.2 Methodology

Initially, they created the database by extracting the image features from the real images of the fish and stored them in a database. For any fish image taken as input,the features will be extracted and identified images populated in database will be selected as the output.

It conjointly tells regarding the bar chart of destined gradients(HOG),which may be a feature descriptor and that is employed for object detection. It uses gradient orientation in localized parts of a picture. It's a method that discovers shapes within the image and extract feature for all location within the image or region of interest.

Next, the paper talks regarding PCA.PCA may be a method for locating correlations between options in information. It's a linear technique for spatial property reduction. It reduces the dimension by eliminating the non-useful data. It's ofttimes employed in machine learning and image process technique. FAST may be a corner detection rule that relies on the intensity that minimizes the process of knowledge and it's conjointly employed in image process. It detects the interest purpose within the image.

Then the author dicusses regarding the neural networks. Neural networks are a series of algorithms that mimic the operations of a human brain to recognize the relationship between vast amount of data. It explains about the connections and the working of neural network in the human brain. The nerve cells are called as neurons. Some of the learning strategies discussed in the paper are supervised learning, unsupervised learning and reinforcement learning.

3.6.3 Proposed Methodology

The input image is taken and then the feature extractor FAST is applied which detects the interest point and after that the dimension is reduced by the PCA and then Neural network is applied which is a classifier and training it for detecting the fish disease and then getting the accuracy.

The steps involved in proposed methodology are discussed, which are Image acquision, image pre-processing, feature extraction and recognition and the last step is classification.

3.6.4 Experimentation and Conclusions

In order to detect EUS disease,combination of FAST-PCA-NN is preferred over PCA-HOG-NN since it gives better accuracy results. Finally the author has achieved associate degree accuracy of eighty six percent for FAST-PCA-NN combination.

3.7 Responsible Aquaculture in 2050

As aquaculture is gaining more ground, we must confirm to avoid mistakes made during intensification of agriculture. so as to style a responsible aquaculture System we must understand environmental impacts and therefore the steps to mitigate them.

3.7.1 Introduction

Aquaculture is an ancient practice which has around 4000 years of history, it is mentioned in the tombs of Egyptian rulers, it is mentioned in the books written by the Chinese and romans. However, aquaculture grew and intensified in the last three decades, so the current practices of aquaculture are different than ancient ones aquaculture has grown faster than agriculture at the rate of 8.3% per year since 1970.

Looking at current trends the world will be vastly different, populations will increase and will mostly be concentrated in the urban areas. Climate change will raise the water levels, water will become scarce and will become more precious, suitable lands for agriculture will be less. Due to increased food demand, agricultural practices need to be made sustainable. One of the challenges in coming years would be to meet increasing sea food demands via aquaculture since production obtained from wild fisheries will remain constant at best. As aquaculture production increases we must make sure that we don’t repeat the same mistakes which were made during green revolution, which although led to increased food production, caused a lot of environmental damages. Therefore, understanding environmental impacts while developing a responsible aquaculture model is necessary.

The Four goals required to develop a sustainable aquaculture model are:

1: Improvement of management practices

There are variety of techniques of growing variety species of crops in different locations, modern production systems aren’t well managed, improvement in feeding techniques can result in overall improvement of management practice improving management techniques can improve production output, which leads to increased profits, another benefit of improving management techniques is that it creates a lot of environmental benefits.

2: Emphasis on local human capital

Dissemination of information can be carried out in a better manner by involving local deciding and local resources available. Much of the aquaculture knowledge is transferred through communication among folks that are engaged within the same profession and thru government outreach, thanks to various factors it's inconceivable to own one practice followed by an outsized number of farmers. The key to achieving more production in aquaculture is to possess the most effective of local resources and human capital.

3: Development of risk management systems

One of the concerns related to aquaculture that needs addressing is that the impact of invasive species, there are secure systems for these new species which have temperature control and water management systems. Most culture systems have a risk of escape which mostly occurs during natural disasters. A number of these foreign species carry different types of viruses, so a good risk management system is vital to avoid these mishaps.

4: Identification of more sustainably grown crops

Differentiation done between products from more sustainable and less sustainable aquaculture fish systems is carried out in a poor manner which results in financial loss to farmers. Standards which define agriculture, husbandry products don't exist for aquaculture products, although certain retailers require the products be labelled, there isn’t a hard and fast standard of branding the products.

3.7.2 Conclusion

Due to lack of available land the potential for expansion of terrestrial food products looks limited, so by 2050 Aquaculture will be predominant source for food, although aquaculture has huge potential, it should be done responsibly and sustainably to preserve the environment, in order to do this business models needs to be changed.

3.8 Fish Diseases in Aquaculture and assessment of economic loss due to disease

3.8.1 Abstract

Aquaculture is recognised as an important sector, it provides a significant amount of protein supplement, besides supporting the rural populace economically. Over 14.5 million people are dependent on aquaculture for their livelihood. However, occurrence of diseases has been a major constraint for the development of aquaculture, there are various reasons behind these health problems: pathogens, environmental factors and poor animal husbandry. So, understanding and avoiding the above-mentioned problems is essential. A survey was carried out in the states of Odisha Andhra Pradesh, Chhattisgarh, Jharkhand. It was found out that the occurrence of bacterial disease has been a great concern for farmers. Parasites, viruses, poor management practices and seasonal climate changes have contributed to diseases in marine life.

3.8.2 Introduction

There has been significant progress in the field of aquaculture, the shift from traditional methods to modern methods have resulted in an increase in production and profits. It has provided employment to the rural people of India. Many Indian states like Andhra Pradesh, Karnataka, Maharashtra, Haryana etc have taken up fish farming as trade and enterprise. The top three States having the largest land under aquaculture are Andhra Pradesh, Karnataka and West Bengal. Despite the phenomenal growth, the progress has caused some unwarranted activities for both the species and the environment. The consequence has been the spread of diseases which has affected the production of aquaculture. These diseases have cost the aquaculture industry in crores. In this paper the effect of diseases and economic loss to diseases have been reviewed.

3.8.3 Materials and methods

The data was collected over a period of two years from the states of Andhra Pradesh,Jharkhand, Odisha and Chhattisgarh. Primary data was collected through field surveys, personal interviews with fish farmers, hatchery owners. Secondary data was collected through published reports, news articles. Data collected from these various sources were entered into a database system and analysed.

3.8.4 Results and discussions

Various information relating to aquaculture were collected from different fish farmers. It was found that the practices were different in each state, Andhra Pradesh mostly follows shrimp culture, Jharkhand follows cage culture. Farms in AP were clustered, in other states the farms were mostly scattered. Most of the freshwater produce are marketed for local purposes brackish water produces like the tiger prawn and white shrimp are marked for exports and not usually sold in the local markets.

3.8.5 Disease problem in aquaculture

Diseases in marine species can cause huge losses in terms of production and profits, it can have negative impacts on the socioeconomic life. A total loss of $1 Billion USD was reported due to diseases in shrimp culture. Stress factors such as non-ideal water quality, higher microbial load, high stocking densities can increase the chances of Infection. Fish are vulnerable to a variety of diseases and if affected can be very difficult to deal with. Nutritional deficiencies, poor water quality and overstocking can accelerate bacterial Infections. Diseases like Red disease, motile aeromondas septicaemia (Aeromonas hydrophila), edwardseillosis (edwardseilla tarda), flexibacteriosis (Flexibacter columnar) etc were observed in carp culture.

Beside bacterial diseases parasitic infections haver been a major cause of concern in most farms, causing significant setback to the aquaculture industry in India, these parasites can multiply rapidly underwater when the water conditions are poor. Oomycete fungi caused the most infections in fishes. Besides bacteria and fungi, the other causes for diseases are low dissolved oxygen which happens during cloudy/winter weather, low alkalinity causes microbial infections. Single or multiple parasites in combination with disease causing bacteria have posed a huge risk to the growth of aquaculture in India.

Viral infections have posed a lower risk in India, they haven't caused huge economical losses, viral infections have been isolated from some species of fishes. Like fishes' diseases were also found in shrimps, due to the significant growth in shrimp culture the number of diseases affecting the shrimp culture has increased.

3.8.6 Economic loss due to diseases

The survey has shown that in places where better management practices were followed a 20%-30% profit was made. Diseases have caused 10-15% loss. Shrimp culture has been a risky business, in which the profits were extremely high or extremely low, viral diseases have significantly affected shrimp culture in countries like china, Brazil and Ecuador. In India the loss due to the disease outbreaks in farms located in nine coastal districts were estimated to be 1000 crores. Additional price losses were also recorded on poor quality of final product like deformed organs, loose shell and muddy smell, these problems have severely impacted socio-economic factors of the people dependent on these farms, in some countries the losses were so huge that some farms never recovered from the blow. Diseases in aquaculture threaten the food security of small holder shrimp farmers and job security of the people working in these farms.

3.8.7 Conclusion

Disease has become a major worry for the rapidly expanding aquaculture industry, which affects the socioeconomic status of coastal rural India. Currently less emphasis is laid on disease management systems and to train people who are can take on these challenges.

In India emphasis needs to be laid on improving the shrimp culture practices to lessen the environmental impact, more efforts are needed to improve production through with adoption of bio secure measures to reduce losses. The participation of Private sector to ensure quality standards to ensure that they meet global safety standards is needed. Better management practices which are scientific, cost effective and simple should be encouraged to be taken up by farmers. Overall better husbandry methods, effective management practices can reduce diseases which can lead to higher production and can reduce losses.

CHAPTER 4

DATA

This section contains details about various datasets which have been used during execution of entire capstone project.

4.1 Aquaculture Production in India Dataset

4.1.1 Overview

This dataset is used for performing time series analysis in R to predict overall Aquaculture production in India over coming 10 years. Dataset consists of Aquaculture production values in India from year 1980 till 2018. This dataset is obtained from ministry of fisheries handbook 2018.

4.1.2 Columns

Dataset consists of 2 columns -

- Year
- Aquaculture production (in tonnes)

4.2 Aquaculture Prices in India Dataset

4.2.1 Overview

This dataset is used for performing time series analysis to predict prices of marine and inland fish in India in next 5 years. Dataset consists of prices of inland and marine fish in India from year 2005 till 2016. Dataset is obtained from ministry of fisheries handbook 2018.

4.2.2 Columns

Dataset consists of following 3 columns -

- Year
- Inland fish price (Rs/Kg)
- Marine fish price (Rs/Kg)

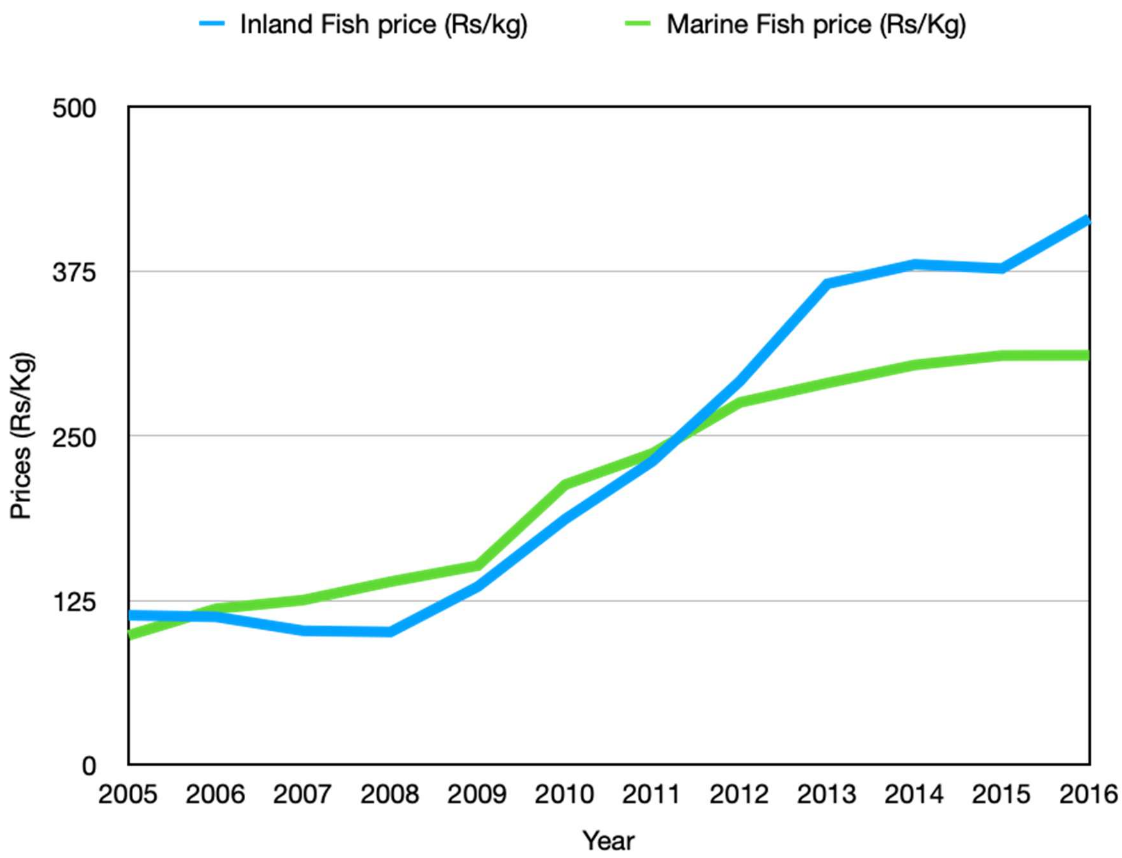


Figure 4.1: inland and marine aquaculture fish prices from 2005 till 2016

4.3 Disease Detection Dataset

4.3.1 Overview

This dataset is used to implement supervised machine models in order to detect spread of diseases in aquaculture farms. Dataset contains information about 250 odd farmers and their respective aquaculture farms spread across 6 districts in the state of West Bengal. Dataset is obtained from data.world website.

4.3.2 Columns

Dataset consists of 45 columns. Some of the important attributes of dataset are -

- Area (in hectare)
- Previous prevalence
- Stocking age
- Feed type
- Stocking density
- Virus detected

4.3.3 Target Class Distribution

62% - Virus not detected

38% - Virus detected

4.4 State wise Land Area and Aquaculture Production Dataset

4.4.1 Overview

This dataset is used to suggesting appropriate species based on various factors. Dataset contains information about Aquaculture land area and total production for 10 coastal states and for 3 different aquaculture species L wannamei, Scampi and Tiger Shrimp for the last 20 years. Dataset is obtained from ministry of fisheries handbook for the year 2018.

4.4.2 Columns

Dataset consists of Land area (in hectares) and Aquaculture production (in tonnes) for all the states on vertical axis. Years from 2001-02 till 2017-18 are represented on horizontal axis.

4.5 State wise Freshwater Fishponds Dataset

4.5.1 Overview

Freshwater fishponds dataset is used to suggest regions where freshwater aquaculture can be carried out at an extensive scale in order in meet the growing demands. Dataset contains information about state wise number of operational and non operational freshwater fishponds. This dataset is obtained from the fisheries survey carried out by GOI.

4.5.2 Columns

Dataset consists of following 4 columns -

- State
- Number of operational freshwater fishponds
- Number of non operational freshwater fishponds
- Total number of freshwater fishponds

4.5.3 Snapshot

State	Operational	Non-Operational	Total
West Bengal	78	84	162
Karnataka	40	232	272
Rajasthan	30	23	53
Bihar	30	35	65
Tamil Nadu	26	35	61
Maharashtra	25	28	53
Orissa	23	4	27
Andhra Pradesh	16	24	40
Madhya Pradesh	9	20	29
Uttar Pradesh	50	112	162
Punjab	2	4	6
Gujarat	2	14	16
Kerala	2	3	5

Total	333	618	951

Table 4.1 : State wise freshwater fishponds

4.6 Aquaculture Production based on Water Type Dataset

4.6.1 Overview

Dataset gives information about Freshwater, Brackishwater and Marine water Aquaculture production in India from year 1980 till 2018. This dataset is used to compare the performance of Aquaculture sector on the basis of water type.

4.6.2 Columns

Dataset consists of following 3 columns -

- Year
- Water type
- Aquaculture production (tonnes)

CHAPTER 5

PROJECT REQUIREMENT SPECIFICATION

5.1 Introduction

The document describes various type of requirements needed in the project. The various category of requirements include software, hardware, functional, non- functional requirements as well as other type of requirements such as data requirements etc.

5.1.1 Project Scope

The objective of the project is to help fish farmers maximise their profits by improving quality and quantity of aquaculture in India. The benefit of the product would be providing employment to local youth, improving GDP through export of aquaculture products etc.

5.2 Product Perspective

Research papers suggest the fact there is a big need for improving aquaculture production in India in response to rising demand in both local as well as global markets. This is the main reason for the origin of the product.

5.2.1 Product Features

Product will help fish farmers gauge the future demand as well as prices of aquaculture products in the future based on forecasted values using the time series analysis performed.

Product will also help fish farmers decide which type of aquaculture to perform ie- Freshwater aquaculture vs marine water aquaculture, shrimp farming vs crab culture based on various factors such as weather, local demand, prices, soli type and capital investment required for each type.

5.3 User Classes and Characteristics

- Fish farmers who are interested in improving quality and quantity of aquaculture production in order to maximise their profit.
- Middlemen who buy in wholesale from farmers and sell it in the market.
- Agencies responsible for export of aquaculture products to foreign land.
- Customers who are interested in prices and availability of species in their region.

5.4 Operating Environment

The system will be able to run in latest versions of any of the major operating systems. For performing time series analysis in R requirements include R programming language and IDE for running R which can be RStudio, VS code etc.

The prices of aquaculture products, aquaculture products datasets must be downloaded before code execution.

For building supervised machine learning models Jupiter notebook or Google Colab can be used in order to run python code.

CHAPTER 6
SYSTEM SPECIFICATION REQUIREMENTS

6.1 Functional Requirements

For time series analysis in R after dataset collection data preprocessing is done by removing commas in aquaculture production column there by converting string to integer. Augmented dickey fuller test is performed in order to check if time series is stationary or not. There time series analysis is done using multiple models and the model having the lowest RMSE value is chosen as the best model and forecasting is performed for the next 10 years using the best model.

For supervised machine learning after data collection columns which do not affect the detection of disease are removed from the dataset. After that PCA can be done do further reduce the number of columns from the dataset. Missing values are handled by replacing with mean or mode value of that particular column. Dataset is than divided into training and testing subsets. Supervised machine learning models such as decision tree and KNN are used to train the model. Model performance is evaluated by testing model. Models can than be used to predict spread of disease from for new row of values.

6.2 External Interface Requirements

6.2.1 User Interfaces

The user can view the forecast values of prices and production through charts, visualisation tools and observe trends pattern and rate of growth.

For detection of diseases fish farmer can enter his own set of values such as species type, water pH, number of days from start of production check whether model predicts spread of disease or not and than take appropriate action.

6.2.2 Software Requirements

- Python
    - Version - 3
    - OS - Any of the major operating systems

- R
    - Version - 4.0.3
    - OS - Any of the major operating systems
- RStudio
    - Version - 1.3
    - OS - Any of the major operating systems
- Jupiter Notebook
    - Version - 6.0.3
    - OS - Any of the major operating systems

6.3 Non-Functional Requirements

In this project safety measures include mainly performance requirements which must be met in order for product to the rendered usable on a large scale.

6.4 Performance Requirement

For time series analysis the performance requirement for every model is that p-values should be less than or equal to 0.05, R-square value should be high and RMSE value should be as small as possible.

In case of machine learning models accuracy of trained model should be high enough and cases where model predicts no disease but in reality disease spread has occurred must be minimised since this is the worst case scenario. Also training accuracy must not be too high compared with testing accuracy in order to reduce risk of overfitting.

CHAPTER 7

SYSTEM DESIGN

7.1 High level design

7.1.1 Introduction

High level system design describes the step by step procedure and the interaction among various modules in order to build the final product. The overall objective or aim of the product is to help fish farmers maximise their profits by improving aquaculture practices in India. This will in turn benefit in economy as aquaculture contribution to GDP can increase through exports and also employment opportunities will be made available to the local youth. Hence the product developed will be a win-win situation for all stakeholders involved.

7.1.2 Current System [if applicable]

Instead of building upon a system already present the product developed from fresh by using core data science, data analysis and machine learning concepts and understanding the research work done in this field by doing literature survey.

7.1.3 Design Considerations

7.1.3.1 Design Goals

Aim of the design is to be easy to use and also that it can be used by various category of users.

Easy of use means fish farmers can easily give their data such as water pH, species category, weather and the model we tell whether there is disease spread possible or not. Product will also suggest species which can be used based on important factors such as local demand, weather etc.

7.1.3.2 Architecture Choices

Incase of detection of diseases in aquatic life since target class is present in Dataset indicating spread or no spread of disease preference is to use supervised machine learning models such as decision tree and KNN over unsupervised machine learning models. Also since the corresponding dataset contains too many columns the requirement is find the columns which cause the most variance in the target class which can be done by applying PCA.

7.1.3.3 Constraints, Assumptions and Dependencies

Assumptions include the fact that there is not going to be a major overhaul in terms of aquaculture reforms in India. For eg - If rules are made stricter in terms of export for the nation which is the largest exporter of aquaculture products in India than naturally the aquaculture production in India in future will go down.

Assumption also includes the fact that all the software requirements mentioned are run in the correct version so that there are no dependency issues involved. For eg - there can be minor issues if python 3 is run instead of python 2.

Constraints include the fact that for Andhra Pradesh before the state was split into Telangana and Andhra Pradesh we have the values for production and area utilized for production as single value each for the entire state.

7.1.4 High Level System Design

Time Series Analysis in R

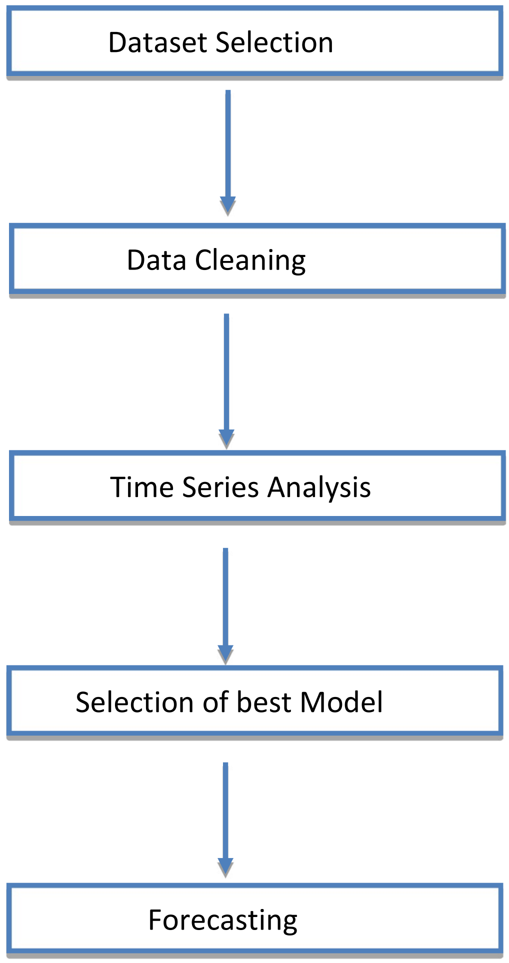


Fig 7.1 Time series analysis in R

Fig 7.2: Sequential state diagram for time series analysis

Dataset Selection

Dataset for time series analysis and forecasting is selected from ministry of fisheries handbook (2018).

Prices of Aquaculture Products Dataset

This dataset consists of 3 columns which are Year, Price of inland fish (Rs/Kg) and Price of marine fish (Rs/Kg). This dataset is split into 2 different time series and time series analysis and forecasting is performed on each.

Aquaculture Production in India Dataset

This dataset consists of 2 columns year, aquaculture production in India (tonnes).Time series forecasting is also performed on this dataset.

Data Cleaning

Aquaculture production in India column is a string consists of numbers and commas in between. This column is converted into integers by removing commas present.

Time Series Analysis

In total time series analysis is performed 3 times ie - twice on price dataset and once on aquaculture production dataset. In all 3 cases augmented dickey test is used to check if time series is stationary or not. If time series is not stationary is not stationary it is converted to stationary by performing differencing in order to remove the trend present in the time series.

The models which are used to perform time series analysis are -

- Naive Model
- AR Model
- MA model
- ARIMA model
- Linear Regression model
- Non linear Regression model
- Exponential Smoothing model
- Holt-Winters model
- Holt-Winters model with Trend Smoothing

Selection of Best Model

Model having the least RMSE value is selected as the best model for performing forecasting. Other criteria include p-value less than or equal to 0.05 and R-square value should be really high.

Forecasting

Using the best model chosen in the previous step forecasting is performed to predict values of prices and aquaculture production in the coming 10 years.

Supervised Machine Learning in Python

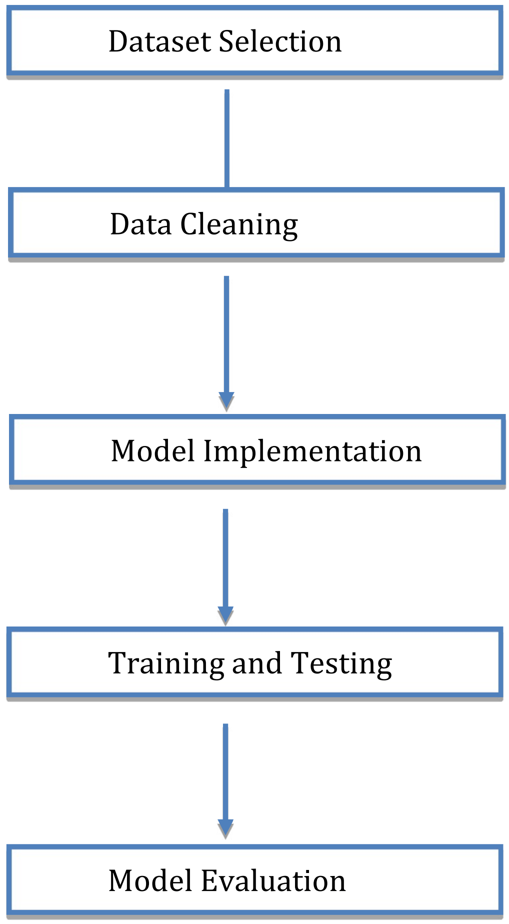


Fig 7.3: Supervised Machine Learning in Python

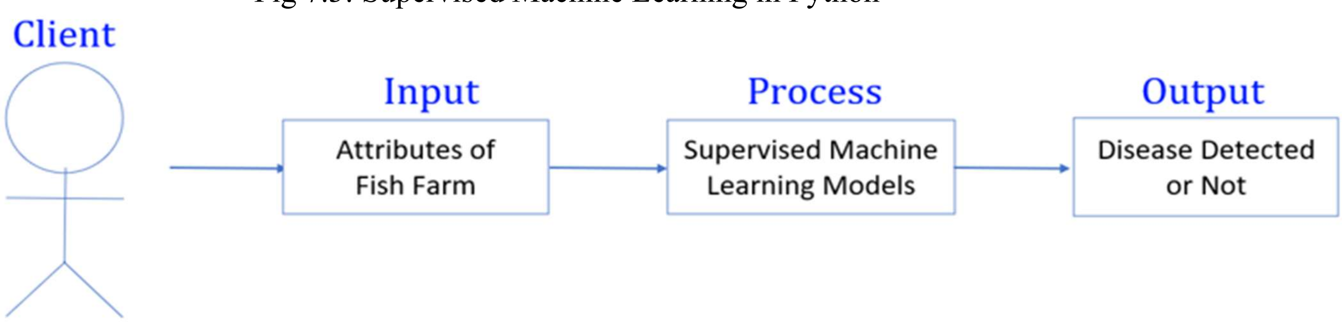


Fig 7.4: Sequential state diagram for supervised machine learning

Dataset Selection

Dataset for detecting spread of diseases in aquatic life is obtained from data.world. This dataset consists of 40 columns.

Data Cleaning and Preprocessing

Columns which do not affect the target class column are removed from the dataset. Further on PCA can be performed in order in remove attributes which cause the least deviation in the target class. Further on missing values are replaced using mean or mode of that particular column.

Model Implementation

Since the target class is available in the dataset in binary form supervised machine learning models are preferred over unsupervised machine learning models. As the target class is binary make can make use of decision tree and KNN models etc.

Training and Testing

Dataset is split into training and testing subsets. Model is trained on the training set and tested on the testing set. For model performance to be good testing accuracy should be close to training accuracy such that overfitting does not occur. False negatives and false positives should be minimum in confusion matrix.

Model Evaluation

Each model can be evaluated based on training and testing accuracy as well as confusion matrix. Later the models can be used to predict disease or no disease on unseen data.

7.1.5 User Interface Diagrams

In case of time series analysis users will be able to view trends and forecasts through line charts, graphs which will be easier to understand for end users rather than showing model performance through parameters such as R-square value or RMSE value.

In case of disease detection users can fill the values applicable to the farm and the model will generate whether disease is detected or not.

Overall the interface will be easy to use for the end user.

7.1.6 Design Details

7.1.6.1 Novelty

We are using known computer science methodologies in a sector in which this kind of technologies have not been used for hence there is huge scope for improvement in this field.

7.1.6.2 Performance
Performance of design will in turn depend on performance of models used in the design. Good time series model will have high R-square value, p-value <= 0.05 and low RMSE value. Supervised machine learning model should not overfit and false positives and false negatives in confusion matrix should be low.

7.1.6.3 Reliability
The design is highly reliable as there is no chance of any fault being developed over time.

7.1.6.4 Maintainability
The product is easy to maintain as there is no maintain cost involved in terms of time, effort and money.

7.1.6.5 Portability
There is no issue involved in terms of portability as the design can be easily moved from one system or another or from one OS to another.

7.1.6.6 Reusability
Sub Systems of the product can be used in other cases hence the design is reusable.

7.1.7 Other Requirements
Other requirements include all the datasets which have been collected for the purpose of this project.

7.1.7.1 Data Requirement
Dataset of prices of aquaculture products as well as production in India are both collected from handbook of ministry of fisheries (2018). Statewide area utilisation and total production dataset is obtained from the same source. Dataset for detecting spread of diseases among aquatic life is obtained from data.world website.

7.2 LOW LEVEL DESIGN

7.2.1 Introduction

7.2.1.1 Overview

Low level design and implementation document consists of details about algorithms, structures used throughout the project at lower abstraction level. We make use of class diagrams, sequence diagrams as well as user case diagrams to illustrate the sequence flow of the work carried out during the duration of the project.

7.2.1.2 Purpose

The purpose of low level design is to provide internal logical design of the actual program code and it describes the class diagrams with methods and relations between multiple classes and their interdependencies.

7.2.1.3 Scope

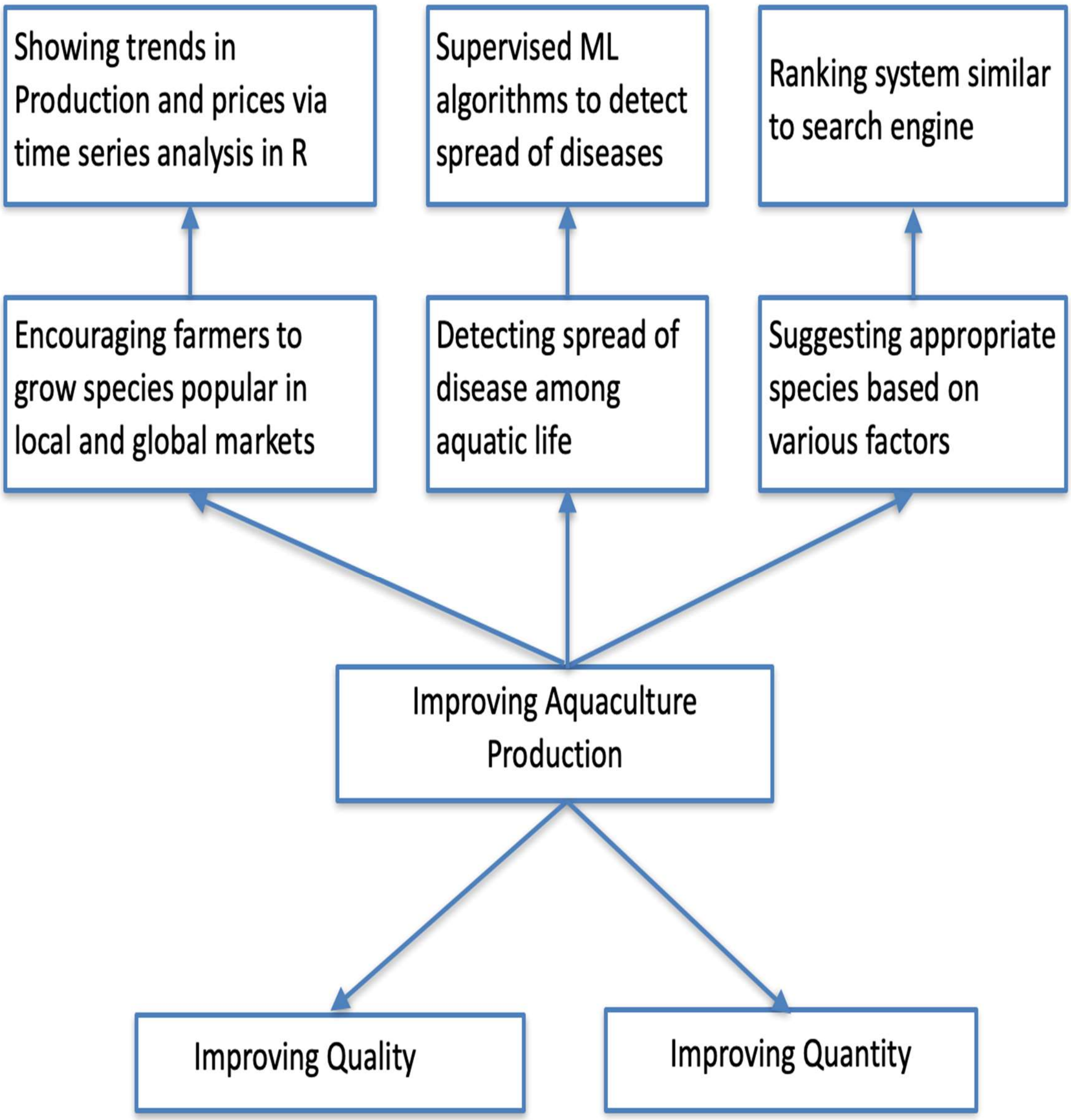
Low level design document consists of component level design process which follows step by step refinement process. LLD document contains master design diagram, use case diagram and class diagram for each subcomponent as well as algorithm, implementation and final results for the code developed for each subcomponent in the entire project.

7.2.2 Design Constraints, Assumptions, and Dependencies

Assumptions include the fact that there is not going to be a major overhaul in terms of aquaculture reforms in India. For eg - If rules are made stricter in terms of export for the nation which is the largest exporter of aquaculture products in India than naturally the aquaculture production in India in future will go down.

7.2.3 Design Description

The HLD and LLD design has been developed keeping aim and objective of the project in mind which is to improve aquaculture production in India.

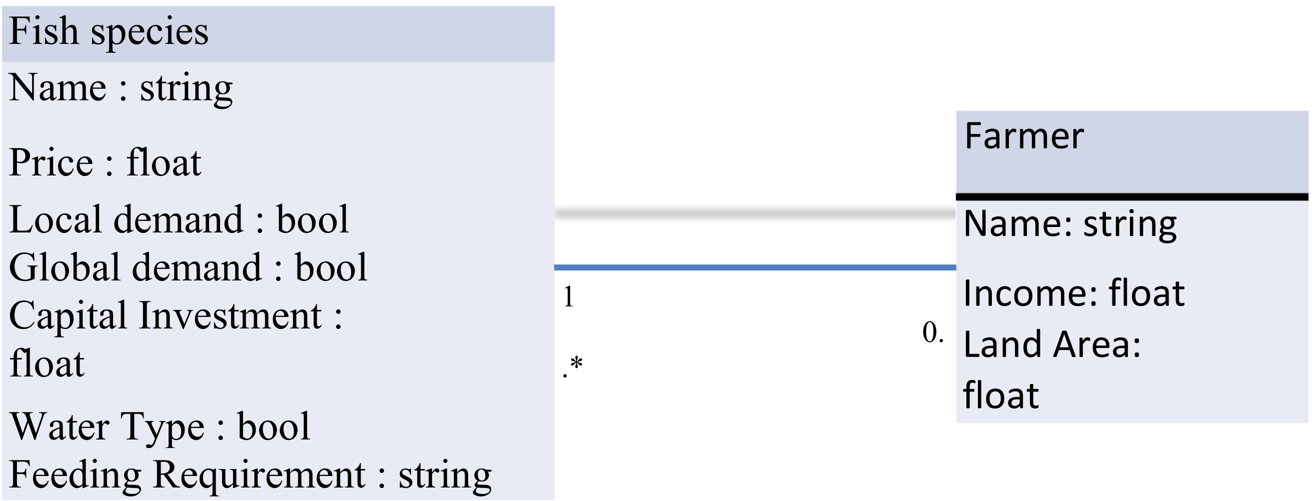


7.2.3.1 Module 1- Forecasting values of Aquaculture production and prices for next 10 years

7.2.3.1.1 Description

Module 1 is based on forecasting values of Aquaculture production and prices for next 10 years.Sharing the information to general public that fish production is rising at a rapid pace in response to growing demand in local and global markets and as well as the rise in prices will encourage more people to take up aquaculture which we result in increase in overall production.

7.2.3.1.2 Class Diagram

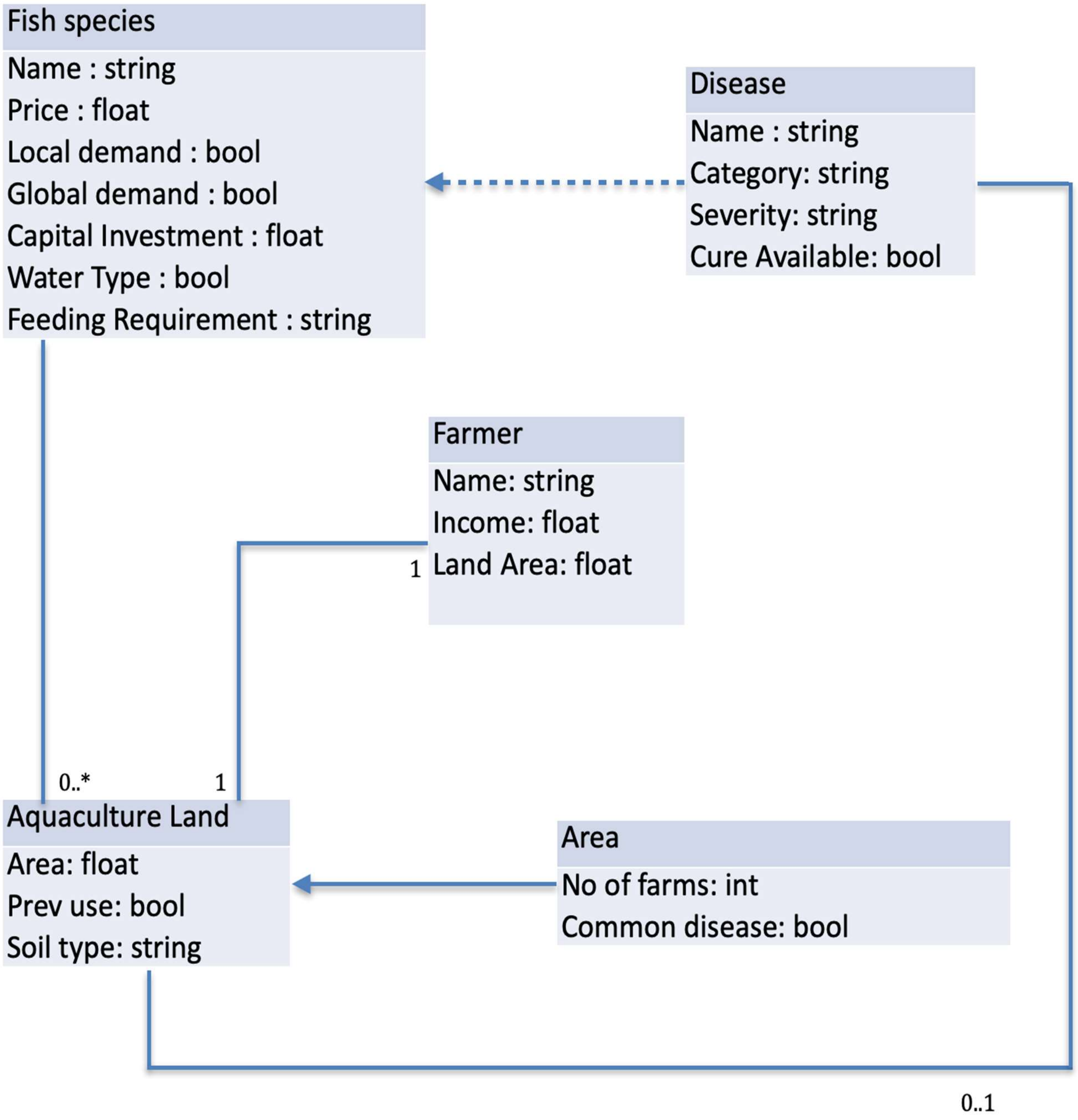


7.2.3.2 Module 2 -Detecting spread of disease among aquatic life

7.2.3.2.1 Description

Module 2 is based on detecting spread of disease among aquatic life. This will help to improve both quality and quantity of aquaculture production in India. Our model and analysis done will help farmers to find out the cause of disease spread so that appropriate action can be taken to prevent disease spread in future.

7.2.3.2.2 Class Diagram

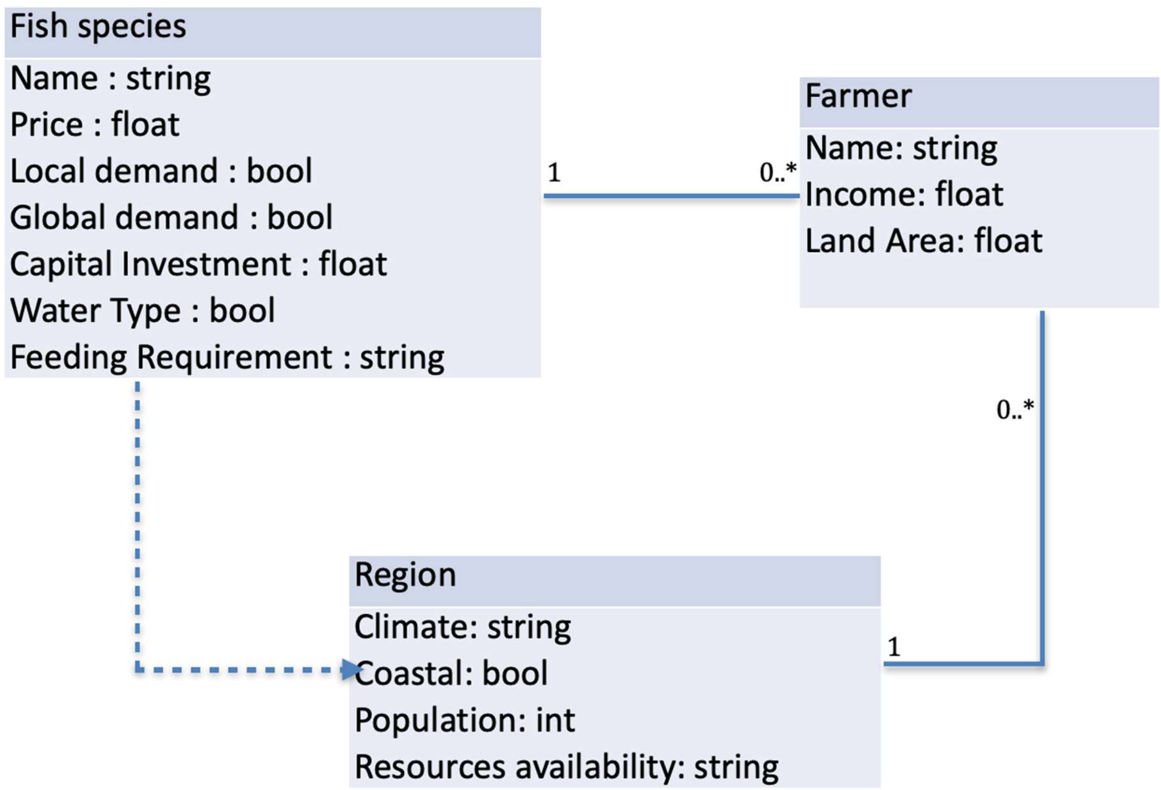


7.2.3.3 Module 3 – Suggestion of appropriate species to end users

7.2.3.3.1 Description

Module 3 is used to suggest appropriate species to end users by considering all factors in order to maximise the profits of fish farmers in the long run. For eg- Fish farmer with low income might have been producing species which has demand in local regional market but overall profit was minimal can shift to crab culture for which capital investment is minimal but profits will be high as demand for crabs is very high in international market and the demand continues to grow.

7.2.3.3.2 Class Diagram



7.2.4 Proposed Methodology / Approach

7.2.4.1 Algorithm and Pseudocode

Time Series Analysis for forecasting Aquaculture Production and Prices
Steps-

- Dataset collected from ministry of fisheries website, 40 years of Aquaculture production values is used to forecast values for next 10 years and 20 years of fish prices are used to predict prices over coming 5 years.
- Data cleaning includes converting Aquaculture production column from string into int in order to perform the required analysis.
- 10 different models are generated by performing time series analysis in R.
- Best model is selected on the basis of RMSE value as well as adjusted R-square.

Disease detection using Supervised Machine Learning Algorithms
Steps-

- Dataset is selected from data.world website and consists of details about aquaculture performed by farmers across 6 different districts in West Bengal.
- Attributes which do not affect target variable such as farmer age are removed, variables having high correlation among them are removed and at the end best 12 attributes are selected using random forest classifier.
- Dataset is divided into training and testing set using 70-30 ratio.
- K fold cross validation method is used in order to remove the bias of the dataset.
- Decision tree and random forest classifier are used as the models in order to perform the required analysis.

Suggesting species to end users
Steps-

- Based on inputs provided such as climate, capital investment available, water type, local as well as global demand aquaculture species will be ranked in accordance with feasibility and maximising the profits of the fish farmers in the long run.

7.2.4.2 Implementation and Results

Time Series Analysis for forecasting Aquaculture Production and Prices

- For aquaculture production time series analysis ARIMA model turns out to be the best among all models having a low value of RMSE and MAPE is which turns out be 4.69.
- For prices of aquaculture species Holt Winters model with trend smoothing turns out to be the best model having adjusted R-square value 0.96.

Disease detection using Supervised Machine Learning Algorithms

We use stratified k fold cross validation technique to make distribution of target variable similar in both training and testing dataset. It reduces the bias involved while analysing performance of the models.

After Stratified K fold cross validation method is added in order to remove the bias training accuracy becomes more than 80% while test accuracy becomes around 60% and 70% for decision tree and random forest classifier respectively.

CHAPTER 8

IMPLEMENTATION AND PSEUDOCODE

8.1 Implementation Overview

Time Series Analysis

Models Used

Naive

MA

AR

ARIMA

Exponential Smoothing Holt-Winters

Holt-Winters with Trend Smoothing Linear

Model Non Linear Model

Libraries Used

ggplot2

forecast

knitr

tseries

Figure 8.1 : Implementation overview of time series analysis

8.2 Time Series Analysis of Aquaculture production in India Dataset Information

8.2.1 Dataset Information

Dataset for aquaculture production in India from the year 1980 till 2018.

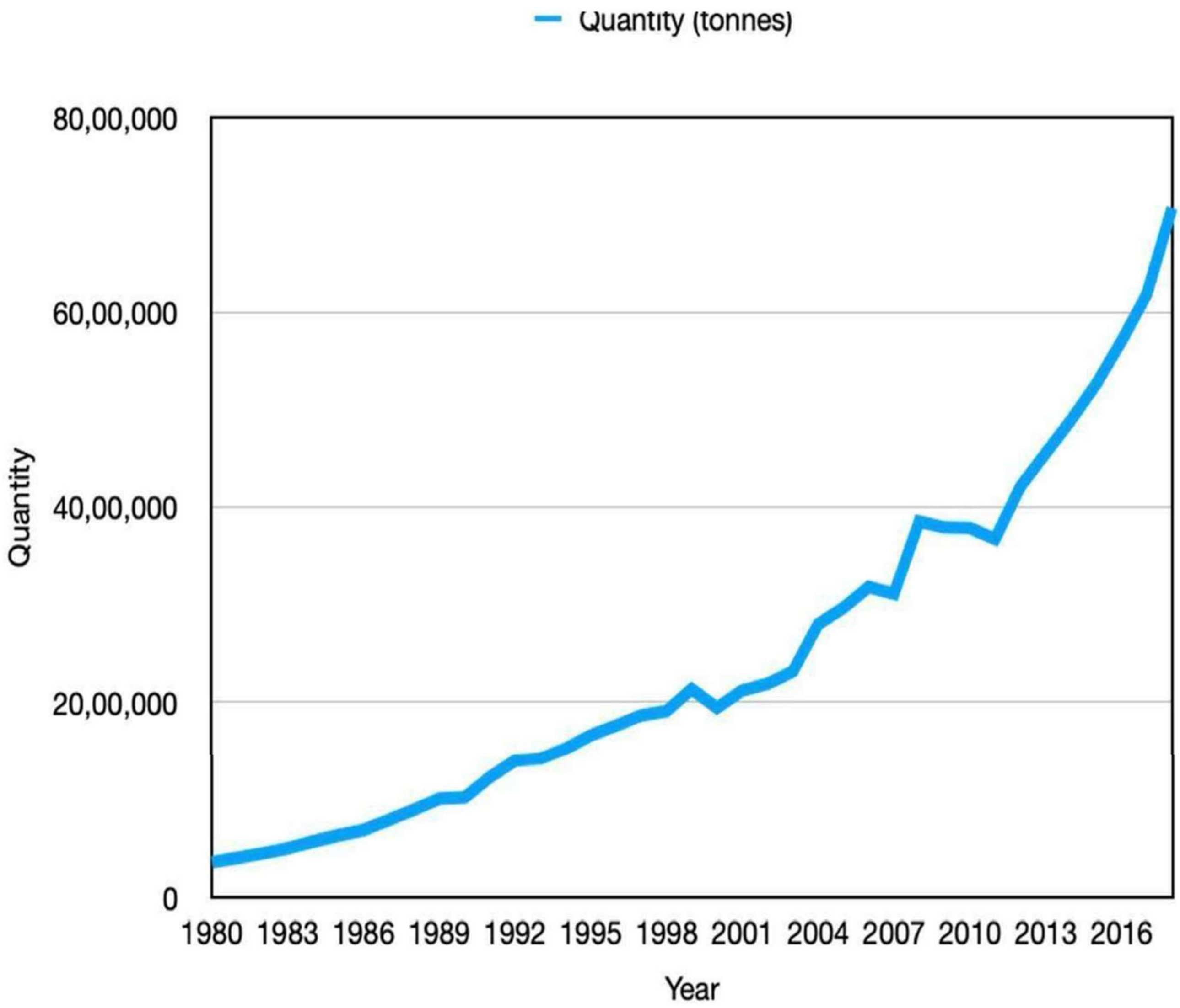


Figure 8.2: Aquaculture production year wise plot

8.2.2 Data Cleaning and Preprocessing

String is converted to an integer ie - aquaculture production quantity column.

8.2.3 Check for Stationarity

Time series is checked for stationarity by using augmented dickey fuller test. P-value less than 0.05. Indicates time series is stationary.

8.2.4 Convert to Stationary Time Series

As we can observe from the plot that time series has trend component but no seasonality component. In order to remove the trend differencing is performed.

8.2.5 Time Series Analysis

Time series analysis is performed on multiple models and model having the least RMSE value is used for forecasting.In this case the the model is non linear regression model.

8.3 Time Series Analysis of Prices of Aquaculture Products

8.3.1 Dataset Information

Dataset consists of prices of marine fish and inland fish in Rs/Kg from year 2005 till 2016.

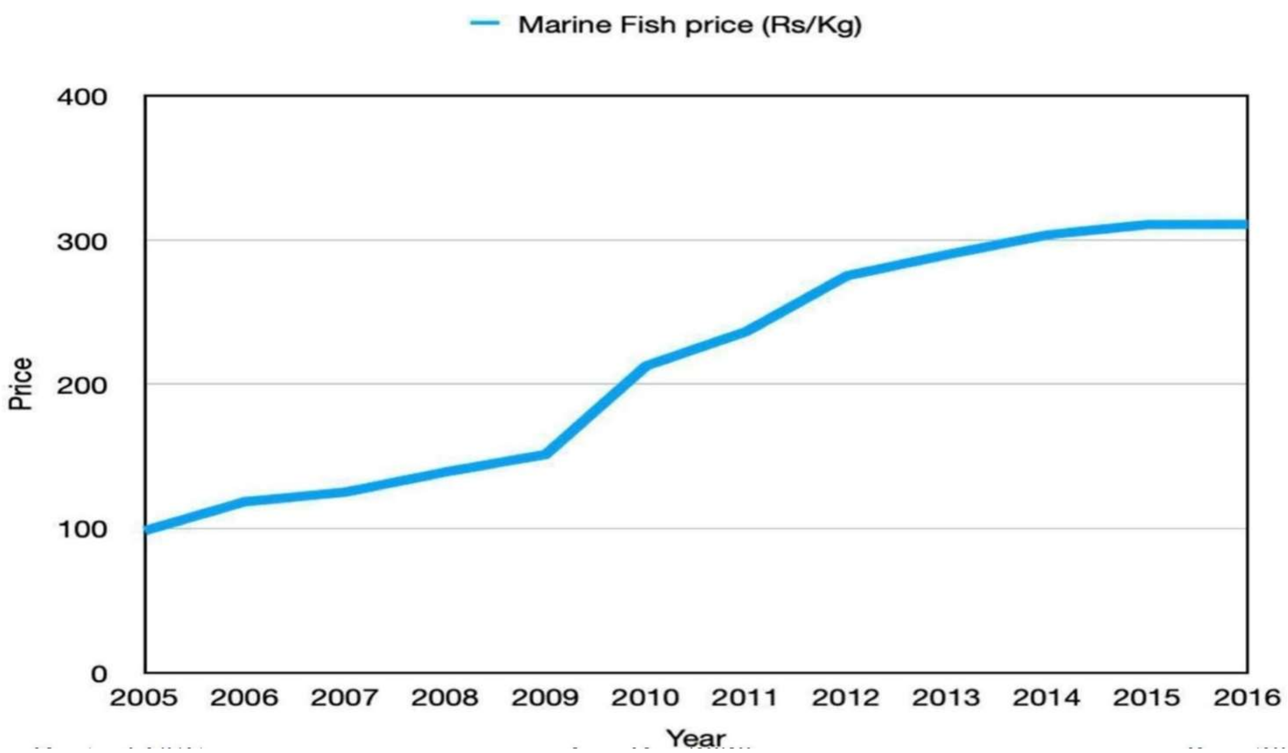


Figure 8.3: Price of marine fish in Rs/Kg year wise

Dataset for prices of marine fish in Rs/Kg from year 2005 till year 2016 is obtained from the handbook of ministry of fisheries (2018).

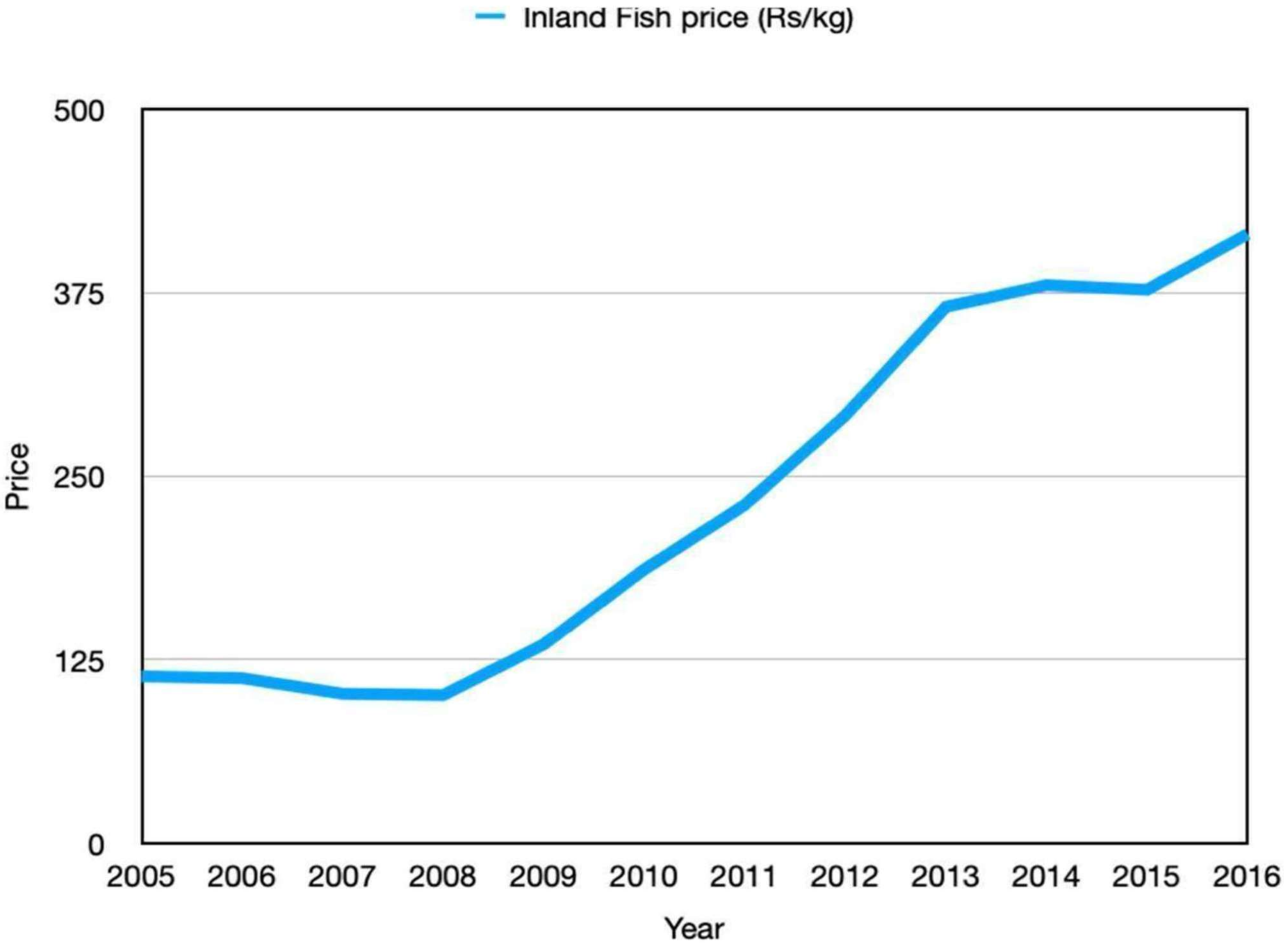


Figure 8.4: Price of inland fish in Rs/Kg year wise

8.3.2 Implementation Details

Time series is checked for stationarity and later converted into stationary time series.
Time series forecasting is done using Naive, AR, MA, ARIMA, ARIMA using auto.arima() function, linear regression, non-linear regression, simple exponential smoothing, Holt-Winters as well as Holt- Winters with trend smoothing model.

The best 2 models turn out to be ARIMA models and Holt-Winters model with trend smoothing. Forecasting to done for year 2017 till 2016. Forecasting is done using the best 2 models.

8.3.3 Check for Stationarity

Time series is checked for stationarity by using augmented dickey fuller test. P-value less than 0.05 Indicates time series is stationary.

8.3.4 Convert to Stationary Time Series

As we can observe from the plot that time series has trend component but no seasonality component.In order to remove the trend differencing is performed.

8.3.5 Time Series Analysis

Time series analysis is performed on multiple models and model having the least RMSE value is used for forecasting. For both inland prices time series and marine prices time series the 2 best models turn out to be ARIMA model and Holt-Winters Model with trend smoothing.

8.4 Supervised Machine Learning to detect Disease spread in Aquaculture

8.4.1 Overview

Supervised Machine learning models are implemented in order to detect spread of disease in Aquaculture farms in 6 districts of West Bengal. Supervised ML is preferred since target variable ie - virus detected or not is present as a boolean variable as part of the dataset.

Supervised Machine Learning

Models Implemented

Used Decision Tree

Random Forest Classifier

K Nearest Neighbour

Logistic Regression

Libraries

random

sklearn

numpy

pandas

matplotlib.pyplot

- build from scratch

Figure 8.5: Implementation overview of supervised machine learning for disease detection

8.4.2 Data Cleaning

- Removing variables such as serial number, farmer age, farmer education and district which do affect the target variable.
- Deletion of pH, salinity and temperature columns from the dataset since 30% of values of these 3 columns are missing and variance of values in these 3 columns is very less. Eg - temperature of all 6 districts is almost same as they lie next to each other.
- Variables such as Crab fench, Aerator which are highly correlated with other variables in the dataset are also removed.

8.4.3 Feature Extraction

Random Forest Classifier is preferred over PCA for extracting top 12 important features from the dataset. More than 85% of the variance of the dataset is retained even after decreasing number of columns from 45 to 12.

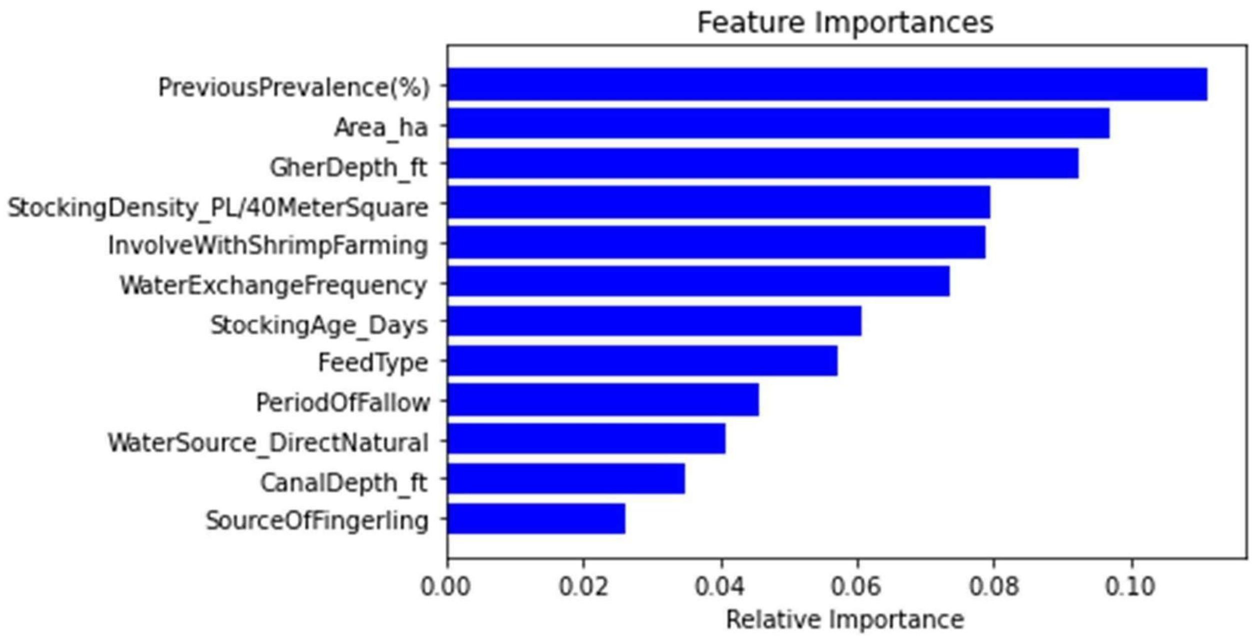


Figure 8.6: graph depicting top 12 features and their respective importance

8.4.4 Stratified K Fold Cross Validation

K fold cross validation is used to reduce the bias involved and make the models implemented robust since the size of the dataset is small. Stratification ensures that the distribution of target variable is similar in both training and testing set.

8.4.5 Splitting Dataset into Training and Testing Set

- For model implementation without stratified K fold cross validation split ratio is set as 0.7
- since dataset size is small.
- For model implementation with stratified K fold cross validation k value is set to be 4. Dataset is divided into 4 groups and each time 1 of the groups is selected as testing set and other 3 as training set. This step is performed 4 times and average of scores of the 4 instances is taken to the overall performance of the model. This makes model more robust against random noise.

8.4.6 Model Implementation

Four different supervised machine learning models are implemented in order to detect spread of disease in Aquaculture Farms.

8.4.6.1 Decision Tree

Decision tree is one of the most common models used for supervised ML. Max depth of decision tree is set as 4 via trial and error method in order to maximise accuracy on test set.

8.4.6.2 Random Forest Classifier

Random forest classifier fits number of decision tree classifiers on sub samples of dataset and uses averaging to improve accuracy and control over-fitting. Similar to decision tree max depth value is set as 4.

8.4.6.3 K Nearest Neighbour

KNN is a simple algorithm which is commonly used for solving classification problems using supervised machine learning. K value should be an odd number and approximately equal to square root of number of columns of the dataset hence K value is set as 3.

8.4.6.4 Logistic Regression

Logistic Regression is a common method used for solving binary classification problems. For Logistic Regression model we have used 'liblinear' solver as it is suitable for smaller datasets.

8.4.7 Predict Target Class for New Instance of Data

8.4.7.1 Algorithm

1. Generate new instance of data by using randon available library available in python. For each of the 12 Columns generate value between min value of that attribute and max value of that attribute.
2. Randint method inside Random module is used in case of integers and Uniform method inside random module is used for floating point variables.
3. Predict target class for each of the 4 models twice ie - once without stratified K fold and again after applying stratified K fold validation. Total 8 binary outputs generated 1 each for all the models involved.
4. Target class which occurs more times in the final output array is selected as the predicted target class for the given unseen instance of data.

CHAPTER 9

RESULTS AND DISCUSSION

9.1 Time Series Analysis of Aquaculture Production

9.1.1 Forecasted Values of AquacultureProduction

On performing time series analysis of Aquaculture production in India dataset best modelturns out to be non-linear regression model having lowest RMSE value. This model has 0.989 adjusted R- squared value.

Year	Aquaculture Production (tonnes)
2019	7030200
2020	7501657
2021	8000718
2022	8528320
2023	9085402
2024	9672900
2025	10291751
2026	10942893
2027	11627263
2028	12345799

Table 9.1 : Forecasted values of Aquaculture Production in India

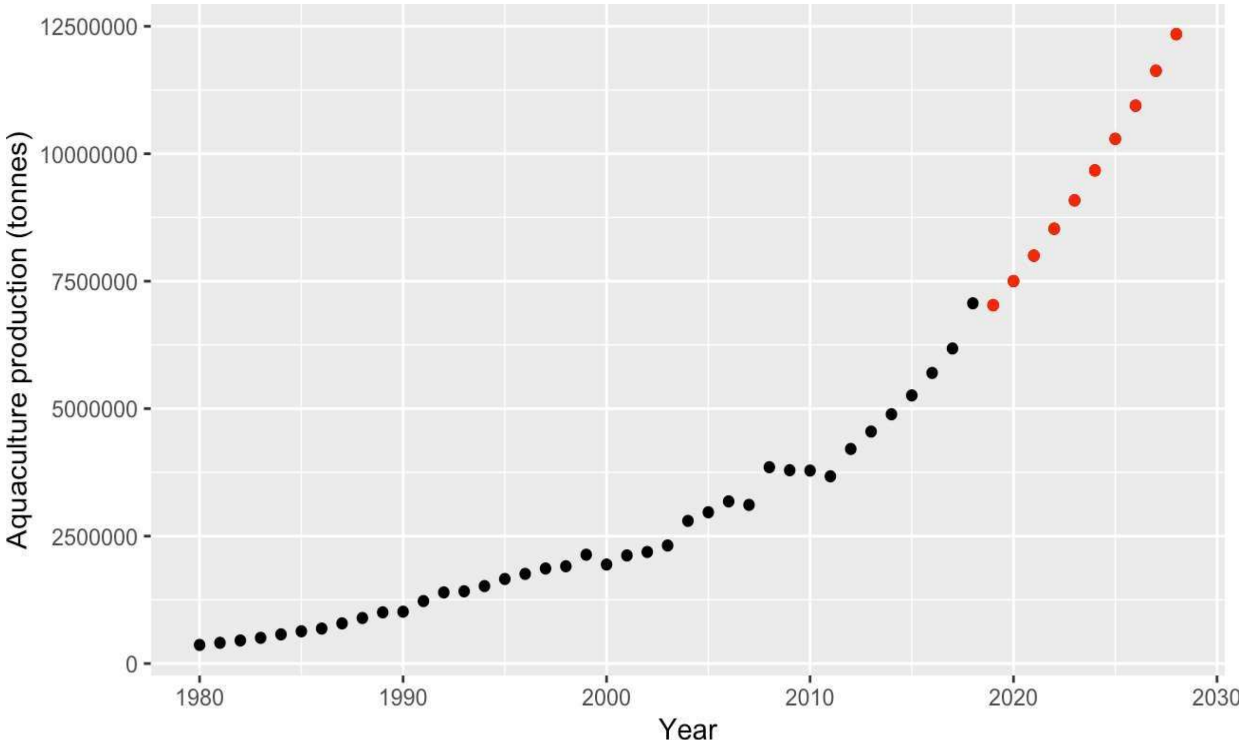


Figure 9.1 : Time series plot of actual and forecasted values of aquaculture production

9.1.2 Validation of Forecasted Values of Aquaculture Production

The performance of the best model for Aquaculture production ie- non linear regression model is validated by comparing actual values vs values predicted by the model for 5 different years. The percentage difference between actual value and predicted value turnsout to be less than 3% for 4 of the 5 years.

Year	Actual Value	Predicted Value	Percentage Difference
1989	1004500	983731.2	2.08%
1995	1658807	1697395	2.29%
2003	2315771	2383008	2.86%
2010	3785779	3848559	1.64%
2016	5700000	5416605	5.1%

Table 9.2: Comparison of actual value vs predicted value of Aquaculture production for 5 different years

9.2 Time Series Analysis of Inland Fish Price

After performing time series analysis on prices of inland aquaculture fish products in India best 2 models turn out to be ARIMA model and Holt-Winters model with trend smoothing. RMSE values of these 2 models turn out to be 25.36 and 27.76 respectively. Forecasted values from both these models turn out to be same.

Year	Inland Fish Price (Rs/Kg)
2017	452.74
2018	490.71
2019	528.68
2020	566.65
2021	604.62
2022	642.59
2023	680.56
2024	718.53
2025	756.50
2026	794.47

Table 9.3: Forecasted values of Inland fish price in India

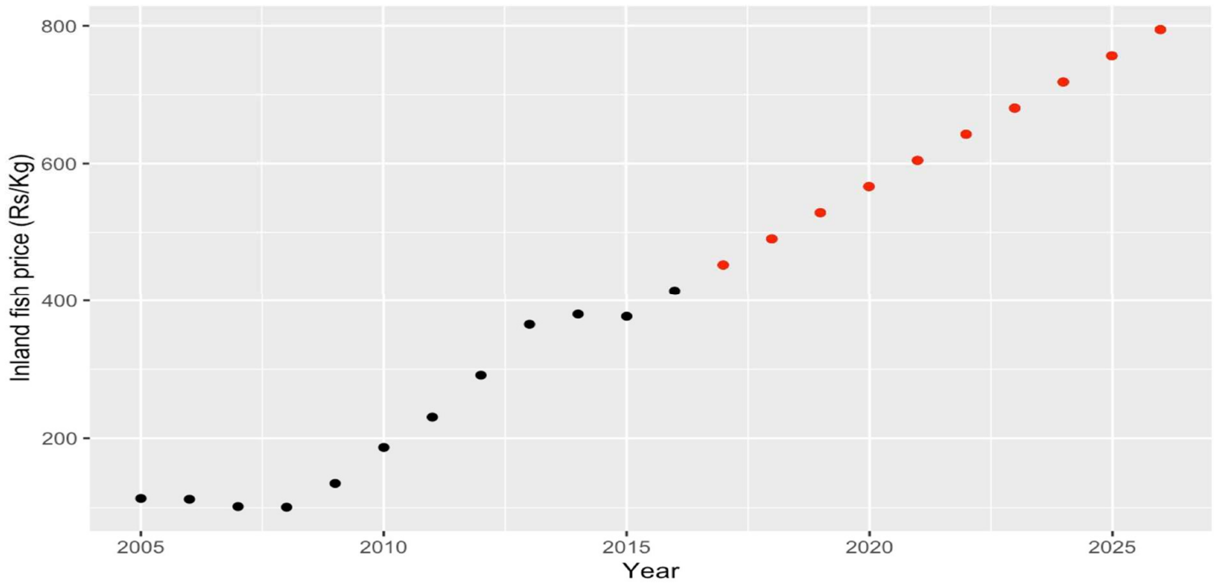


Figure 9.2 : Time series plot of actual and predicted values of inland fish price

9.3 Time Series Analysis of Marine Fish Price

After performing time series analysis in R on prices of marine aquaculture products overall the best Model turns out to be Holt Winters model with trend smoothing. This model has the lowest RMSE value among all models and the second lowest MAPE among all the models.

Year	Marine Fish Price (Rs/Kg)
2017	331.18
2018	351.38
2019	371.58
2020	391.78
2021	411.98
2022	432.18
2023	452.38
2024	472.58
2025	492.78
2026	512.98

Table 9.4 : Forecasted values of marine fish price in India

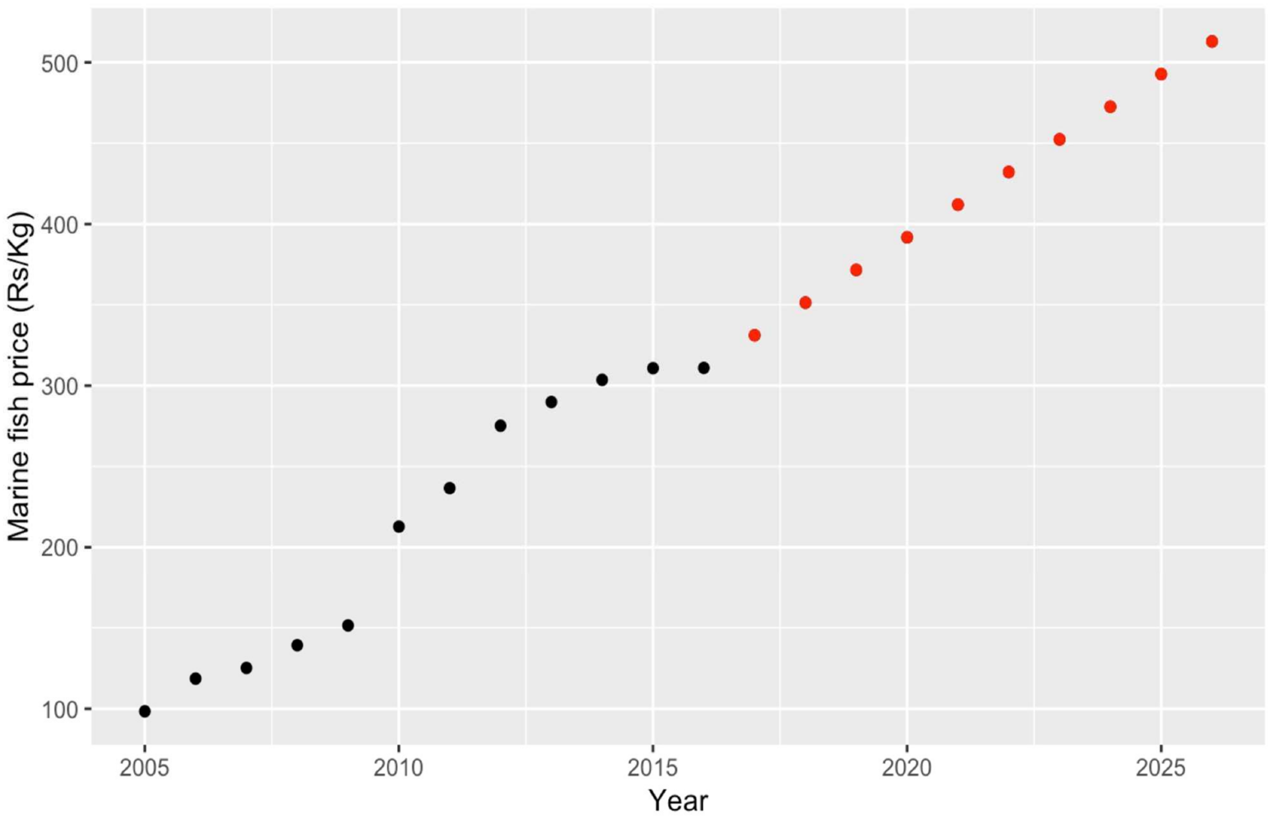


Figure 9.3 : Time series plot of actual and predicted values of marine fish price

9.4 Supervised Machine Learning for Disease Detection

9.4.1 Overview

Supervised Machine Learning technique is used to predict disease spread in aquaculture farms in six different districts of West Bengal. 4 different models implemented are Decision tree, Random forest Classifier, K Nearest Neighbour and Logistic Regression respectively. Stratified K fold cross validation method is used to increase the robustness of the models implemented.

9.4.2 Model Performance without Stratified K Fold Cross Validation

Decision Tree

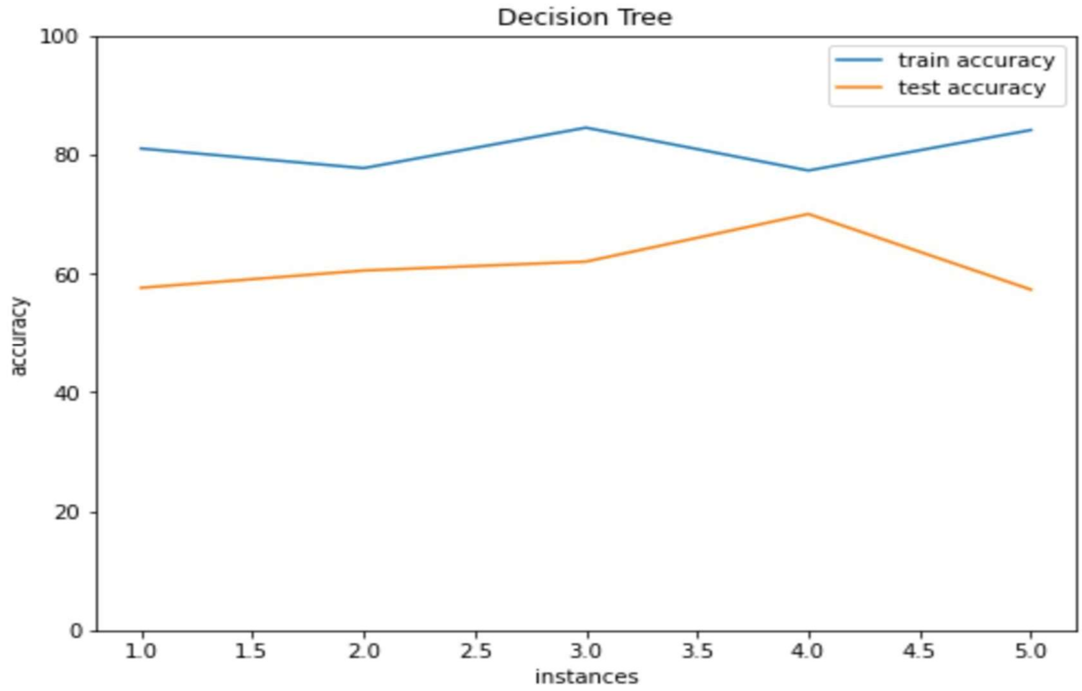


Figure 9.4: Decision Tree model without stratified k fold

Random Forest Classifier

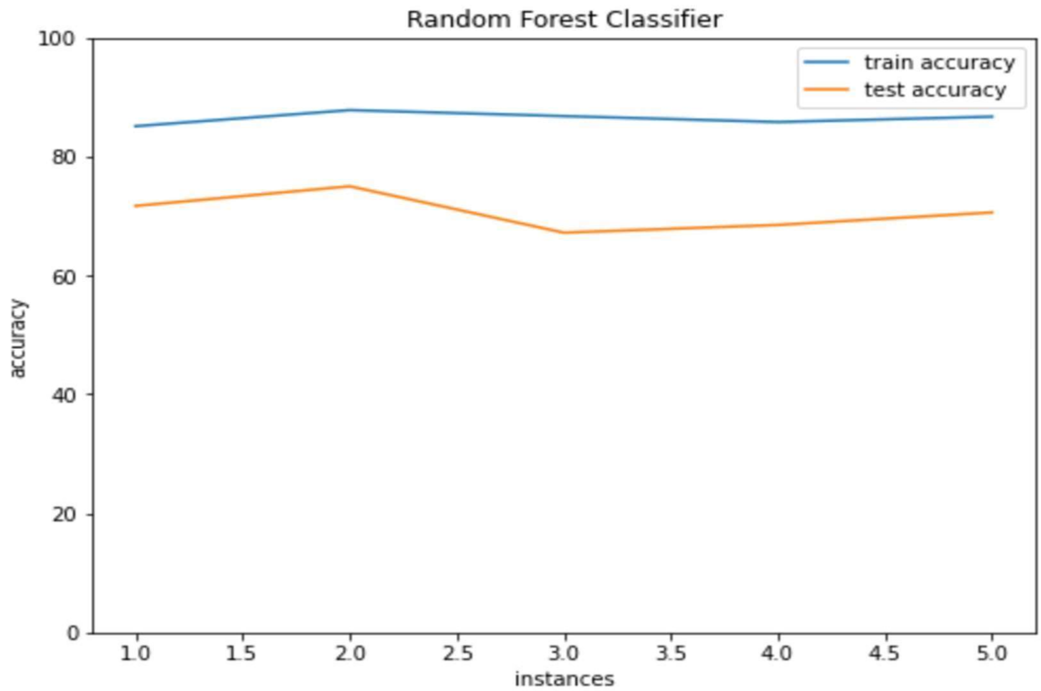


Figure 9.5: Random Forest Classifier without stratified k fold

K Nearest Neighbour

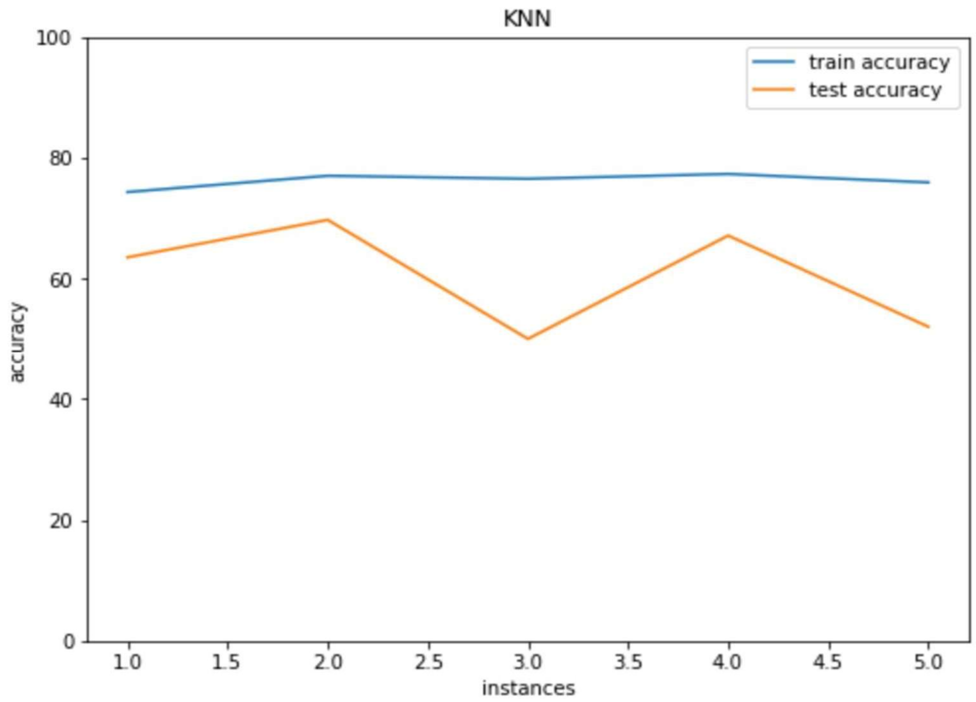


Figure 9.6: KNN model without stratified k fold

9.4.3 Model Performance with Stratified K Fold Cross Validation

Logistic Regression

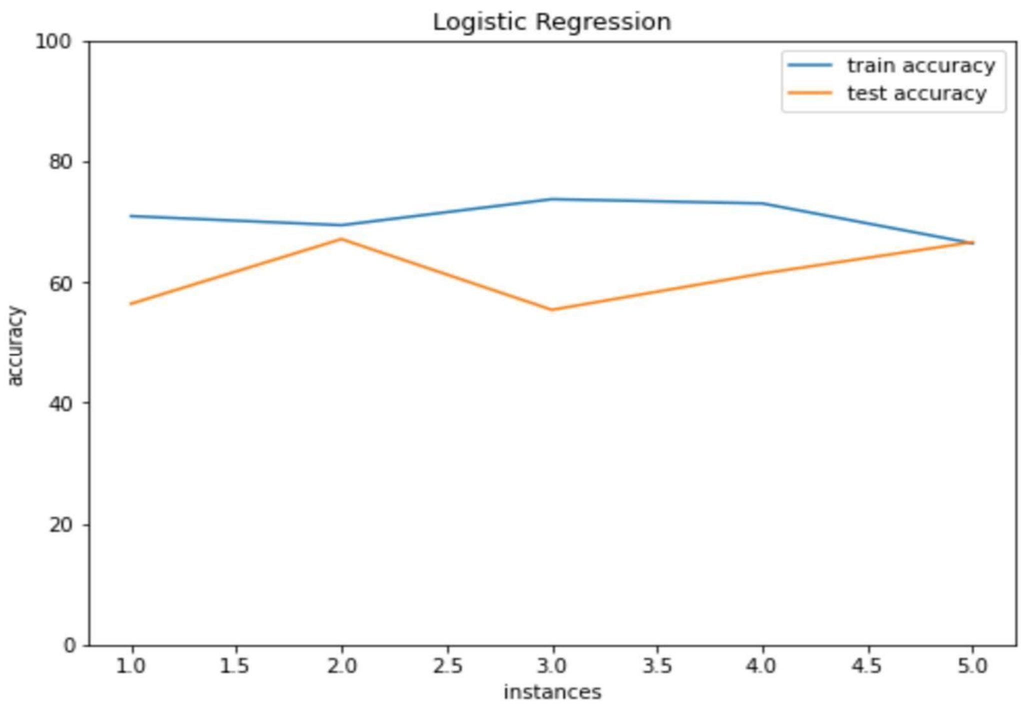


Figure 9.7: Logistic Regression model without stratified k fold

Decision Tree

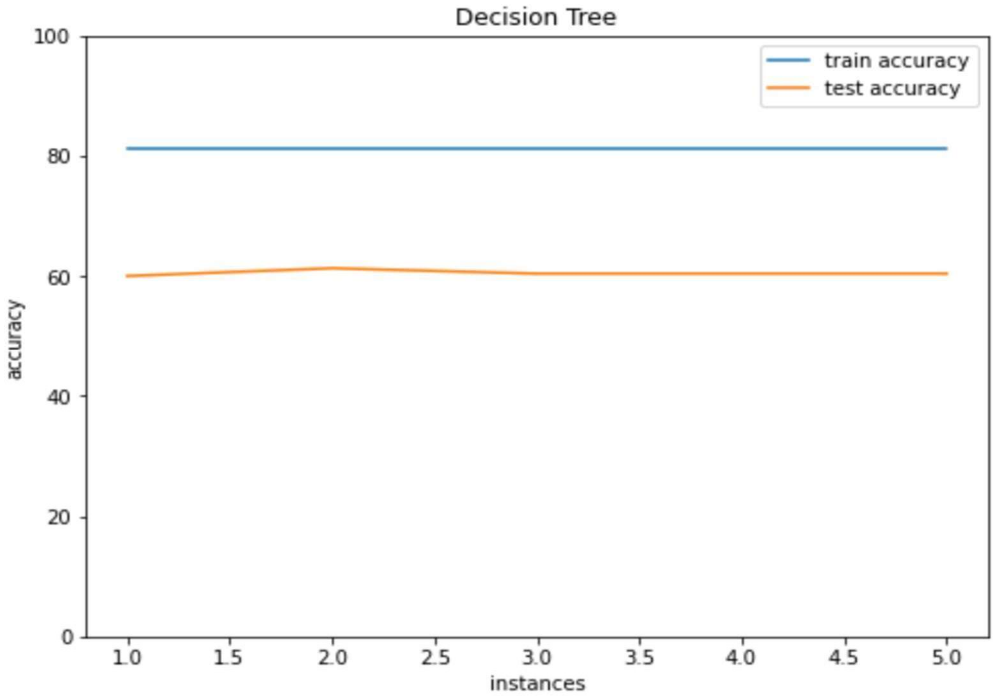


Figure 9.8: Decision Tree model with stratified k fold

Random Forest Classifier

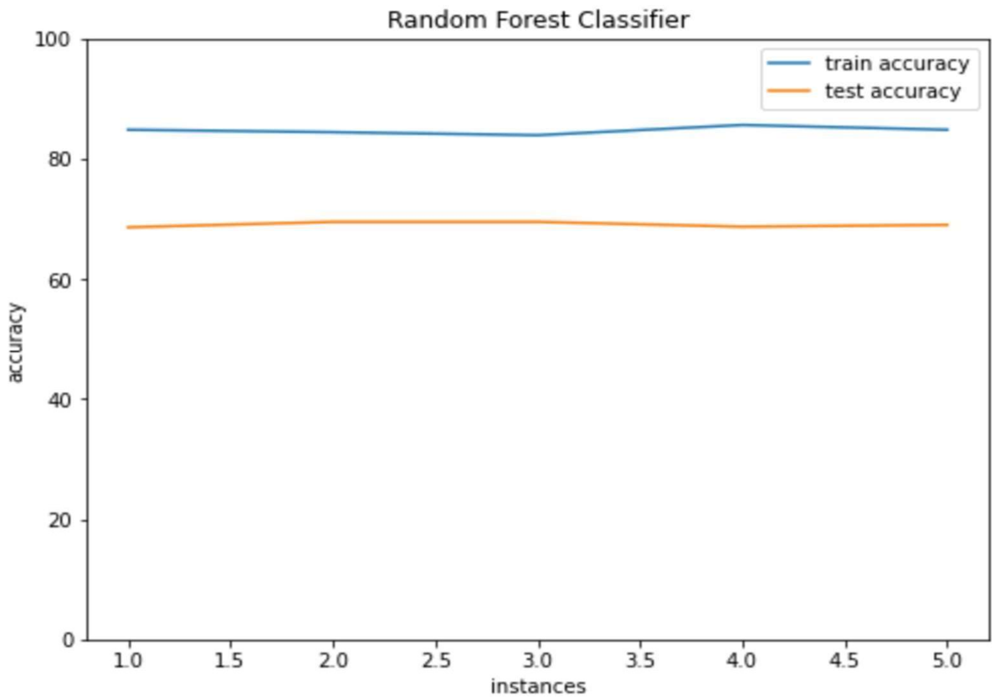


Figure 9.9: Random Forest Classifier with stratified k fold

K Nearest Neighbour

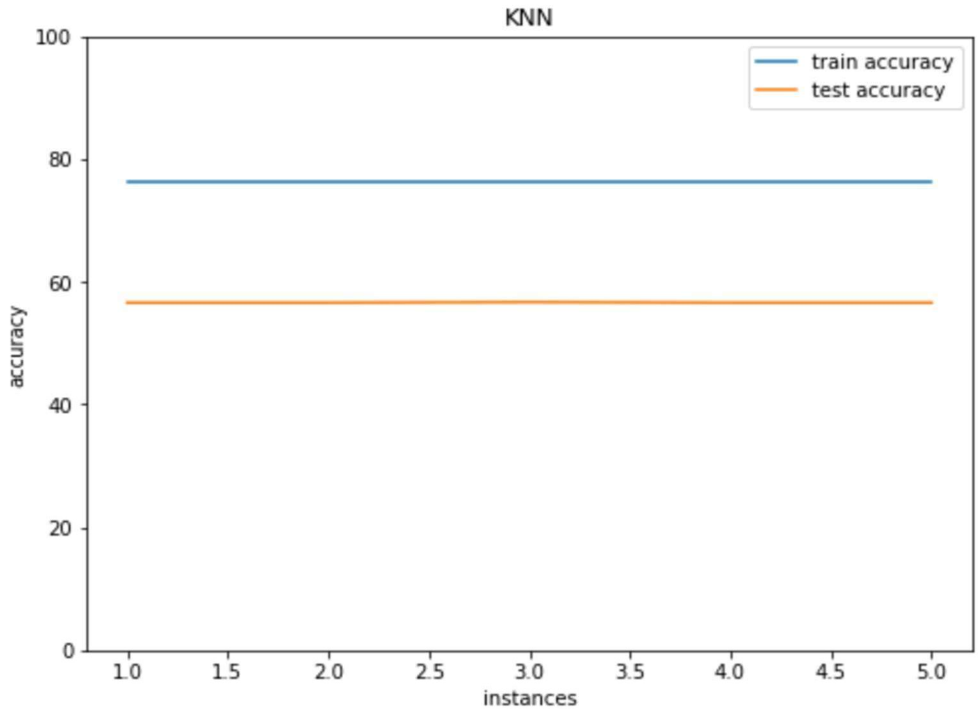


Figure 9.10: KNN model with stratified k fold

Logistic Regression

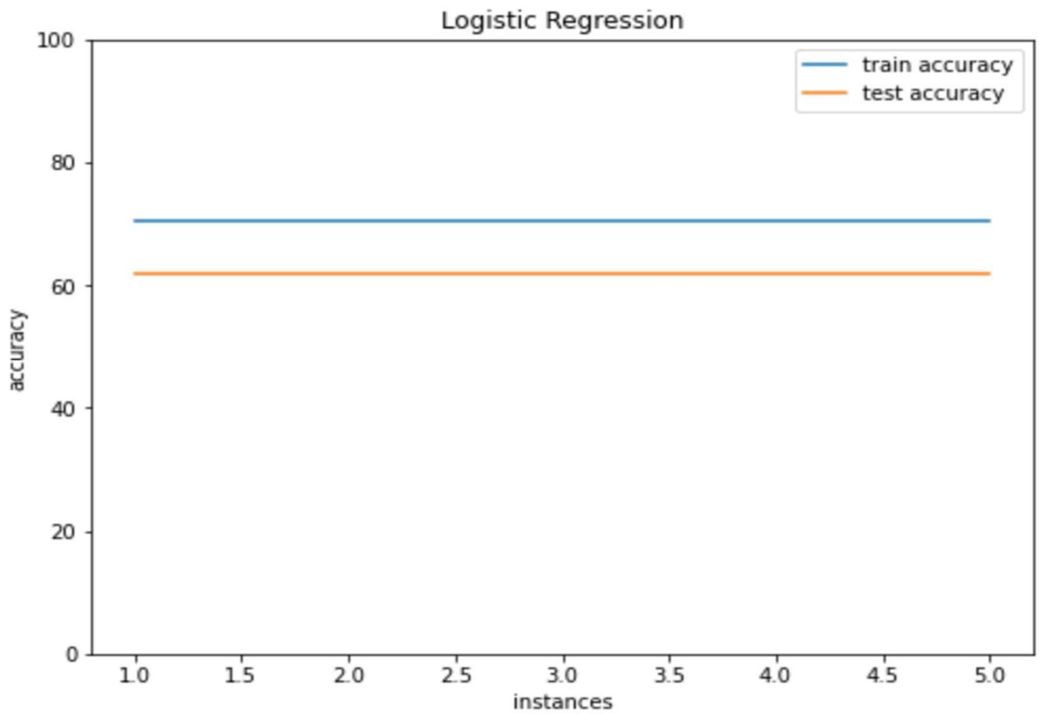


Figure 9.11: Logistic Regression model with stratified k fold

9.4.4 Inferences

- Random Forest Classifier is the best model in terms of both train and test accuracy.
- K Nearest Neighbour performs worst among all the 4 models for disease prediction.
- Adding stratified K fold cross validation reduces the variance of train/test accuracies for all 4 models.
- The gap between train and test accuracy is least for Logistic Regression indicating that logistic Regression model works well with unseen data.

9.4.5 Prediction for New Instance of Data

In total 8 models - 4 without stratified K fold cross validation and 4 with stratified K fold cross validation are used to predict target class for new instance of data. The target class value having more votes is selected as the predicted value.

```
Prediction for simple models -
Decision tree: No Disease
Random forest classifier: Disease
KNN: Disease
Logistic Regression: Disease

Prediction for models after adding stratified k fold cross validation -
Decision tree: No Disease
Random forest classifier: Disease
KNN: Disease
Logistic Regression: Disease
```

Figure 9.12 : Predicted values of target variable for new instance of data

9.5 Comparing Freshwater, Brackishwater and Marine water Aquaculture Production

9.5.1 Overview

This section is compare performance of Freshwater, marine water and Brackishwater Aquaculture production in India from year 1980 till 2018.

9.5.2 Percentage distribution of Aquaculture production in India

Year	Brackishwater	Freshwater	Marinewater
1980	1.06%	98.94%	0%
1990	3.44%	96.56%	0%
2000	4.98%	94.94%	0.08%
2010	3.15%	96.40%	0.48%
2018	11.21%	88.60%	0.18%

Table 9.5: Percentage distribution of Aquaculture production in India

9.5.3 Decade wise average growth rate in Aquaculture production

Years	Brackishwater	Freshwater	Marinewater
1980-89	69.32%	18.92%	NA
1990-99	13.83%	12.13%	NA
2000-09	14.71%	10.26%	128.7%
2010-18	70.44%	8.95%	-3.48%

Table 9.6: Decade wise rate of change in Aquaculture production based on water type

9.5.4 Inferences

- Most of the Aquaculture production in India is carried out using freshwater.
- Contribution of Brackishwater to the overall production increased by almost 4 times from year 2010 to 2018.
- Marine water Aquaculture production in India decreased from year 2010 to 2018.

CHAPTER 10

CONCLUSION AND FUTURE WORK

Forecasted values of Aquaculture production and fish prices obtained after performing time series analysis indicates steep growing trend. This will encourage more people from rural areas to get into Aquaculture and reap benefits in the long run. This will result in increase in quantity of Aquaculture production in India. In future species wise time series analysis of Aquaculture production and prices can be performed which will provide more details to fish farmer regarding which species is likely to be in high demand in 20 years time. In addition to that data analysis can be performed on export of Aquaculture products per species and per country.

Supervised Machine Learning models implemented for disease detection will help fish farmers detect spread of disease at an early stage and they can take appropriate action in order to prevent further loss. This will also help fish farmers to identify factors which caused disease spread in the first place. Future work in this regard includes generating ML models for disease detection in Aquaculture farms for all the states in this country. Factors causing disease spread vary across length and breadth of our country hence single model is not suitable in this regard. This will reduce per year economic loss due to disease spread in Aquaculture farms in India.

Aquaculture production in India and its contribution to economy via export is going to continue to increase in years to come hence the need of the hour is to make sure Aquaculture sector remains healthy in years to come.

REFERENCE/BIBLIOGRAPHY

[1] Dr. P. E. Vijay Anand, “The fish farming industry of India”, https://www.aquaculturealliance.org/advocate/the-fish-farming-industry-of-india/

[2] B. Ganesh Kumar, K. Dutta, P. K. Joshi, “Growth of fisheries and aquaculture sector in India”, https://www.researchgate.net/publication/285280707_Growth_of_Fisheries_and_Aquaculture_Sector_in_India_Needed_Policy_Directions_for_Future

[3] Marine Products Export Development Authority, “State wise aquaculture productivity”, https://mpeda.gov.in/MPEDA/cms.php?id=YXJlYS11bmRlci1jdWx0dXJl#

[4] R Sathiadas, T. M. Najmudeen, Sangeetha Pratap, “Breakaway analysis and profitability of aquaculture in India”, https://www.researchgate.net/publication/228651117_Breakeven_Analysis_and_Profitability_of_Aquaculture_Practices_in_India

[5] T. J. Abraham, S. K. Sil, P. Vineetha, “Comparative study of aquaculture practices in West Bengal and Andhra Pradesh”, https://www.researchgate.net/publication/267969481_A_comparative_study_of_the_aquaculture_practices_adopted_by_fish_farmers_in_Andhra_Pradesh_and_West_Bengal

[6] M. Krishnan, Pratap Birthal, “Aquaculture development in India: overview with special reference to coastal aquaculture”, https://www.researchgate.net/publication/247507461_Aquaculture_development_in_India_An_economic_overview_with_special_reference_to_coastal_aquaculture

[7] Ministry of fisheries, “Handbook on fisheries statistics 2018”, http://dof.gov.in/sites/default/files/2020-08/HandbookonFS2018.pdf