AI for Aquaculture in India

Sujay Gad
Dept. of Computer Science
PES University Bangalore, India
gadsujay6026@gmail.com

Mayur R B
Dept. of Computer Science
PES University Bangalore, India
rbmayur88@gmail.com

Priyadith T P
Dept. of Computer science
PES University Bangalore, India
priyadithgowda@gmail.com

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.

I.INTRODUCTION

A. BACKGROUND

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.

B. 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 d 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.

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)

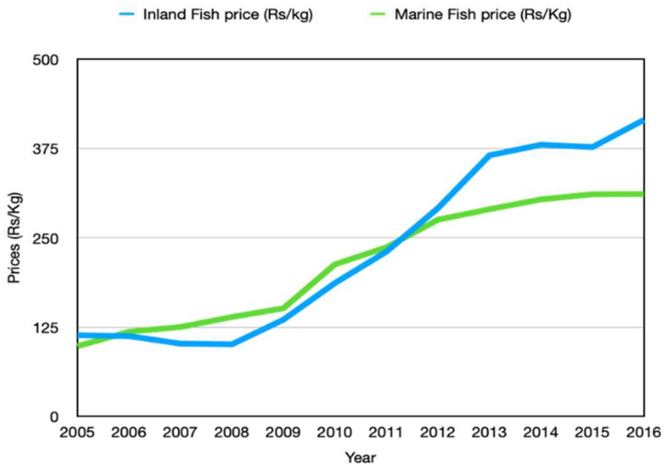


Figure 4.1: inland and marine aquaculture fish prices from 2005 till 2016

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.

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.

SYSTEM SPECIFICATION REQUIREMENTS

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.

SYSTEM DESIGN

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.

High Level System Design

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

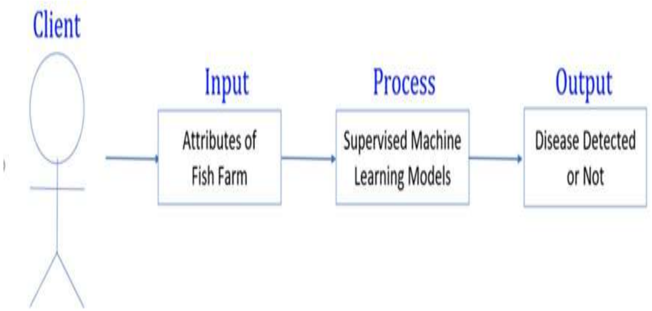


Fig 7.4: Sequential state diagram for supervised machine learning

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.

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

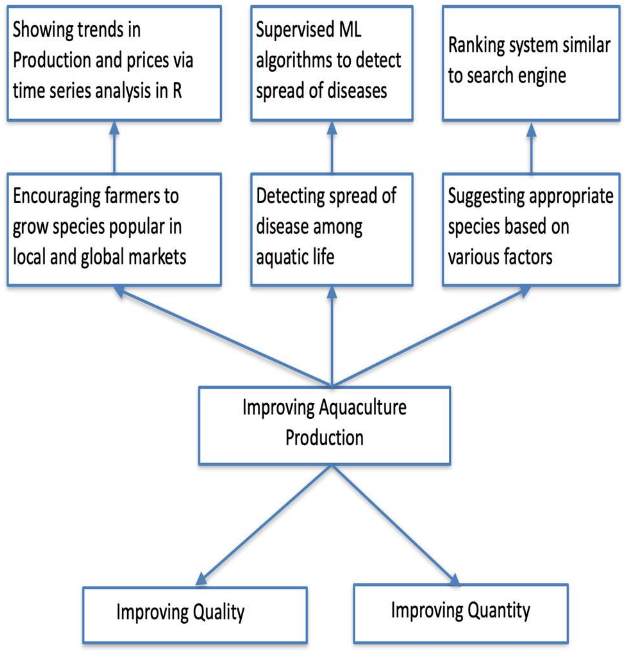
Low level design document consists of component level design process which follows step by step refinement process. LLD document contains master design diagram, use
case diagram and class diagram for each subcomponent as well as algorithm, implementation and final results for the code developed for each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



7.2.4 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.

IMPLEMENTATION AND PSEUDOCODE

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

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

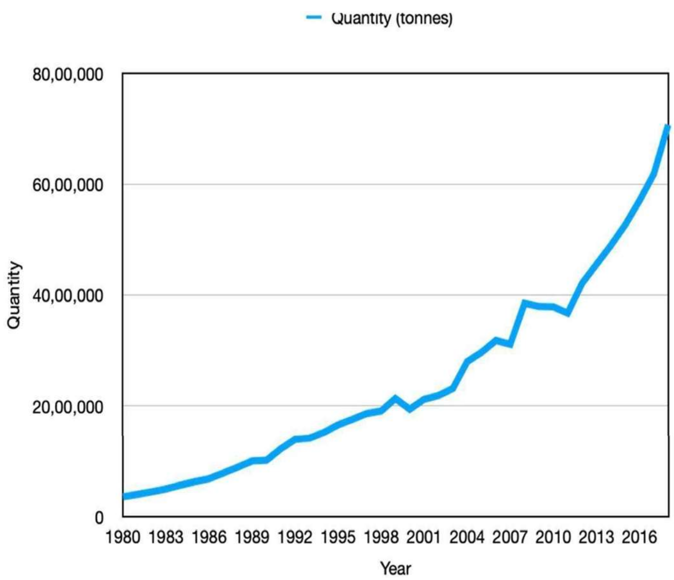


Figure 8.2: Aquaculture production year wise plot

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.

Time Series Analysis of Prices of Aquaculture Products

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

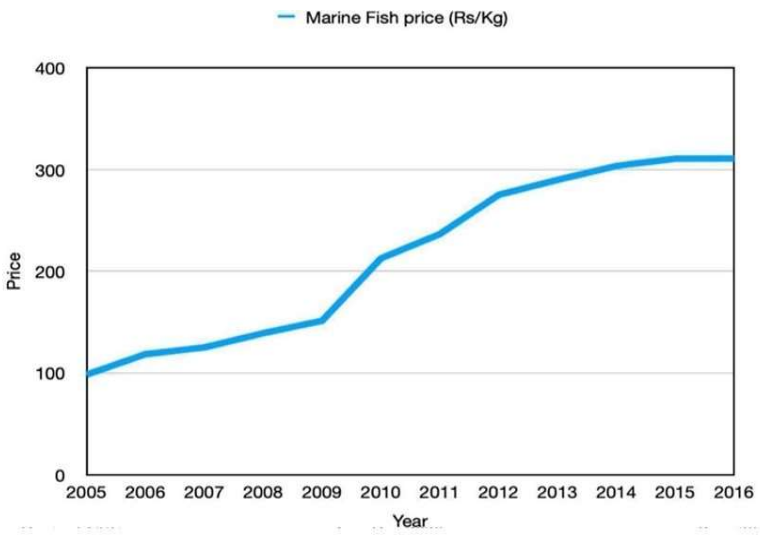


Figure: Price of marine fish in Rs/Kg year wise

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.

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.

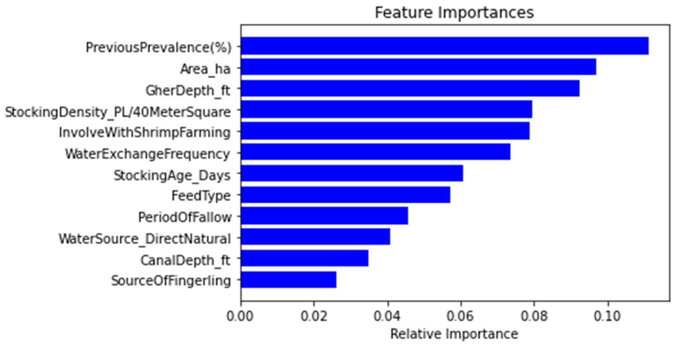


Figure 8.6: graph depicting top 12 features and their respective importance

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.

RESULTS AND DISCUSSION

Time Series Analysis of Aquaculture Production

9.1.1 Forecasted Values of Aquaculture Production

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

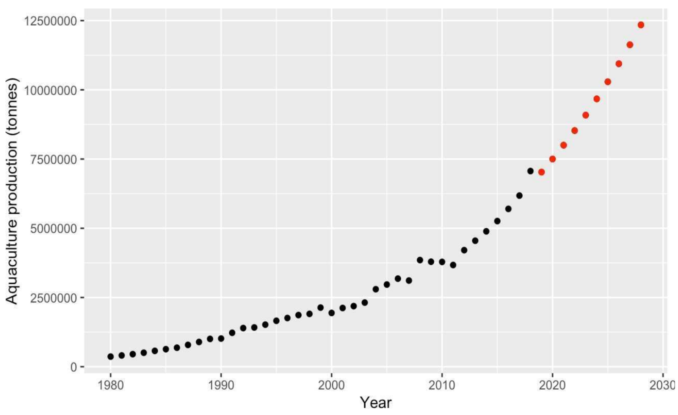


Figure 9.1 : Time series plot of actual and forecasted values of aquaculture production

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

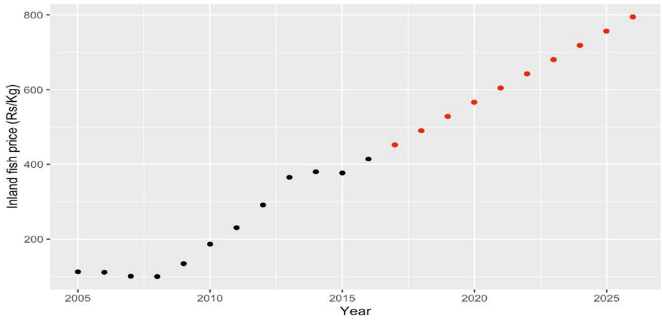


Figure 9.2 : Time series plot of actual and predicted values of inland fish price

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.

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_Aquacultur e_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