UE17CS490B – Capstone Project Phase – 2

SEMESTER - VIII

END SEMESTER ASSESSMENT

Project Title : AI for Aquaculture
Project ID : PW21UD03
Project Guide : Dr Uma D
Project Team : 177_714_1324

Outline

- Abstract
- Team Roles and Responsibilities.
- Summary of Requirements and Design (Capstone Phase - 1)
- Summary of Methodology / Approach (Capstone Phase - 1)
- Design Description
- Modules and Implementation Details
- Project Demonstration and Walkthrough
- Test Plan and Strategy
- Results and Discussion
- Lessons Learnt
- Conclusion and Future Work
- References

Abstract

- Using AI to improve quality and quantity of production of Aquaculture in India.
- Time series Analysis is used to predict prices of Aquaculture products and also overall production in near future in India.
- Using AI to detect spread of diseases in organisms in order to benefit the fish farmers.
- Suggest appropriate species to fish farmers on the basis of factors such as capital investment available, weather conditions, water type available(freshwater vs brackish water) as well as the overall profits.
- We will also suggest ways to improve Aquaculture production in India in future.

Team Roles and Responsibilities

Tasks carried out by individual members during capstone 1.

NAME	TASKS
SUJAY	• Added 3 models for time series analysis of Aquaculture production in India. • Analysis of factors which determine whether to go for Freshwater or Marinewater Aquaculture. • Worked on design details and high level design diagram, state sequence diagram which were added to HLSD.
MAYUR	• Worked on time series analysis of prices of marine water Aquaculture products, performed preprocessing of Aquaculture production dataset and added 2 models for time series analysis of Aquaculture production. • Analysis of 2 research papers as part of literature survey carried out. • Added introduction, problem statement sections to the final report.
PRIYADITH	• Worked on time series analysis of prices of freshwater Aquaculture products and added 4 models for time series analysis of Aquaculture production. • Analysis of 2 research papers as part of literature survey carried out. • Prepared project requirements specification document.

Team Roles and Responsibilities

Tasks carried out by individual members during capstone 2.

NAME	TASKS
SUJAY	• Implemented 4 supervised machine learning models in order to perform virus detection in Aquaculture farms. • Testing/validation carried out of the best model found in the time series analysis performed. • Prepared low level design document.
MAYUR	• Reduced number of attributes in disease detection dataset by removing columns having high % of missing values, columns which have high correlation with other attributes in the dataset. • Analysis of 2 research papers as part of literature survey carried out. • Results and discussion, conclusion and future work sections added to final report.
PRIYADITH	• Added functionality to predict target class for new instance of data, graphs added to compare and contrast performance of the 4 models implemented. • Analysis of 2 research papers as part of literature survey carried out. • Comparison of performance of freshwater, Brackishwater and marine water Aquaculture in India for the last 40 odd years.

Summary of Requirements and Design

Design Constraints, Assumptions & Dependencies

- Interoperability Requirements
  - For Time Series Analysis system should have R, Rstudio installed. R libraries such as series, forecast must also be downloaded.
  - For Supervised Machine Learning Anaconda must be downloaded to run Jupyter Notebook with Python.

- Data Repository and Distribution Requirements
  - Aquaculture Production and Prices of Aquaculture products dataset is collected from Ministry of Fisheries handbook.
  - Dataset for disease prediction is collected from data.world.
  - Dataset for comparison of performance between freshwater, marinewater and brackishwater aquaculture production is taken from Food and Aquaculture Organisation of UN website.

Summary of Requirements and Design

Design Constraints, Assumptions & Dependencies

- Discuss Performance related issues as Relevant
  - Performance of the design will depend on the performance of the models which will use. Good time series model should be low RMSE, MAPE value, p<=0.05 and high R-square value.
  - Supervised Machine learning models implemented should not overfit and false positives and false negatives should be less in the confusion matrix.

- End User Environment
  - Client will be able to enter values related to his fish farm and model will give binary output suggesting virus detected or not.

Summary of Requirements and Design

Design Constraints, Assumptions & Dependencies

- Hardware Requirement
  - System must be able to run all software applications required.

- Assumptions
  - Assumptions include the fact that datasets collected are trustable and can be used to perform the required analysis.

- Dependencies
  - There are no dependencies involved between subsystems of the design.

Summary of Methodology / Approach

Design Details

- Novelty
  We are using known computer science technologies in a sector in which they haven't been used before and there is huge scope for using modern technologies to improve aquaculture production.

- Reliability
  The design is highly reliable since there is no chance of fault being developed over time.

- Maintainability
  Easy to maintain as there is no maintenance cost in terms of time, effort and money.

Summary of Methodology / Approach

Design Details

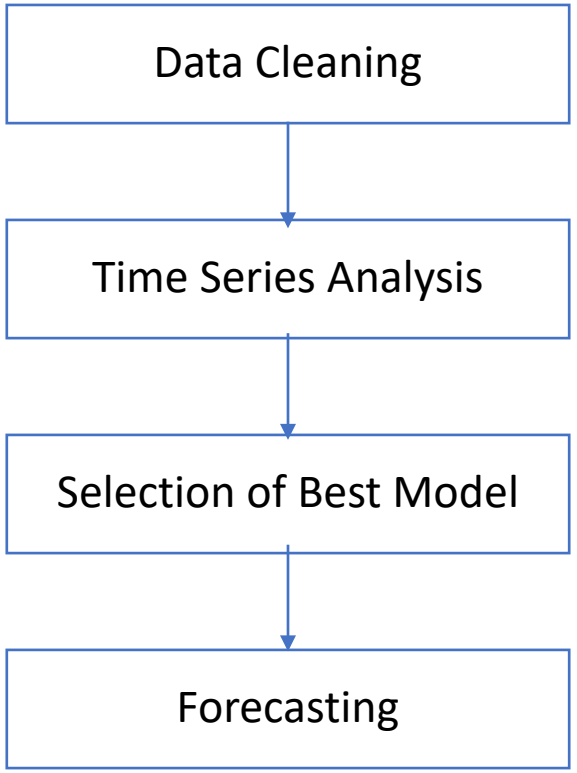
- Portability
  There is no issue in terms of portability since design can be moved from one system to another or from one OS to another.

- Reusability
  Subcomponents of the product can be used in other cases hence design is reusable.
  Eg - ML models to detect spread of disease can be used in other scenarios

Summary of Methodology / Approach

Architecture

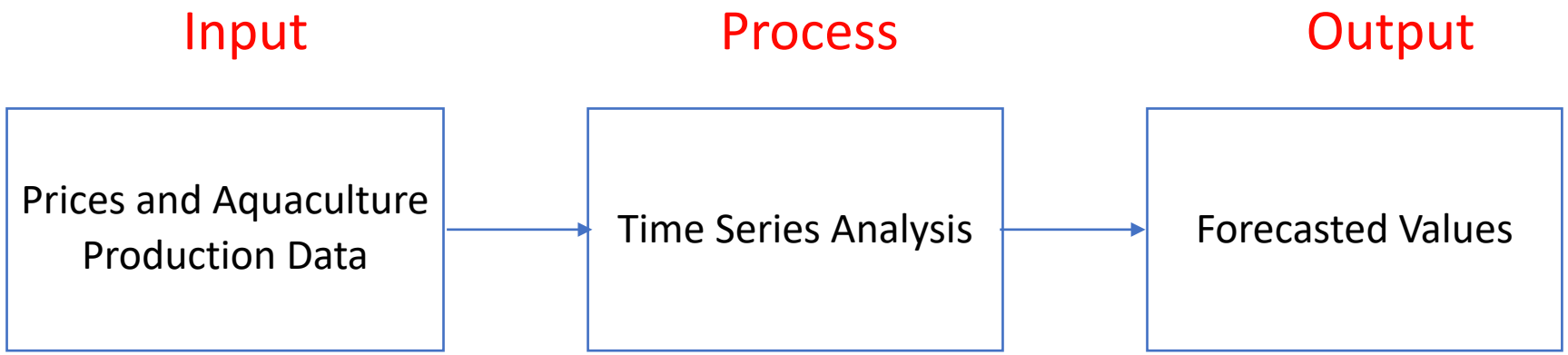
- Time Series Analysis



Summary of Methodology / Approach

Architecture

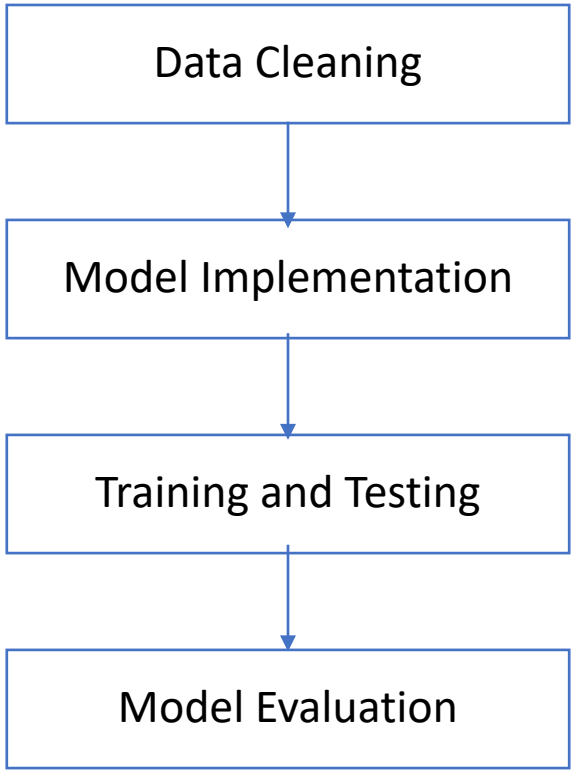
- State Sequence Diagram



Summary of Methodology / Approach

Architecture

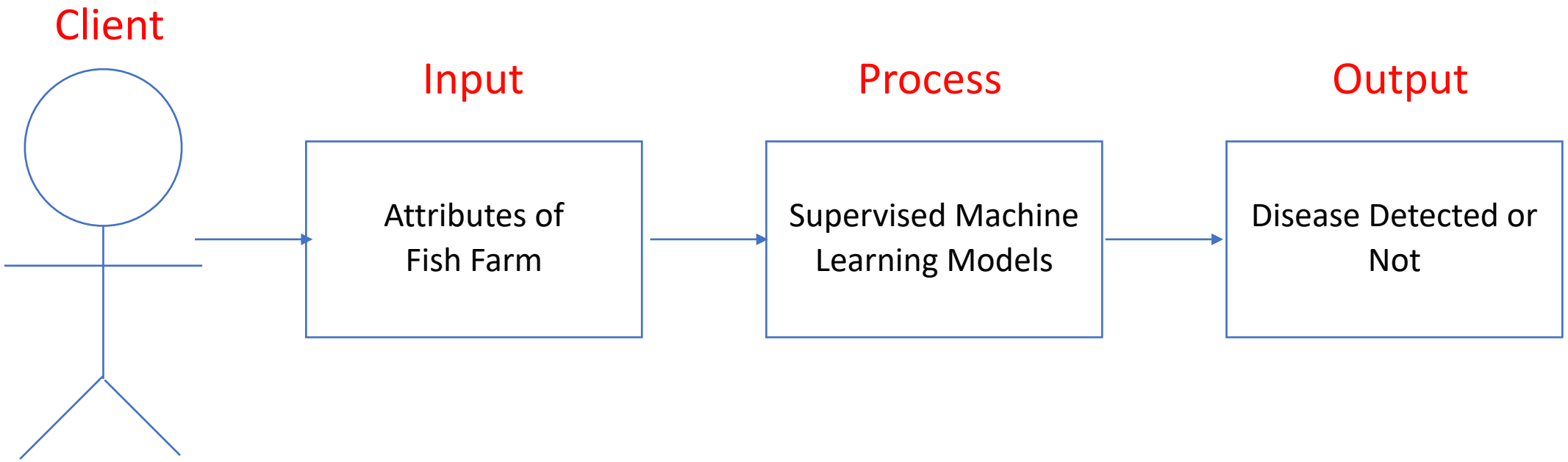
- Supervised Machine Learning



Summary of Methodology / Approach

Architecture

- State Sequence Diagram



Summary of Methodology / Approach

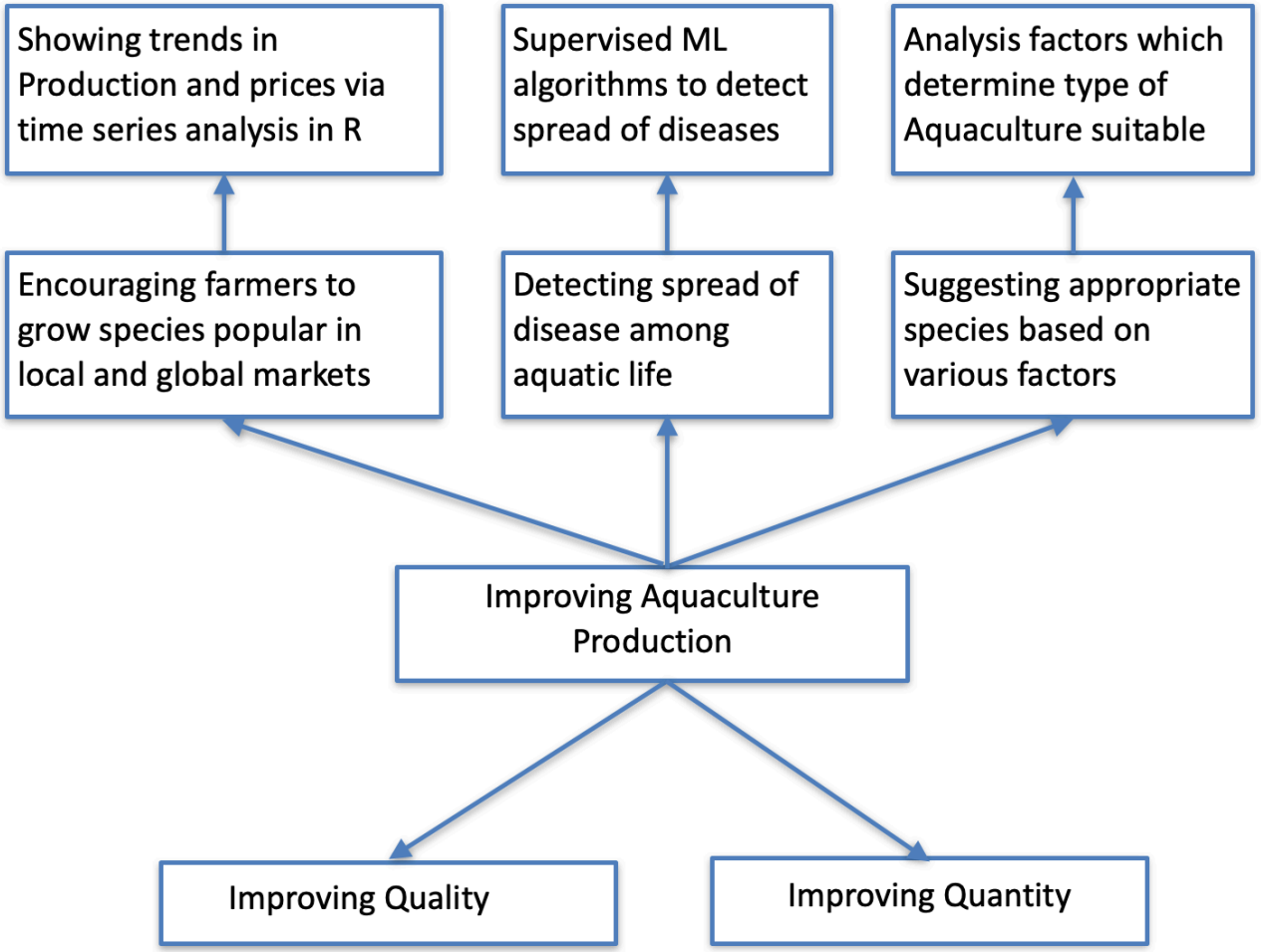
Technologies Used

- R, Studio
  Used to perform time series analysis to forecast Aquaculture production and prices of Aquaculture products. Following R libraries have been used.
    - ggplot
    - forecast
    - knitr
    - tseries

- Python, Jupyter Notebook
  Used to implement models for disease detection in Aquaculture farms. Following Python have been used
    - random
    - sklearn
    - numpy
    - pandas
    - matplotlib.pyplot

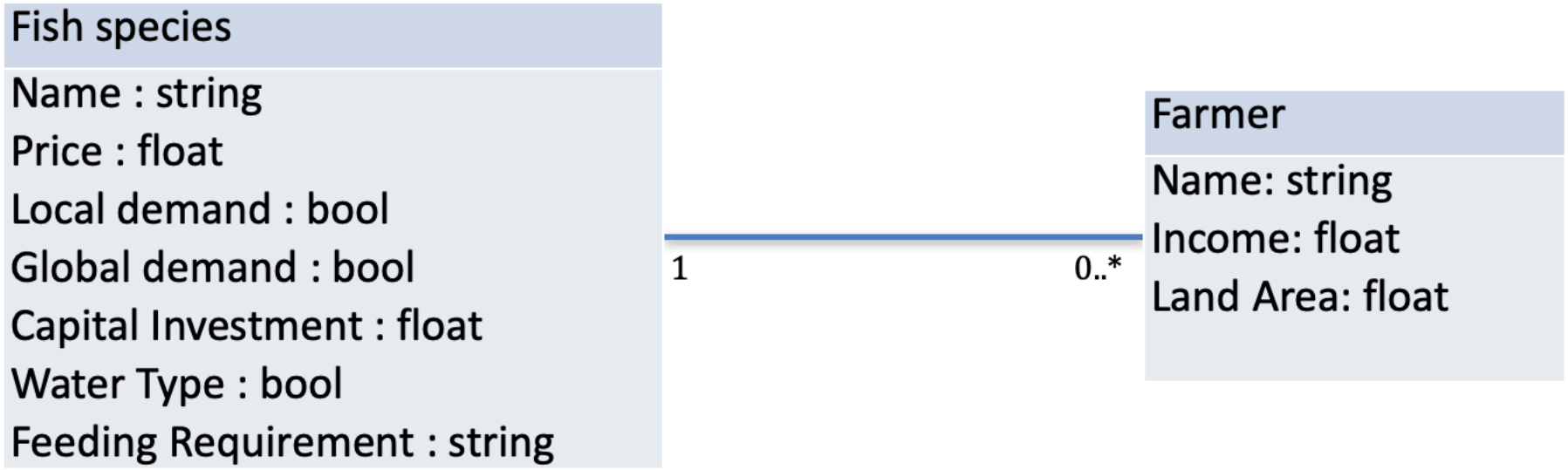
Design Description

Master Design Diagram



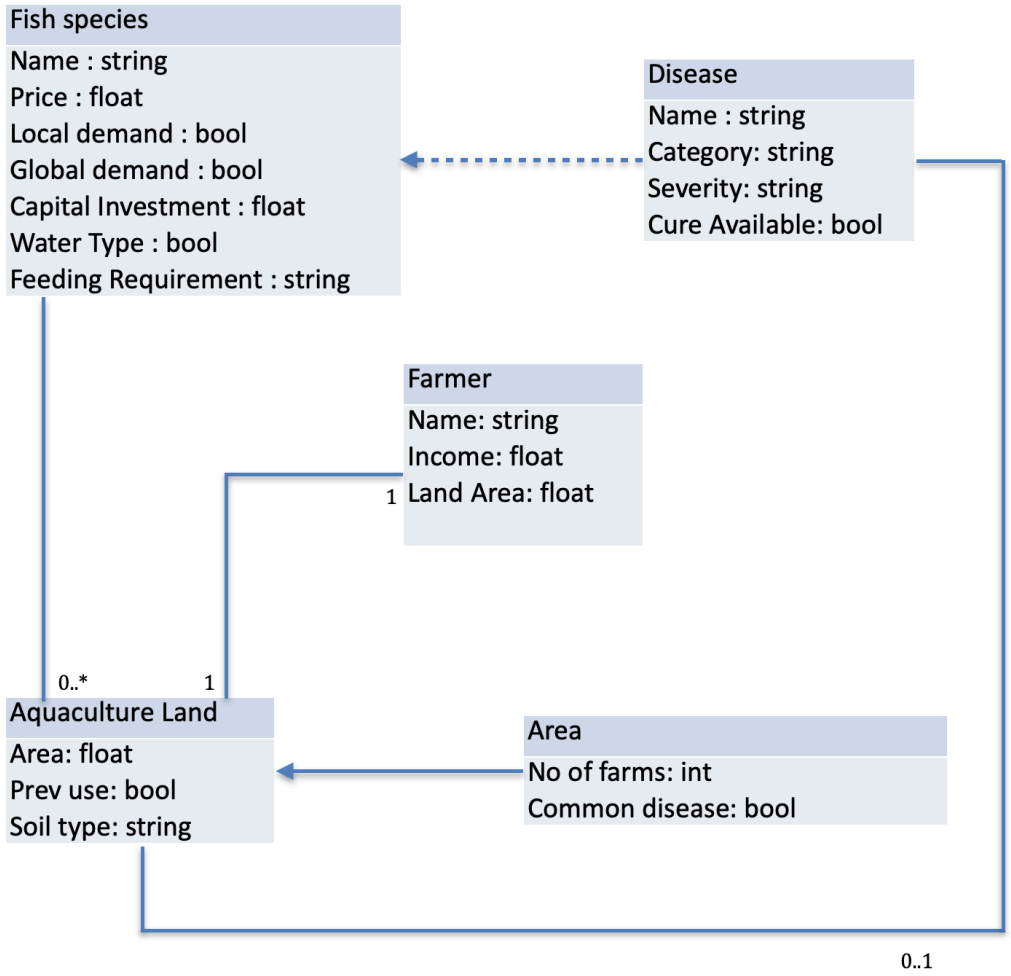
Design Description

Class Diagram for Time Series Analysis



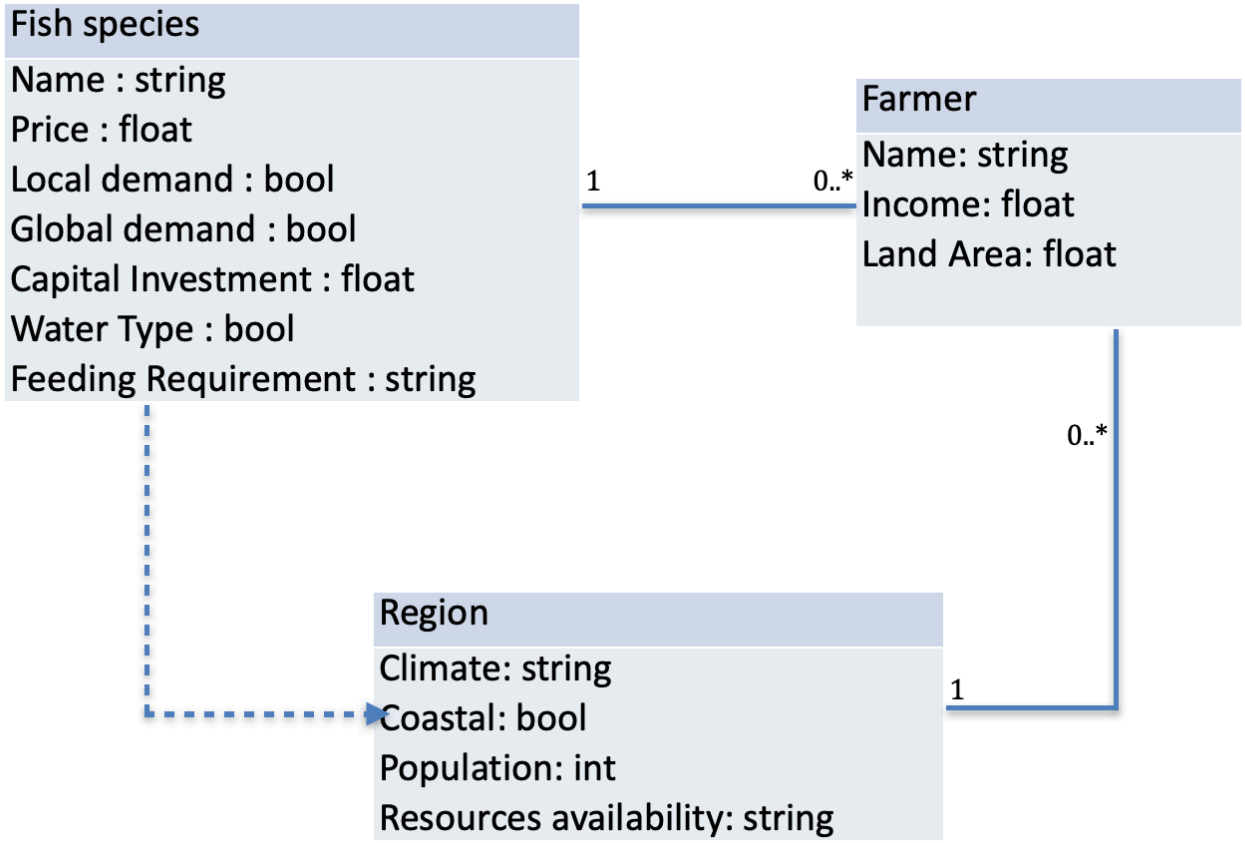
Design Description

Class Diagram for Disease Detection



Design Description

Class Diagram for Suggesting Appropriate Species



Modules and Implementation Details

Time Series Analysis for forecasting Aquaculture production and prices

- Dataset is collected from ministry of fisheries website. 40 years of Aquaculture production values is used to forecast for the next 10 years, 20 years of Aquaculture product prices are used to forecast for the next 5 years.
- Data preprocessing involves convert Aquaculture production column from string to int, converting non stationary time series to stationary time series.

Modules and Implementation Details

Time Series Analysis for forecasting Aquaculture production and prices

- 10 different models are generated by performing time series analysis in R.
- Best model is selected on the basis of RMSE/MAPE and adjusted R-squared value.
- The performance of the best model is validated by comparing actual value vs predicted value for multiple different years.

Modules and Implementation Details

Disease detection using Supervised Machine Learning Algorithms

- Dataset is taken from data.world website and consists of details about Aquaculture performed by farmers across 6 different districts of West Bengal.
- Attributes which do not affect target variable, have high percentage of missing values are removed. Attributes having high correlation with some other variables in the dataset are also discarded. At the end top 12 features are selected.
- Dataset is divided into training and testing set using 70-30 ratio.

Modules and Implementation Details

Disease detection using Supervised Machine Learning Algorithms

- Stratified K fold cross validation method is used to remove bias of the dataset and make training and testing set target class distribution similar.
- Decision tree, Random forest classifier, KNN and Logistic regression are the 4 models used in order to perform the required analysis.
- Performance of all 4 models is compared using graphs.
- Functionality is added to predict target class for new unseen instance of data.

Modules and Implementation Details

Suggesting appropriate species to fish farmer

- Analysis of factors which affect whether to go for freshwater, marine water or Brackishwater Aquaculture.
- Comparison of performance of freshwater, brackishwater and marine water Aquaculture production in India for the last 40 odd years.

Project Demonstration

Demonstration of the complete project.

Test Plan and Strategy

- Performance of the best model for Aquaculture production is validated by comparing actual value vs predicted value for 5 different years.
- Validation is also done by calculating R-square for best model in each category.

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%

Results and Discussion

Time Series Analysis for forecasting Aquaculture production and prices

- Non linear regression model turns out to be the best model for forecasting Aquaculture production in India. ARIMA model and Holt Winters for trend smoothing turn out to be the best models for forecasting prices of inland and marine fish price.
- All 3 models have R-squared value greater than 0.9.
- Forecasted values confirm the hypothesis that Aquaculture production and prices of Aquaculture products are set to grow at a fast rate in coming years.

Results and Discussion

Disease detection using Supervised Machine Learning Algorithms

- Random Forest Classifier turns out to be the best model for disease detection in Aquaculture farms. Accuracy on test set for this model turns out to be around 70%.
- This will help reduce economic loss due to disease spread by certain proportion.

Results and Discussion

Suggesting Appropriate Species to fish farmer

- Local climate, resources available, demand, environmental impact are some of the factors in deciding whether to go for freshwater, marine water or brackishwater Aquaculture.
- Freshwater Aquaculture has dominated percentage distribution of Aquaculture production in India but brackishwater Aquaculture has made a big jump in the last decade.

Documentation

Following documentation has been done

- Report
- Poster
- Video
- IEEE draft
- Github Repository

  Link - https://github.com/Sujay2611/Capstone

Lessons Learnt

- More detailed analysis work would have been possible if the datasets contained more in depth information.

Conclusion and Future work

- Increasing trend in Aquaculture production and prices of Aquaculture products will encourage more people to take up fish farming in near future thereby increasing quantity of Aquaculture production in India.

- Future work which can be carried out in this regard is performing data analysis of species wise Aquaculture production export which will help fish farmers maximise their profits and also boost the economy.

Conclusion and Future work

- Performing supervised machine learning for disease detection will help fish farmers reduce losses incurred due to disease spread in aquaculture farms and also identify factors which resulted in disease spread such that situation is not repeated in future.

- Future work can include implementing different models for disease detection for all states in which Aquaculture is carried out at a large scale so that economic loss due to disease spread in Aquaculture farms can be reduced country wide.

References

- Dr. P. E. Vijay Anand, “The fish farming industry of India”, https://www.aquaculturealliance.org/advocate/the-fish-farming-industry-of-india/
- 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
- Marine Products Export Development Authority, “State wise aquaculture productivity”, https://mpeda.gov.in/MPEDA/cms.php?id=YXJlYS11bmRlci1jdWx0dXJl#
- R Sathiadas, T. M. Najmudeen, Sangeetha Pratap,“Break away analysis and profitability of aquaculture in India”, https://www.researchgate.net/publication/228651117_Breakeven_Analysis_and_Profitability
- Ministry of fisheries,“Handbook on fisheries statistics 2018”, http://dof.gov.in/sites/default/files/2020-08/HandbookonFS2018.pdf

References

- 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
- James Diana et al, “Responsible Aquaculture in 2050” , https://academic.oup.com/bioscience/article/63/4/255/253239
- Ayalew Aseefa and Fufa Abunna, “Maintenance of fish health in aquaculture”, https://www.hindawi.com/journals/vmi/2018/5432497/
- Department of Computer Science Western Michigan University, “Low level design document”, https://nest.cs.wmich.edu/report/AGORA-Low-Level-Design-Document-rev2.pdf
- Food and Agriculture Organisation of UN, “National Aquaculture sector overview: India” , http://www.fao.org/fishery/countrysector/naso_india/en

Thank You