INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN ACCOUNTING, FINANCE AND MANAGEMENT SCIENCES

OPEN ACCESS JOURNAL

Vol 14, Issue 4, (2024) E-ISSN: 2225-8329

Using Machine Learning to Identify Factors of Importance in Aquaculture Management in Malaysia

Sulochana Nair, Sagaran Gopal, Bipinchandra Mavani

Binary University of Management& Entrepreneurship, Malaysia

To Link this Article: http://dx.doi.org/10.6007/IJARAFMS/v14-i4/23555 DOI:10.6007/IJARAFMS/v14-i4/23555

Published Online: 06 November 2024

Abstract

The aquaculture business has expanded over the years in Malaysia. It provides a source of income and self-employment for many small-time players besides providing the protein requirement and a contributor to food security. However, the sustenance of this important sector is at stake as many small-timers are not making sufficient profits, citing increasing costs and dumbing from neighbouring countries. Hence, the focus of the study was to investigate various factors that influence this sector and identify important factors to help make this sector a viable and profitable venture in Malaysia. The various factors for the investigation were identified from the literature as influencers, and through the use of machine learning techniques on logit regression, significant factors were identified in the local context. The responses for the factors were adduced from a questionnaire survey directed at 268 farm operators and owners of aquaculture. The data collected from this survey form the primary input for the analysis. The important factors identified as drivers of Profitability are the Provision of Extension Services, Climate Change, Innovative Technologies, Farming Practices, Institutional Influences, Environmental, Learning and Development, and Economics. However, Societal, Supply Chain, Risk Management Culture, and Feed are not significant drivers in the local context.

This study is limited in scope as it involves only the farm operators and owners and no other stakeholders and uses a single model in the machine learning process namely logit regression. However, findings are not diminished in the sense that the factors identified can be a source of input for policymakers in the local context for any future blueprint for this sector.

Keywords: Aquaculture, Profitability, Machine Learning, Logit Regression

Introduction

Aquaculture farming involves 628 species, that are endemic, native, introduced, and reintroduced. However, the Department of Fisheries Malaysia has concentrated its effort on red tilapia and catfish as well as giant freshwater prawn to be farmed throughout Malaysia as they view it as more sustainable and lucrative. Aquaculture business in Malaysia was once a robust industry, showing an annual growth of 10% with Malaysia being a net exporter of the surpluses to countries like Singapore and Indonesia. The Department of Fisheries has gone to great lengths to see the growth and viability of this sector, through its various activities from how to start a farm, to the best breeding methods, funding, training, and support system (Department of Fisheries Malaysia, 2024). Yet, there are rumbles of discontent, from small timers, of dumping from Thailand, cost issues, labour, land, and, the competition that affect their profitability and sustenance. There are some indicators in this direction, as RAS Aquaculture (2023) has pointed out, that we need to be worried about the drop in current production capacity. There is a high likelihood that if there are real issues and if left alone, can lead to the closure of many farms and the effort of the government this far is wasted. Therefore, it is important to investigate these issues in perspective as the industry is an important source of food security for the nation. There are broader factors that can play an important role in the success of this industry. This study will extract significant factors that are central to the strengthening and success of the industry. It can be useful input for any blueprint for managing and developing this industry further in Malaysia.

Problem Statement

The aquaculture business is important in Malaysia as it provides self-employment opportunities to many youths and rural populations with a reasonable income, besides being a source of protein and food security. Of late, the sustenance of this important sector is at stake as many small-timer operators are not making sufficient profits due to increasing costs and dumbing from neighbouring countries. However, the success of this sector involves many variables as identified from the literature to be significant influencers. Hence, the research intends to investigate which of these factors from literature reviews are important in the local context. The factors identified can be a source of input for policymakers in the local context for any future blueprint for the development of this important sector.

Limitation

This study is limited in scope as it involves only the farm operators and owners from key states in Malaysia who operate freshwater aquaculture practices on a small scale and no other stakeholders' feedback is taken. Besides only a single model, logit regression is used in the machine learning process namely logit regression to extract factors of importance.

Literature Review

Aquaculture has expanded rapidly over the years contributing significantly to food security and protein sources that are localized (Galappaththi et al., 2020) with more than 400 aquatic species found throughout the world (ibid). Entry into the business of aquaculture has fewer barriers, as prospects learn fast from friends, family members, and neighbours to indulge in this activity, as an Indonesian study pointed out. However, initial financing and sustaining over time becomes an issue when cost and other issues come into play which needs to be looked into for the successful development of this sector (Miyata and Manatunge, 2004). Cost sinks are found across all stages, and become crucial over the years as healthy harvest requires continuous monitoring of the environment (Wang and Mendes, 2022), feeding habits, and necessary nutrients (Zhou et al., 2003; Jegede and Olorunfemi, 2013; Daudpota et al., 2016; Ludson et al., 2020) together with other factors. According to Wang and Mendes (2022), in China, the income of these farmers is closely linked to the environmental factor. Any negative influence drags down incomes as their cross-region study under different environments shows. Technologies are known to help with some of these factors. According to Kumar et al. (2018), technological factors play a key role in the development of this sector. However, the authors pointed out its adoption are not straightforward. The drivers are many in its adoption, albeit complex as it involves many interacting variables. Some of the important factors identified by these authors include; the method of information transfer, characteristics of the technology, farm characteristics, economic factors, and sociodemographic and institutional factors. They observed not all factors are applied in toto, but adoption is based on what is perceived as important in a specific environment. To overcome such selective use, Olivier et al. (2017) suggested a system approach to the problem of technology and through better management, both at the farm level and policy level. Such an approach can also help avoid negative consequences as found in some countries (ibid). The system approach through information technologies can be in the form of instrumentation and process control, data management, computerized models, decision support systems, artificial intelligence and expert systems, image processing and pattern recognition, geographical information systems, and information centers and networks and with the right management attitude, aquaculture can be very successful (El-Gayar, 1997). The active influence of government and its agencies is the primary driver, its involvement or otherwise leads to success or failure. An academic field study on aquaculture companies, chairmen, and, senior aquaculture teachers in universities, identified the role of government as being the most influential factor in aquaculture businesses. They opined that the right government support and policy can help close the current gaps in its performance (Chen, et al, 2010). Government involvement can be as diverse as providing funds, training, monitoring, and implementing fishery projects (Dube and Mabika, 2022) and arranging collaboration activities in the area (op cit). It can also manage necessary infrastructure and help in storage and marketing through its agencies. A well-managed supply chain, through using digital technologies and government involvement can help with sustainable development goals (Rowan, 2023)

Research Questions

Based on the literature review and the context of this research, the following research question is formulated.

What are the significant factors that can contribute to the success of aquaculture farming in Malaysia?

Research Objectives

The research question is adduced from the objective identified below.

The research aims to extract significant factors that are central to the success of aquaculture in Malaysia.

Research Methodology

A quantitative study is relevant in this survey as it involves collecting empirical data from field studies and the use of machine learning in model building. Relevant factors are identified from the literature reviews and used for the questionnaire design. Twelve factors of importance are used in the design, eleven being drivers and one being the driver as shown below.

A1: Provision of Extension Service
A2: Climate Change
A3: Innovative Technologies
A4: Farming Practices

A5: Profitability (Target variable)
A6: Institutional Influences)
A7: Supply Chain Risk Management Culture
A8: Cost of Feed
A9: Learning and Development
A10: Economic
A11: Societal
A12: Environmental
Two hundred and sixty-eight eligible participants were selected from an original data list provided by Senarai Syarikat Aqua. Those who are involved in aquaculture and meet the criteria set are invited to participate. The sample size is based on a margin set at 5% with a confidence of 95%. A survey questionnaire, with a Likert scale of 1 to 5, (strongly disagree to strongly agree) was then used to extract the necessary feedback, both in person and through mobile calls.

Data Analysis Plan

After initial screening for outliers, missing values, wrongly entered values, and overall distribution, the data will be taken through logit regression for factor identification and model-building purposes.

The training and testing of data in the model will be set with a ratio of 80:20 respectively. The logit model is suitable, as the study involves a cause-effect regression-based analysis where relationship matters, and, therefore, ideal machine learning techniques where a relationship is involved can be logit regression.

The output from the model, after all the necessary testing procedures, is observed for their predicting ability and accuracy, with relevancies of factors identified. The factors (causes/independent variables) from the model are taken as input drivers for the model. As for the target variable, the realization of profit, the information gathered, is identified as, 'not profitable', and 'yes profitable'. This dual classification is necessary for the use of log-logit regression modelling. A cut point is set as more than 3.5 as 'yes' it had led to Profitability is given a value of '1' otherwise it is given a value of '0' for duality purposes. This is ideal on a Likert scale of 5 points, where the mid-value is 2.5, and 2.5 to 3.5, which can be assumed to be borderline on profitable gains' opinion and more than 3.5 as profitable.

Data Analysis

The following discussion is on Model Building Using Machine Learning Techniques on Log-logit Regression.

Training and Testing

Train-Test split is set at an 80:20 ratio as it is usually practiced, 80 percent of the total dataset is used for training on the Jupiter platform using Python programming. The model learns on the training set and assigns weights and biases for the model. For the Test, 20 percent of the data is used for final evaluation. 'MinMaxScaler ()' scalar is used to scale all the features, MinMaxScaler scales the data to a fixed range, between 0 and 1. The following variance output results are presented for 12 PCA components of the drivers. PCA (Principal Component Analysis) is a dimensionality reduction approach to reduce factors by dropping/bringing together unnecessary correlated variables.

Table 1
Variance output for the PCA

```
           A1         A2         A3         A4         A6     A7         A8  \
A9  \
237  0.000000  0.777778  1.000000  0.477124  0.428571  0.50  0.300752
0.50
63   0.333333  0.555556  0.142857  0.346405  0.142857  0.75  0.751880
0.25
209  0.666667  1.000000  0.571429  0.215686  0.000000  0.75  0.751880
1.00
149  0.111111  0.444444  0.428571  0.084967  0.428571  0.25  0.751880
0.00
139  0.111111  0.444444  0.000000  1.000000  1.000000  0.75  0.751880
0.00

      A10  A11       A12
237  0.50  0.4  0.000000
63   0.75  0.2  0.000000
209  0.50  0.2  0.333333
149  0.00  0.8  0.500000
139  0.75  0.8  0.166667
```

Variance Ratio For Each Component

The variance ratio explains the amount of variance explained by PCA. The array below is the sum from the earlier PCA output.

Table 2
Cumulative Sum of Variance Output

```
array([0.19698988, 0.14005957, 0.11353766, 0.109435  , 0.0865748 ,
       0.08438895, 0.07225522, 0.05533291, 0.04799094, 0.03683013,
       0.0301641 , 0.02644086])

var_cumu = np.cumsum(pca.explained_variance_ratio_)
```

The results of the above values can be summarized as follows in table format.

Table 3
Summary of Explained Variance

PCA	Variance explained (%)	Cumulative variance
PCA1	19.60%	19.6
PCA2	14.00%	33.60%
PCA3	11.40%	45.00%
PCA4	10.90%	55.90%
PCA5	8.60%	64.50%
PCA6	8.40%	72.90%
PCA7	7.20%	80.10%
PCA8	5.50%	85.60%
PCA9	4.80%	90.40%
PCA10	3.70%	94.10%
PCA11	3.00%	97.10%
PCA12	2.60%	99.70%

To explain the above result, PCA 1 explains 19.6% of the variance, and PCA 2 explains 14% of the variance. Together, PCA 1 and 2 explain 33.6% of the variance.

Scree-Plot For the Explained Variance

A graphical representation called scree-plot, cumulative of PCA, of the above variance data is presented below. The first 10 PCAs explain 94.1% of the variance

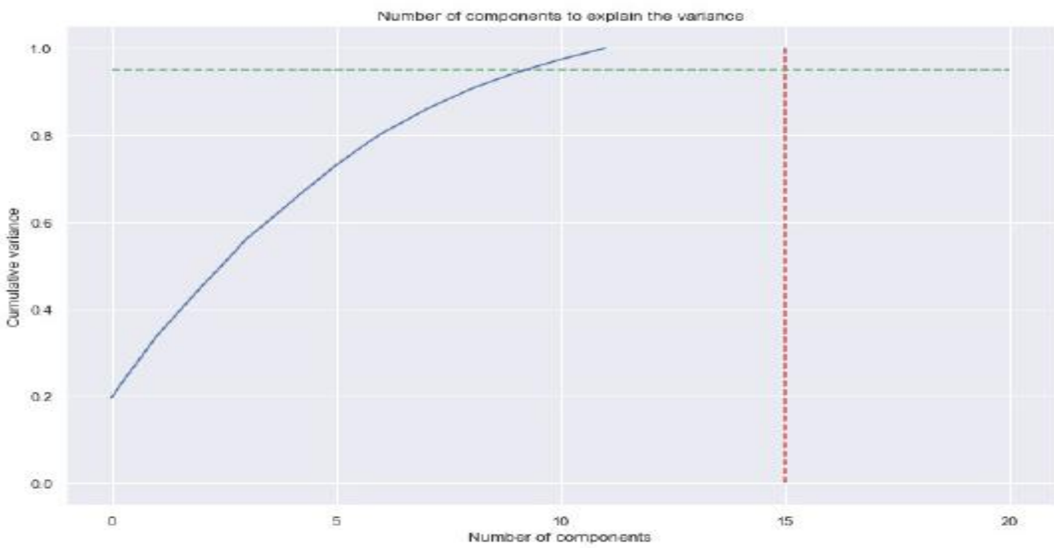

Figure 1: Scree-plot for explained variance

Evaluating the Models on the Trained Test and the Test Set For the Pca

The training accuracy, 61.6%, and the observation of the Test confusion matrix are summarized below from the out give the following.

Accuracy = (TP+TN) / Total = (97+35) / (214) = 0.6168 (as shown in the output) = 61.6%.
Precision = TP/predicted yes = 35 / (95) = 0.3684 (36.8%)

Table 4

Confusion Matrix Summary

N=214	Predicted Neg	Predicted Yes	
Actual Neg	TN=97	FP=60, Type I error	157
Actual Yes	FN=22, Type II error	TP=35	57
	119	95	

Table 5

Evaluating the Model Result

```
evaluate_model(Model_PCA)

Train Accuracy : 0.616822429906542
Train Confusion Matrix:
[[35 22]
 [60 97]]
--------------------------------------------------
Test Accuracy : 0.6296296296296297
Test Confusion Matrix:
[[ 0 20]
 [ 0 34]]
```

Testing Accuracy

Based on the table above, the test accuracy for the prediction is 62.96%.

Table 6
Evaluating Prediction Result

N=54	Predicted Neg	Predicted Yes
Actual Neg	TN=0	FP=20
Actual Yes	FN=0	TP=34
		54

Plotting ROC Curve
The ROC curve (receiver operating characteristic curve) plots the True Positive Rate vs. False Positive Rate for all classification thresholds. The closer the curve comes to the 45-degree diagonal line (FPR=TPR) of the ROC space, the less accurate the test. As the curve is only slightly out of this line the accuracy is somewhat lesser

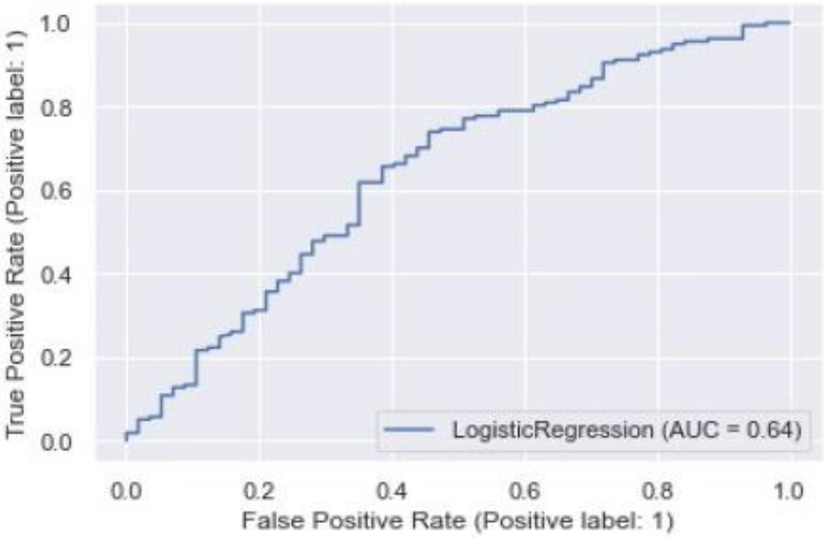

Figure 2: ROC curve

Using RFE to Select the Variables For the Model
Using, RFE (Recursive Feature Elimination) to select features (columns) in a training dataset, the following are more and most relevant in predicting the target variable.

Table 8
RFE Result

```
[('A1', True, 1),
 ('A2', True, 1),
 ('A3', True, 1),
 ('A4', True, 1),
 ('A6', True, 1),
 ('A7', True, 1),
 ('A8', True, 1),
 ('A9', True, 1),
 ('A10', True, 1),
 ('A11', True, 1),
 ('A12', True, 1)]
```

These variables should be used for modelling

```
Index(['A1', 'A2', 'A3', 'A4', 'A6', 'A7', 'A8', 'A9', 'A10', 'A11',
'A12'], dtype='object')
```

Adding the Constant and the Regression Output to Create a Fitted Model
The final model, together with their coefficients, and errors, is given below.

Table 9
The Model with Coefficients and Errors

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.3103	1.082	-0.287	0.774	-2.431	1.810
A1	0.2192	0.531	0.413	0.680	-0.822	1.260
A2	0.3206	0.770	0.416	0.677	-1.189	1.830
A3	-0.0527	0.608	-0.087	0.931	-1.245	1.140
A4	-1.1783	0.573	-2.057	0.040	-2.301	-0.055
A6	-0.0999	0.604	-0.166	0.869	-1.283	1.083
A7	-0.2114	0.643	-0.329	0.742	-1.472	1.049
A8	1.2625	0.639	1.976	0.048	0.010	2.515
A9	0.3828	0.519	0.738	0.461	-0.634	1.400
A10	0.2143	0.547	0.392	0.695	-0.858	1.287
A11	-0.2280	0.645	-0.354	0.724	-1.492	1.036
A12	1.5177	0.557	2.725	0.006	0.426	2.609

Variance Inflation Factor (VIF)
Variance inflation factor measures the amount of multicollinearity in regression analysis, i.e., a correlation between multiple independent variables. Most research papers consider a VIF (Variance Inflation Factor) > 10 as an indicator of multicollinearity, but some choose a more conservative threshold of 5 or below. Here, taking a conservative approach of dropping any factors having a VIF of more than 5, hence, A7, A8, and A11 are dropped from the model, and the remaining 8 factors are used.

Table 10
Factors with VIF Values

	Features	VIF
9	A11	6.48
6	A8	6.41
5	A7	5.85
10	A12	4.90
2	A3	4.37
8	A10	4.13
7	A9	3.96
0	A1	3.92
1	A2	3.84
3	A4	3.47
4	A6	3.32

Findings and Conclusion
The study aims to adduce factors of importance to productively manage aquaculture contextually. Eleven drivers and one target variable were used for the purpose. Machine learning techniques were used on logit regression for the model-building purpose and isolating factors of importance. Of the eleven drivers investigated, all drivers were found to be important; the Provision of Extension Services, Climate Change, Innovative Technologies, Farming Practices, Institutional Influences, Environmental, Learning and Development, Economic. However, Societal, Supply Chain, Risk Management Culture, and Feed are not significant drivers in the local context

Some of these findings are a departure from other similar research, for example, the importance of management (Olivier et al., 2017) is only of secondary importance, likely due to the local business requirement and regular briefing from government agencies. However, the surprising outcomes are the cost (Miyata and Manatunge, 2004) and supply chain issues. Cost seems to be not an issue with local aquaculture participants as pointed out by the outcome. The reason can be traced to some form of subsidy from the government and good supply chain management and facilities and infrastructure provided by the government (El-Gayar, 1997; Chen et al, 2010). However, Environmental (Wang and Mendes, 2022), innovation (Kumar et al, 2018) and related knowledge infuse and the right education and training are still an issue (Rowan, 2023). These issues are common outcomes of many past studies where the involvement of many stakeholders is expected but is lacking. The uncontrollable economic climate and weather patterns (Kumar et al., 2018) also influence the viability of the business, this study, which was undertaken just outside the Covid-19 period seems to affect aquaculture farms' viability. However, more support is required, especially on the continuity of any services rendered from the outside as identified by Dube and Mabika (2022).

Recommendation

The study identified important factors that are important in the local context for the sustenance and development of this important sector namely the provision of extension services, climate change, innovative technologies, farming practices, institutional influences, environmental, learning and development, and economic. These factors identified can be a source of input for policymakers in the local context for any future blueprint for this sector. Hence it is recommended that any future directions the government and policymakers intend to take emphasize the factors.

Acknowledgement

The grant for this study comes from the Ministry of Higher Education Malaysia, Fundamental Research Grant Scheme (FRGS). However, the authors declare that they have no conflicts of interest in undertaking this research project. The data that support the findings of this study are available from the corresponding author upon reasonable request. The study has taken all necessary ethical practices, however, if there are any issues due to omission or commission do contact the authors for ratification.

References

Chen, S. C., Huang, C. C., & Chiang, H. C. (2010). Factors affecting the participation of the aquaculture industry in university-industry co-operation. *World Transactions on Engineering and Technologies Education,* 8(4), 510-516.

Daudpota, A. M., Kalhoro, H., Abbas, G., Kalhoro, I. B., & Shah, S. S. A. (2016). Effect of Feeding Frequency on Growth Performance, Feed Utilization and Body Composition of Juvenile Nile Tilapia, Oreochromis niloticus (L.) Reared in Low Salinity Water. *Pak J Zool* (48), 171-177.

Dube, S., & Mabika, N. (2022). Factors affecting the adoption of modern fish farming technologies in Gokwe South District, Zimbabwe. *International Journal of Scientific Reports*. (9), 11-16.

El-Gayar, O. F. (1997). The use of information technologies in aquaculture management. *Aquaculture Economics & Management*, 1(1-2), 109-128.

Galappaththi, E. K., Galappaththi, C. J., Aubrac, S. T., Ichien, A. A., & Hyman, J. D. (2020). Climate change adaptation in aquaculture. *Reviews in Aquaculture* (12), 2160-2176.

Jegede, T., & Olorunfemi, O. T. (2013). Effects of Feeding Frequency on Growth and Nutrient Utilization of (O. niloticus) fingerlings. *Global Journal of Science Frontier Research Agriculture and Veterinary* (13).

Kumar, G., Engle, C., & Tucker, C. (2018). Factors Driving Aquaculture Technologies Adoption. *Journal of World Aquaculture Society,* 49(3), 447-476.

Ludson, G. M., Marcos, A. S., Érika, R. A., Gabriel, F. O., & Nadille, H. F. (2020). Effects of different stocking densities on Nile tilapia performance and profitability of a bio-floc system with a minimum water exchange. *Aquaculture,* (530), Article 735814.

Miyata, S., & Manatunge, J. (2004). Knowledge Sharing and Other Decision Factors Influencing Adoption of Aquaculture in Indonesia. *International Journal of Water Resources Development*, 20(4), 523-536.

Department of Fisheries Malaysia (2024). Official Portal, Department of Fisheries Malaysia.

Olivier, M. J., Laurens, K., Malcolm, D., & Marc, V. (2017). How is innovation in aquaculture conceptualized and managed? A systematic literature review and reflection framework to inform analysis and action. *Aquaculture* (470), 129–148.

RAS aquaculture. (2023). Aquaculture Industry In Malaysia: Challenges And Issues. *Interview*.

Rowan, N. J. (2023). The role of digital technologies in supporting and improving fishery and aquaculture across the supply chain – Quo Vadis? *Aquaculture and Fisheries,* 8(3), 365-374.

Wang, P., & Mendes, I. (2022). Assessment of Changes in Environmental Factors Affecting Aquaculture Production and Fisherfolk Incomes in China between 2010 and 2020. *Fishes*, 7(4), article 192. https://doi.org/10.3390/fishes7040192

Zhou, Z., Cui, Y., Xie, S., Zhu, W. & Lei, W. (2003). Effect of feeding frequency on growth, feed utilization, and size variation of juvenile gibel carp (C. auratus gibelio). *J Appl Ichthyol* 19, 244-249.