Journal of Bioresource Management

Volume 7 | Issue 2 Article 9

6-29-2020

Prediction of Feed Utilization Performance in Clarias gariepinus Using Multiple Linear Regression in Machine Learning

Adekunle Oluwatosin Familusi
Ekiti-State University, Ado-Ekit, Nigeria, familusiadekunle@gmail.com

Recommended Citation
Familusi, A. O. (2020). Prediction of Feed Utilization Performance in Clarias gariepinus Using Multiple Linear Regression in Machine Learning, *Journal of Bioresource Management, 7* (2).
DOI: https://doi.org/10.35691/JBM.0202.0134

Prediction of Feed Utilization Performance in Clarias gariepinus Using Multiple Linear Regression in Machine Learning

Cover Page Footnote

A great deal of appreciation goes to IBM coursera team for dexterity in knowledge dissemination. Dr. Familusi, E.B. also played a huge role in facilitating my hunger for knowledge and persistence. To the Fisheries and Aquaculture Management department, Ekiti-State University; No one has ever motivated me more, I'm very grateful.

PREDICTION OF FEED UTILIZATION PERFORMANCE IN AFRICAN SHARPTOOTH CATFISH (*Clarias gariepinus*) USING MACHINE LEARNING

ADEKUNLE OLUWATOSIN FAMILUSI

Ekiti-State University, Ado-Ekiti, Nigeria

Email: familusiadekunle@gmail.com

ABSTRACT

Machine learning models can be used to make predictions about nutrient utilization performance index using available proximate analysis data on feed composition. Data from similar experiments on nutrient utilization performance was used to fit a multiple linear regression model for the prediction of four performance indexes. The Specific Growth Rate and percentage inclusion with strength of 0.57 was noted along with a negative relationship between protein efficiency and protein content. A negative relationship between Nitrogen Free Extract (NFE) and Protein Efficiency Ratio (PER) at NFE content ≥25 % was observed. PER was predicted with 85 % accuracy, while Weight Gain (WG), Feed Conversion Ratio (FCR) and Specific Growth Rate (SGR) were predicted at 48 %, 7.6 % and 4.2 % respectively. WG model showed highest coefficient value to ash content (1.23) which is less likely to contribute to fish weight compared to values of fat content (-0.34) and crude protein (-1.02). FCR and SGR models appeared to be dependent on variables outside those included in the proximate analysis data for this study.

Keywords: Machine learning, protein efficiency ratio, weight gain, specific growth rate, feed conversion ratio.

INTRODUCTION

African Sharptooth Catfish (*Clarias gariepinus)* is a major cultivated fish in parts of West-Africa, with an increasing market demand due to growing human population that intensifies pressure on the aquaculture industry to increase production. Feed alone accounts for nearly 70 % of the total cost which decides profitability of Aquaculture. High-quality feeds are available but at relatively high prices.

Several farms depend on manufacture of feed using locally available feedstuff including agricultural by-products to reduce the cost. Akhlaqur and Sumaira (2014) described the relevant points of data collection in aquaculture including farm yields and environmental data, and how this information can improve decision-making using specific algorithms. A breakdown of machine learning terminology, algorithm, and applicability to specific areas in agriculture was also done by Konstantinos et al. (2018).

Machine learning is a branch of computer science that focuses on the use of historical data to predict, cluster, and classify datasets. While this branch of science is very popular in other fields including health sciences, engineering, pharmacy, and systems biology for sound and factual decision-making, it can be regarded as being in its early stage of use in fishery due to low cross cross-field interaction. Machine learning uses several types of models for specific purposes and a knowledge of the conditional-specific application is required as well as programming skills.

Currently, research on aquaculture feeding practices is aimed at reducing feed conversion efficiency, optimizing environmental and economic benefits by reducing feed using image analysis (Breiman et al., 1993). Machine learning has also been applied in aquaculture for algal bloom farm closure prediction as reported by Shahriar and Rahman (2013).

The representation and quality of the instance data is essential for the success of machine learning on a given task. Moreover, knowledge discovery during the training phase is more difficult if there is irrelevant, noisy, or unreliable information. Identifying and removing irrelevant information, selecting a subset of features, reduces the dimensionality of the data which often allows learning algorithms to operate more effectively Mendoza et al, (2011).

MATERIALS AND METHOD

Data Acquisition

Research data on *Clarias gariepinus* feedstuff performance was selected from five papers based on similarities in Chemical proximate analysis procedure (AOAC), statistical analysis and measurement of parameters (Percentage Weight Gain, Feed Conversion Ratio, Specific Growth Rate, and Protein Efficiency Ratio according to Olvera-Novoa et al., 1990).

Table 1. Proximate analysis data

	Source code	Inclusion (%)	Culture period (week)	Fish weight (mg)	Moisture (%)	Fiber (%)	NFE (%)	Protein	Fat	Ash
Wei-Kang	FTM0	56.65	4	2.85	4.78	4.811	25.85	39.82	11.45	86.55
Wei-Kang	FTM20	45.32	4	2.85	5.12	4.811	25.85	39.51	11.23	87.43
Wei-Kang	FTM40	33.9	4	2.85	5.03	4.811	25.85	41.25	11.87	89.1
Wei-Kang	FTM60	22.6	4	2.85	5.89	4.811	25.85	38.95	12.14	89.7
Wei-Kang	FTM80	11.33	4	2.85	4.43	4.811	25.85	40.34	11.5	90.7
Wei-Kang	FTM100	0	4	2.85	5.37	4.811	25.85	40.67	11.26	91.27
Adetaranmi	AMGT0	31.73	10	21.73	9.96	3.74	23.98	44.2	4.76	13.37
Adetaranmi	AMGT50	22.77	10	21.8	9.9	3.76	25.47	43.5	4.6	12.7
Adetaranmi	AMGT33	15.95	10	21.67	9.8	3.74	24.52	44.23	4.5	13.11
Adetaranmi	AMGT66	35.57	10	21.8	9.94	3.85	24.8	43.48	4.76	13.16
Adetaranmi	AMGT75	43.65	10	21.77	4.9	4.1	28.54	44.03	4.85	13.6
Alphonsus O.	CMGT0	25	10	10	9.55	4.1	28.54	40.76	9.2	10.59
Alphonsus O.	CMGT12	12.5	10	10	9.22	4.87	28.54	40.59	8.98	9.1
Alphonsus O.	CMGT25	25	10	10	9.87	5.22	28.54	40.74	8.51	8
Falaye	FMAIZE0	13.63	12	5	9.96	4.2	28.45	37.42	4.36	15.78
Falaye	FMAIZE25	13.3	12	5	9.78	5.53	28.2	35.96	4.98	15.88
Falaye	FMAIZE50	13.3	12	5	9.89	5.51	27.3	36.42	5.24	16.02
Falaye	FMAIZE75	13	12	5	9.93	5.6	24.92	38.25	5.46	15.95
Falaye	FMAIZE100	13	12	5	9.88	5.58	22.63	40.2	6.66	16.22
Oyekanmi	OFSHML100	62	13	10.73	8.88	3.44	39.96	38.3	6.81	8.61
Oyekanmi	OFSHML75	47	13	10.57	9.41	3.56	34.2	37.6	6.45	8.78
Oyekanmi	OFSHML50	31	13	10.56	9.83	3.73	30.47	38.9	7.63	9.44
Oyekanmi	OFSHML25	15	13	10.58	9.46	3.63	32.09	37.8	7.49	9.53
Oyekanmi	OFSHML0	0	13	10.61	9.49	3.69	33.3	36.6	7.33	9.59

Source code- representation of variable index as presented in data source.

Inclusion- Percentage composition of protein in the feed.

NFE- Nitrogen Free Extract.

Fish Weight (mg)- initial weight at start of feeding trial

Protein- Crude protein content

Table 2. Feed Utilization performance Data

Source code	WG	FCR	SGR	PER
FTM0	999.62	1.24	8.56	2.16
FTM20	788.46	1.34	7.79	2.04
FTM40	79.06	2.9	2.08	0.88
FTM60	87.98	3.08	2.24	1.03
FTM80	36.05	2.85	1.09	1.22
FTM100	14.04	19.91	0.47	0.79
AMGT0	193.26	5.07	0.69	0.43
AMGT50	298.16	3.96	0.96	0.58
AMGT33	367.08	3.57	1.12	0.62
AMGT66	273.35	4.16	0.9	0.57
AMGT75	506.72	3.13	1.37	0.75
CMGT0	248.61	1.15	2.53	2.55
CMGT12	259.02	1.17	2.55	2.6
CMGT25	263.98	1.16	2.56	2.6
FMAIZE0	127.5	0.65	0.97	0.52
FMAIZE25	122.6	0.68	0.95	0.5
FMAIZE50	150.9	0.7	1.09	0.56
FMAIZE75	150.1	0.68	1.09	0.56
FMAIZE100	152.8	0.62	0.94	0.55
OFSHML100	38.4	2.33	2.3	2.74
OFSHML75	38.4	2.33	2.3	2.74
OFSHML50	43.5	2.16	2.53	2.85
OFSHML25	38.18	2.3	0.47	2.16
OFSHML0	36.91	2.62	2.23	2.72

WG- Weight Gain, FCR- Feed Conversion Ratio, SGR- Specific Growth Rate, PER- Protein Efficiency Ratio

Pre-Processing

Each feature was recorded within a single excel spreadsheet and loaded into Jupyter notebook on the anaconda package. Missing values under crude fiber and NFE were treated using feature means.

Model Development

Some factors were expected to bear more relevance to the model. Hence, correlation was examined between the features, those with relatively odd values were viewed using the regression plot. Proximate analysis features were loaded into a single variable (Z), then split into a training and testing set using the train_test_split function to get more accurate results out of sample accuracy. Half of the data size was used for training the model, and the other half for testing. The total number of correct predictions were considered as the model accuracy.

The regression uses several variables to predict a single variable for the prediction of each of the four dependent variables for which we would like to know or predict, these included PER, SGR, FCR, and WG. The models were visualized using simple line plots. The values of R^2 are shown for each model to show how accurately the model prediction is when compared to existing data. The goal was to understand if values of proximate analysis for an already compounded feedstuff were accurate enough to predict feed performance.

Model Accuracy

The models are compared based on the value of R^2. R^2 is an evaluation metric that indicates the level of accuracy. It has a maximum value of 1 which is equivalent to 100%.

Obtaining a Prediction

The model can be used for predictions of the Protein Efficiency Ratio of a feed to be fed to *Clarias gariepinus species* using the proximate analysis data of the experimental feed with 85% certainty of accuracy level. Using the coefficients presented in Table 3, predictions follow the equation:

h = + 1 1+ 2 2+ 3 3+ 4 4 (1)

Where Y-hat= feature to be predicted (PER)

a= intercept (Starting point of the slope)

b= coefficients of the variable (as presented in Table 3 for PER)

x= predictor variable (proximate analysis values obtained).

A copy of Jupyter notebook and dataset used for modeling is also made available on Github for further analysis, modifications or extension and can be accessed here.

RESULTS AND DISCUSSION

Correlation

Below are scatter plots with fitted regression lines to get an estimate of relationship between the variables and the direction of the correlation. Regression plots showing linear relationships between features indicating an existing relationship.

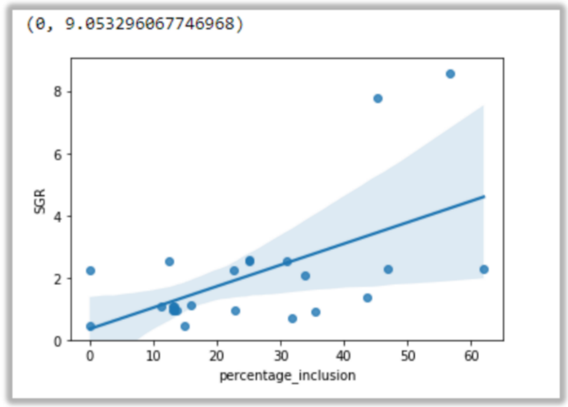

Figure 1: Positive correlation plot showing reasonable relationship between specific growth rate and the percentage of protein inclusion strength of the correlation is 0.57. (Coefficients are indicated on each plot).

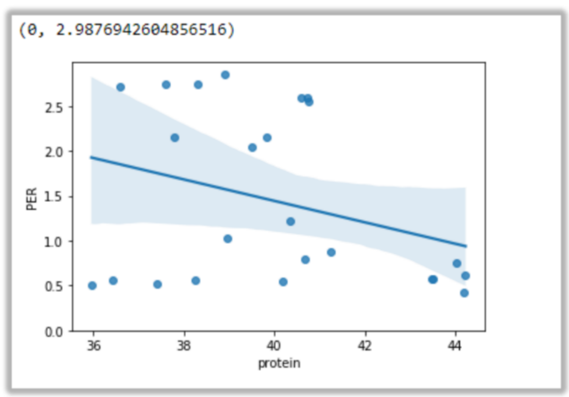

Figure 2: Negative correlation plot indicating relationship between Protein Efficiency and Protein Content.

With a strength of -0.13. Data points are highly scattered which indicates protein content may not be a very good predictor of protein efficiency.

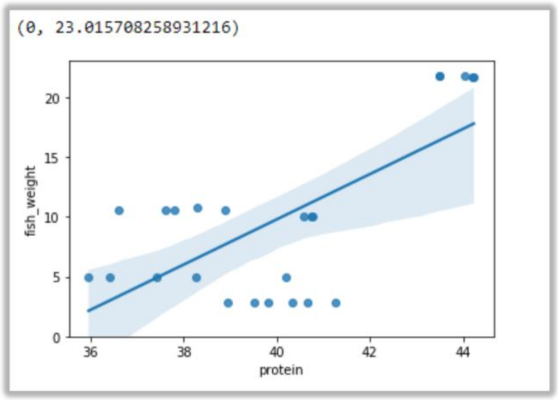

Figure 3: Positive regression Plot of Fish weight and Protein Content.

Coefficient strength was estimated at 0.68. This indicates increasing inclusion of protein with fish weight in feed used for experimental trials. The plot shows the inclusion of more protein in feed of fish with higher weight. The relationship can be attributed to increasing weight of experimental fish during the course of experiments.

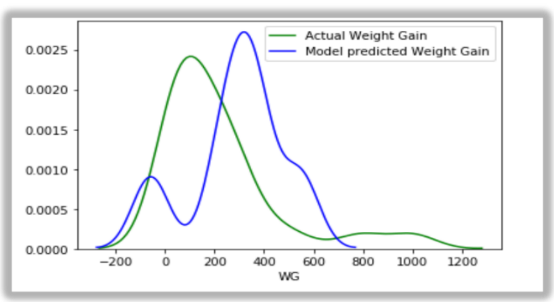

Figure 4: Weight Gain Prediction Model. Model accuracy = 48%. Green line represents the actual weight gain values, while the blue line indicates the predicted weight gain.

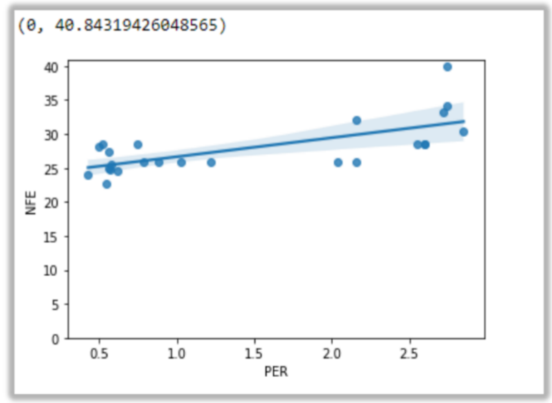

Figure 5: Indication of NFE's impact on protein utilization. Correlation strength = 0.69.

The plot shows high NFE levels increases the Protein Efficiency. Safe NFE levels fall between 20- 26% beyond which it highly influences the efficiency of Protein negatively.

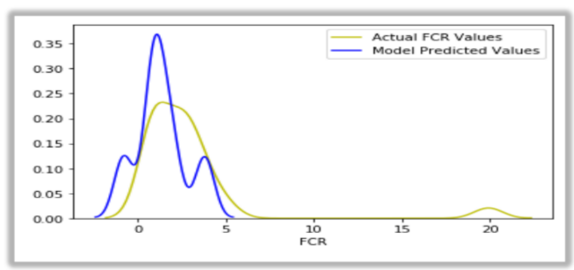

Figure 6: Feed conversion Ratio Prediction Model. Model accuracy= 7.6%.

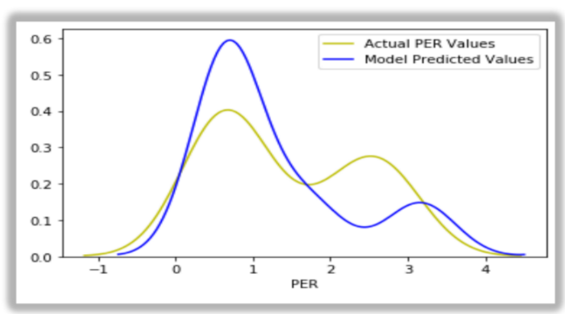

Figure 7: Protein Efficiency Ratio Prediction Model. Model accuracy= 85%.

Table 3. Regression coefficient used to derive model predictions

Feature	Model accuracy	Reg. Coeff Inclusion	Reg. Coeff. Culture period (wk)	Reg. Coeff Fish weight (mg)	Reg. Coeff Moisture (%)	Reg. Coeff Fiber (%)	Reg. Coeff NFE (%)	Reg. Coeff protein	Reg. Coeff Fat	Reg. Coeff Ash
WG	48%	0.07	0.047	-0.2	-0.13	-1.08	-0.34	-1.02	0.31	1.23
FCR	7.6%	0.01	0.28	1.51	0.12	0.58	0.19	0.08	0.25	0.15
PER	85%	0.01	0.00	-0.04	-0.04	-0.44	-0.00	-0.15	0.47	-0.13
SGR	4.2%	0.09	-2.4	0.25	-0.622	-13.86	-1.02	-0.34	-2.4	-5.92

Feature- Dependent variables predicted and visualized as plots in figure 3.5 – 3.8

Reg. Coeff Inclusion- Percentage composition of protein in the feed.

Reg. Coeff Culture period- Weeks over which the feeding trial lasted

Reg. Coeff Fish Weight- Initial weight of experimental fish

Reg. Coeff NFE- Nitrogen Free Extract obtained as part of proximate composition

Reg. Coeff Fiber, Reg. Coeff Protein, Reg. Coeff Fat, Reg. Coeff Ash, Reg. Coeff Moisture-

Weight of Independent variables to be substituted into multiple linear regression equation to derive prediction for proximate analysis data with accuracy levels as reported in figure 5-8.

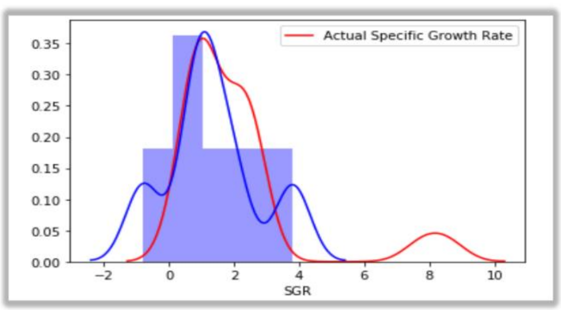

Figure 8: Specific Growth Rate Prediction Model. Model accuracy = 4.2

All predictions were obtained using multiple linear regression. Linear models of WG, FCR, and SGR made poor predictions making them unfit for performing experimental simulations. However, the predictions for protein efficiency ratio had an 85% accuracy indicating that the coefficients are acceptable for future predictions.

Weight Gain prediction with an accuracy of 48% attributed the highest regression coefficient value to Ash content (1.23) followed by Fat (0.31), percentage inclusion (0.07) and culture period (0.047). Since ash content indicates the amount of minerals like calcium in the feed, it was expected that it would not contribute much to the weight of the fish. Features like Protein content and NFE as reported by Adetarami and Akinlade (2013), were observed to be the major contributors to fish weight, this misrepresentation may account for the 52% error in the model.

In Figure 1, the correlation plot indicated that a higher percentage of protein correlates with the growth rate and this observation was supported by variance observed in the data. In Figure 2, the protein efficiency reduced with increasing protein content. Increase in protein efficiency, which is undesirable was also seen to be positively influenced by Nitrogen-free extract. This interconnectedness that makes use of a single predictor variable, as is the case with single linear regression, is quite unacceptable. Therefore, combining all the independent variables into a multiple linear regression equation to generate values for slope, intercept and coefficients is expected to generate more predictions with high sample accuracy. Accuracy of the models can be influenced by other parameters including environmental factors, genetics and anti-nutritional factors.

FCR is estimated as total feed fed (g) / net weight gain (g). Accuracy of 7.6% was observed with the highest coefficients attributed to fish weight, crude fiber and the fat-content. FCR as reported by Charo-Karisa et al. (2013), depends largely on growth stage and premixes which weren't included in the study. Jamabo and Dienye (2017) also observed that feed with highest value for protein and fat and the lowest value for fiber returned the best FCR.

PER showed a prediction accuracy of 85%, with the highest coefficient being the Fat content in the feed. Olukunle and Ekundayo (2016) reported that lipids are included in diets to spare protein as an alternative energy source. However, PER was expected to also be influenced by other variables including the culture period which had a coefficient of 0.00 (Table 3), optimizing this variable is expected to account for part of the 25% error in the model.

SGR was estimated to find the daily growth of experimental fish. As opposed to the strong correlation observed between SGR and percentage inclusion with a strength of 0.57 in Figure 1, an accuracy of 4.2% and a coefficient of 0.09 was attributed to percentage inclusion under multiple linear regression. The low accuracy can be attributed to the estimation of highest coefficient as the weight of the experimental fish at the start of the experiment or culture period rather than nutrient factors. SGR is one of the variables that depends on other factors outside, percentage inclusion, fish weight, fiber, NFE, protein, ash, and moisture for its prediction. Inclusion of variables such as the amino acid profile of utilized feed, band of possible anti-nutritional factors, vitamins, minerals and premixes might develop a more accurate model.

The high protein content in feather meal reported in historical data by Wei-Kang et al. (2013) showed high protein content but very low nutrient utilization performance without anti-nutritional factor representation. With the collation of quality data, prediction models are expected to be more accurate for easy application on-site.

CONCLUSION AND RECOMMENDATION

The model poorly predicted weight gain, the same trend was also repeated in the Feed conversion and SGR. However, PER showed a high accuracy (85%). Further regression modeling can also be carried out using feedstuff as predictor variables. Accuracy of model prediction is dependent on the quantity and quality of available historical data. Hence, using these data for future purposes especially in the emerging field of Big Data requires consistency. Experimental procedures, analytical techniques, and standard units of measurement for each experiment need to conform to specific standards for subsequent use.

REFERENCES

Adetarami D and Akinlade S (2013). The use of maggot meal in African cat fish feeding. Adv Aquac Fish Manage.**,** 1 (5): 49-51.

Alphonsus O and Ebere E (2009). Replacement of fish meal with maggot meal in African Catfish (*Clarias garepinus*) diets. Revista

Cientifica UDO Agricola., 9 (3): 666-671.

Akhlaqur R and Sumaira T (2014). Application of Machine Learning techniques in aquaculture. Int J Comput Trends Technol., 10 (14): 214-215.

AOAC (1990). In: Official Methods of Analysis, 13th ed., Association of Official Analytical Chemists, Arlington, Virginia.

Breiman L, Friedman J, Olshen R and Stone C (1993). Classification and regression trees. In: Information Reuse and Integration in Academia and Industry. Illustrated ed., Chapman and Hall, New York: pp 199-215.

Charo-Karisa H, Opiyo M, Munguti J, Marijani E and Nzayisenga L (2013). Cost-benefit analysis and growth effects of pelleted and unpelleted on-farm feed on African Catfish (*Clarias gariepinus*, Burchell 1822) in Earthen ponds. Afr J Food Agric Nutr Dev., 13 (4): 8019-8033.

Falaye A, Eyiwunmi, Omoike A and Adesina S (2015). Growth performance and nutrient utilization of Catfish *Clarias gariepinus* fed varying inclusion level of fermented unsieved yellow maize. Cont J Biol Sci., 8 (1): 14-23.

Jamabo N and Dienye H (2017). Growth performance of *Clarias gariepinus* fed with different commercial feeds. J Nat Sci Res., 7 (6): 125-129.

Konstantinos L, Patrizia B, Dimitrios M, Simon P and Dionysis B (2018). Machine Learning in Agriculture: A Review. Sensors, 18 (8): 2674. doi: 10.3390/s18082674

Mendoza M, Pennino M, and Bellido J (2011). Tree-Based Machine Learning Analysis fornFisheries Research. In: Fishery Management. Nova Publisher, New York: pp 337.

Shahriar M and Rahman A (2013). Spatio–temporal Prediction of Algal Bloom, Proc. IEEE International Conference on Natural Computation (ICNC), China: pp. 968–972

Olukunle O and Ekundayo I (2016). Assessment of some commercial feed brands in nigeria on growth performance of *Clarias gariepinus* fingerlings. Researcher, 8 (5): 19-28.

Oyekanmi B, Omoniyi I, and Akegbejo S (2013). Whole rocky fresh-water prawns, *Caridina africana* as replacement for fish-meal in diets for African catfish (*Clarias gariepinus)*. Proceedings of 28th Fisheries Society of Nigeria Annual Conference. pp. 25-30.

Olvera-Novoa M, Campos G, Sabido G and Martinez – Palacios C (1990). The use of alfalfa leaf protein concentrate as protein source in diets for Tilapia (*Oreoclronics mossambicus*). Aquaculture, 90: 291-302.

Thomas M, Matteo G, James I, Stewart O, Nicolas B and Leon B (2019). Prediction of bio-concentration factors in fish and invertebrates using machine learning. Sci Total Environ., 648: 80–89.

Wei-Kang C, Leong-Seng L, and Rossita S (2013). Evaluation of feather meal as a dietary protein source for African Catfish fry, *Clarias gariepinus.* J Fish Aquat Sci., 8 (6): 697-705.