

## Development of a prediction model based on linear regression to estimate the success rates of seafood caught from different catching centers



Vinu Sherimon<sup>1,\*</sup>, P. C. Sherimon<sup>2</sup>, Alaa Ismaeel<sup>2</sup>

<sup>1</sup>Department of Information Technology, University of Technology and Applied Sciences, Muscat, Oman

<sup>2</sup>Faculty of Computer Studies, Arab Open University, Muscat, Oman

### ARTICLE INFO

#### Article history:

Received 17 July 2021

Received in revised form

23 October 2021

Accepted 23 October 2021

#### Keywords:

Prediction of seafood quality

Linear regression-based prediction

Least-square method

LSM

Fish quality

### ABSTRACT

For businesses and organizations that aim to be efficient and competitive on a worldwide basis, food quality assurance is extremely important. To maintain constant quality, global markets demand high food hygiene and safety standards. Intelligent software to assure fish quality is uncommon in the fishing industry. Most seafood processing industries utilize Total Quality Management (TQM) systems to ensure product safety and quality. These protections ensure that significant quality risks are kept within acceptable tolerance limits. However, there are no ways for calculating the success rates of seafood obtained from different catching centers. The purpose of this study is to develop algorithms for predicting the success rates of seafood caught at different catching centers. To determine the best model to match the data, the algorithms employ the Least-Square Curve Fitting approach. The success rates are predicted using the best-fit model that results. The bestFitModelFinder algorithm is used to find the best model for the input data, while the prediction of quality algorithm is used to predict the success rate. The algorithms were tested using data obtained from a seafood company between January 2000 and December 2019. Statistical metrics such as mean absolute deviation (MAD), mean square error (MSE), root mean square error (RMSE) and mean absolute percentage error (MAPE) are used to evaluate the prediction accuracy of the presented algorithms. The algorithms' performance analysis resulted in lower error levels. The proposed algorithms can assist seafood enterprises in determining the quality of seafood items sourced from various fishing areas.

© 2021 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### 1. Introduction

Food quality assurance is particularly important to efficient and internationally competitive businesses and organizations. To ensure consistent quality, international markets demand high food hygiene and safety standards. Food-borne infections are on the rise in developed countries, according to the World Health Organization (WHO), due to a lack of effective procedures to check the quality and safety of food (WHO, 2015). As a result, assuring food quality has become an extremely critical concern in recent years, and it is the first move toward drawing worldwide attention to a country's food goods. In many parts of the world, seafood has

long been a favored element of the diet, and in some places, it has even served as the primary source of animal protein (Huss, 1994). Fish is becoming a healthier option for red meat for an increasing number of individuals (Huss, 1994). The consumption of fish, as an excellent source of omega-3 fatty acids, proteins, and vitamins, is increased among people and became an essential part of a balanced human diet (Roos et al., 2003). However, if handled improperly, these items can cause a variety of infectious diseases including food poisoning in humans. Seafood goods, in general, suffer numerous quality control issues across the product spectrum, particularly in export markets.

Intelligent software to assure the quality of fish is not very common in the seafood industry. To assure product safety and quality, most seafood processing businesses use Total Quality Management (TQM) systems. Good Manufacturing Practices (GMP), Sanitation Standard Operating Procedures (SSOP), and Hazard Analysis Critical Control Point (HACCP) processes are all included in such systems (Oliveira

\* Corresponding Author.

Email Address: [vinusherimon@yahoo.com](mailto:vinusherimon@yahoo.com) (V. Sherimon)

<https://doi.org/10.21833/ijaas.2021.12.013>

Corresponding author's ORCID profile:

<https://orcid.org/0000-0003-4923-2841>

2313-626X/© 2021 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

et al., 2016). All these measures ensure that significant dangers that compromise the quality are kept under acceptable tolerance levels. However, no methods are available for analyzing the results of the numerous tests or predicting the success rates of seafood acquired from specific or distinct catching centers. The purpose of this research paper is to suggest algorithms to predict the success rates of seafood caught from different catching centers. This is accomplished by analyzing the past success rates of seafood from the catching center and forecasting future success rates. The proposed algorithm examines previous data, determines the best model to fit the data. The fitted model is then used for future predictions.

The two types of prediction approaches are qualitative and quantitative prediction. Qualitative prediction methods are sometimes known as "subjective" procedures since they rely on people's ability to extrapolate and generalize. Quantitative prediction methods rely on historical data to construct mathematical extrapolations of future data. When historical data is available, such techniques are applied. The information can be specified in mathematical terms, and those characteristics of the standard that have been validated in the past can be assumed to continue in the future (estimation of continuity).

While conducting literature studies, it has been observed that there are no studies in the linear regression-based prediction of quality of seafood based on the catching centers. We found a 2019 study, where the numerous fish rotting indicators were assessed throughout 12 days of storage at 4 ± 2 °C using a basic multispectral imaging (430–1010 nm) system, as well as linear and non-linear regressions (Khoshnoudi-Nia and Moosavi-Nasab, 2019). Total-Volatile Basic Nitrogen (TVB-N), Psychotropic Plate Count (PPC), and sensory score in fish fillets were used as markers (Khoshnoudi-Nia and Moosavi-Nasab, 2019). Different chemometric models, such as partial least-squares regression (PLSR), multiple linear regression (MLR), least-squares support vector machine (LS-SVM), and back-propagation artificial neural network (BP-ANN), were examined in terms of prediction performance. For simultaneous prediction of PPC, TVB-N, and sensory score, all models performed well ( $R^2_P$  0.853 and  $RPD$  2.603) (Khoshnoudi-Nia and Moosavi-Nasab, 2019). The study of García et al. (2017) included the creation of a smart quality sensor that can be used to assess and predict quality over time. The sensor uses biochemical and microbiological deterioration markers, as well as dynamic models, to predict quality according to QIM and EU grading criteria.

The rest of the paper is organized as follows: Section 2 is concerned with the materials and methods used in this research. It includes the general literature about the different stages involved in the development of an effective prediction model, the theory of curve fitting with the focus on least-square curve fits. The proposed algorithms to predict

the quality, the find the best model to fit the data are also included in this section. The next section presents the results and the interpretation of the results. Performance analysis of the proposed algorithms is given in Section 4. The conclusion is given in Section 5 followed by references.

## 2. Materials and methods

### 2.1. Development of an effective prediction model

Model selection, Model fitting, and Model validation are the three essential phases in constructing an effective prediction model (Pham, 2006). The available data points are plotted in the model selection step to establish the shape of the model to fit the data (Pham, 2006). Most of the models use linear, polynomial, or simple nonlinear functions. As a result, the ideal technique to choose an initial model is to plot the data, examine the shape, and then choose the best model that matches the data (Pham, 2006). Following the selection of a basic functional model, the next stage in the model-building process is to determine the unknown parameters in the function using an acceptable model-fitting method (Pham, 2006). Maximum likelihood and least-squares are the two most common methods for parameter estimation. Both methods generate parameter estimators with a variety of useful characteristics. The most crucial phase in the process is to validate the model. Several statistical indicators are used to determine the accuracy of the chosen model in this step. The analysis of residuals is the primary metric. The variations between the actual values and the matching anticipated values produced using the regression function are the residuals from a fitted model (Picard and Cook, 1984). If the chosen model appears to be appropriate, it is utilized for prediction. If the model validation reveals any flaws in the chosen model, the modeling process is repeated to choose a better model (Kuhn and Johnson, 2013).

### 2.2. Curve fitting

Curve fitting is the method of determining the equation of the best-fit curve that can be used to forecast unknown variables (Guest and Guest, 2012). It is also referred to as regression analysis, and it's used to identify the "best fit" line or curve for a set of data points (Freund et al., 2006). The data points are shown, and the basic form of the model is seen during curve fitting. Any point along the curve can be found using the equation generated. To verify shape similarity, the displayed data can be compared to a series of curves. The data must be converted to resemble the curve once a curve that matches the general shape of the data has been determined (Freund et al., 2006). The equation is derived by the form and is used to interpolate or extrapolate the

results. The statistical process of regression varies from curve fitting in that the latter is often the most sensible technique of obtaining the former (Freund et al., 2006). Curve fitting places a higher emphasis on the shape of the curve that will be used to match the data. However, regression is frequently used without much consideration for curve selection.

Curve fitting has the advantage of allowing us to reliably estimate parameter values if we know the mathematical model of the process that created the given data. The technique is known as parametric regression (Hardle and Mammen, 1993). However, obtaining a satisfactory fit necessitates a strong representative system model and correct beginning parameter values, which is the system's drawback. The least-square curve fits, non-linear curve fits, and smoothing curve fits are the three types of curve fits (Motulsky and Ransnas, 1987). The least-square curve fits are the most common of these.

### 2.2.1. Least-square curve fit

The least-square curve fit minimizes the sum of the squared vertical deviations between the original data and the predicted values (Howell, 1971). This method of the curve fit is relatively straightforward and easy to compute and understand, but it is not the most statistically robust method of fitting a function. This method is sensitive to outliers in the data (Howell, 1971). If a given data point is widely different from the rest of the points, the regression results can be false. The commonly available least-square curve fits are linear, quadratic, exponential, logarithmic, and power (Howell, 1971).

Linear regression is a straightforward method of supervised learning. It is the most exact method for fitting a linear regression model. Among the different approaches for determining parameter values, the least-squares method is the most reliable (Steel and Torrie, 1960). The parameter values are selected using the least-square approach so that the total of the squared vertical deviations between data points and the curve is as small as possible. LSM is a method for fitting a unique curve through a set of data points (Fig. 1).

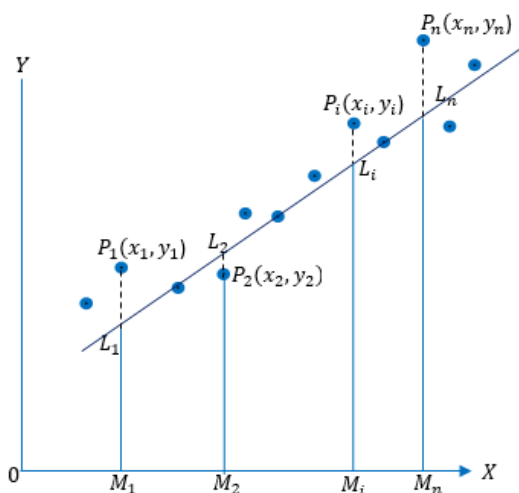


Fig. 1: Least-square method

Let the curve  $y = a_0 + a_1x + a_2x^2 + \dots$  be fitted to the set of  $n$  data points,  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ . The observed value of  $y$  coordinate at  $x = x_i$ , is  $y_i = (P_i, M_i)$ , while the predicted value is  $y_i = (L_i, M_i)$ . The difference between the observed value and the expected value is the error  $e_i = \{(P_i, M_i) - (L_i, M_i)\}$ . The errors can be positive or negative.  $S = e_1^2 + e_2^2 + \dots + e_i^2 + \dots + e_n^2$  is the sum of the squares of the errors, and the best fit curve is the one that has the smallest sum.

Because there is a linear relationship between the dependent variable and the parameters, LSM is easily utilized to identify parameters of linear equations. If the model is not linear, it is either translated into a linear equation or alternative non-linear procedures are used. To convert from non-linear to linear forms, employ transformations like logarithms, inversions, and exponentials.

Here, we consider linear, quadratic, and exponential curves to find the line of best fit. A linear equation is  $y = ax + b$ , quadratic takes the form  $y = ax^2 + bx + c$  and exponential is represented as  $y = ae^{bx}$ .

### 2.3. Proposed algorithms

The proposed prediction algorithm predictionOfQuality estimates the success rate of fish. The algorithm accepts the category of the fish, details of the landing center, the years (start and end years) where data is available, and the year of prediction as input. getCountOfHighQuality and getTotalPurchase are two procedures used by this algorithm to retrieve the success count and the total purchase of the fish from the data store. Suitable queries are used by these procedures to retrieve the concerned data. Subsequently, the success rate of each year is calculated. Then, the bestFitModelFinder algorithm is used to find the best fit model for the given input data.

The bestFitModelFinder algorithm accepts the years ( $x$ ) and the corresponding success rates ( $y$ ). Initially, half of the data input is passed to the bestFitModelFinder algorithm to find the best model. The Least-Square Method is used in the algorithm. The algorithm compares the  $y$  values of a line, a parabola, and an exponential curve to the deviations of the  $y$  values from the equations. The squares of each of the variances are computed. Then the lowest of the three values is chosen. The fitStraightLine algorithm will be called, for example, when the squares of the deviations are the smallest from a line compared to others. The predicted success rate values are returned to the bestFitModelFinder algorithm which in turn returns the values to the predictionOfQuality algorithm.

The standard error is determined, and if it is below acceptable limits (less than 5%), the entire dataset is sent into the bestFitModelFinder algorithm, which finds the best-fit curve. Otherwise, the initial data point is removed, and the remaining values are sent to the bestFitModelFinder algorithm, which compares the input values to a line, parabola, or exponential curve.

The value that results in the least-squares of deviations is then picked, and the predictionOfQuality algorithm gives the projected success rate values, as previously. The first two data points are then removed, and the bestFitModelFinder method utilizes the remaining values if the standard error is still outside the permitted range. This technique is done until the standard error is acceptable. The best-fit model is defined by the curve that results. The resulting curve equation is used to predict the success rate of fish for a given year. The proposed algorithms are given below:

**algorithm predictionOfQuality**

**input:** integer startYear, endYear, landingCenterCode, predictedYear

string fishCategory

**output:** integer predictedFishQualityRate

**begin**

**var** yearDiff, year[], highQualityCount[], totPurchase[], highQualityRate[], yearHalf[], highQualityRateHalf[], SEValue

yearDiff=(endYear – startYear)

**for each** integer i from 0 to (yearDiff – 1) **do**

year[i]=startYear

highQualityCount[i]= algorithm

getCountOfHighQuality

(landingCenterCode, fishCategory, year[i])

totPurchase[i]= algorithm

getTotalPurchase(landingCenterCode, fishCategory, year[i])

highQualityRate[i]=  $\left( \frac{\text{highQualityCount}[i]}{\text{totPurchase}[i]} \times 100 \right)$

startYear=(startYear + 1)

**end for**

yearHalf[]=copy(year[],  $\frac{\text{year}[\text{.length}]}{2}$ )

highQualityRateHalf[]=copy  
(highQualityRate[],  $\frac{\text{highQualityRate}[\text{.length}]}{2}$ )

predictedFishQualityRate[] = algorithm

bestFitModelFinder(yearHalf[], highQualityRateHalf[])

SEValue =

StandardError

(year[], highQualityRate[], yearHalf[], length, predictedFishQualityRate[])

**if** SEValue is acceptable **then**

predictedFishQualityRate[] = algorithm

bestFitModelFinder (year[], highQualityRate[])

**else**

**for each** integer i from 1 to year[].length **do**

XX[] = copy(year[], i, year[].length)

YY[] = =

copy(highQualityRate[], i, highQualityRate[].length)

predictedFishQualityRate[] = algorithm

(XX[], YY[])

SEValue =

StandardError

(XX[], YY[], XX[].length, predictedFishQualityRate)

**if** SEValue is acceptable **then**

break

**end if**

**end for**

**end if**

predictedFishQualityRate[] = algorithm

bestFitModelFinder(XX[], YY[])

Use the resulting best model equation to find the predictedFishQualityRate for the predictedYear

Plot the curve

**end**

**algorithm bestFitModelFinder**

**input:** year[], highQualityRate[]

**output:** integer predictedFishQualityRate[]

**begin**

**var** totDevLine=0, totDevParaB=0, totDevExp=0, DevLine[], DevParaB[], DevExp[]

**for each** integer i from 0 to year[].length **do**

DevLine[i]= highQualityRate[i] – LineEquation(year[i])

totDevLine= totDevLine + DevLine[i]<sup>2</sup>

**end for**

**for each** integer i from 0 to year[].length **do**

DevParaB[i]= –  
highQualityRate[i]

ParabolaEquation(year[i])

totDevParaB = totDevParaB + DevParaB[i]<sup>2</sup>

**end for**

**for each** integer i from 0 to year[].length **do**

DevExp[i]= –  
highQualityRate[i]

ExponentialEquation(year[i])

totDevExp = totDevExp + DevExp[i]<sup>2</sup>

**end for**

**if** ((totDevLine < totDevParaB) andand(totDevLine < totDevExp))

**then**

predictedFishQualityRate[]=algorithm

fitStraightLine(year[], highQualityRate[])

**else if** ((totDevParaB < totDevLine) andand(totDevParaB < totDevExp))

**then**

predictedFishQualityRate[]=algorithm

fitParabola(year[], highQualityRate[])

**else**

**for each** integer i from 0 to year[].length **do**

year[i]=log10 year[i]

**end for**

predictedFishQualityRate[]=algorithm

fitStraightLine(year[], highQualityRate[])

**end if**

**end**

**algorithm fitStraightLine**

**input:** year[], highQualityRate[], n

**output:** integer predictedFishQualityRate[]

**begin**

$X \leftarrow \sum_{i=1}^n \text{year}[i]$

$\bar{Y} \leftarrow \frac{\sum_{i=1}^n \text{highQualityRate}[i]}{n}$

$m \leftarrow \frac{\sum_{i=1}^n (\text{year}[i] - \bar{X})(\text{highQualityRate}[i] - \bar{Y})}{\sum_{i=1}^n (\text{year}[i] - \bar{X})^2}$

$b \leftarrow \bar{Y} - m\bar{X}$

**for each** integer i from 0 to year[i].length **do**

predictedFishQualityRate[i] = m × year[i] + b

**end for**

return

**end**

**algorithm fitParabola**

**input:** year[], highQualityRate[], n

**output:** integer predictedFishQualityRate[]

**begin**

calculate  $\sum \text{year}, \sum \text{highQualityRate}, \sum \text{year} \times$

highQualityRate,  $\sum \text{year}^2, \sum \text{year}^2 \times$

highQualityRate,  $\sum \text{year}^3, \sum \text{year}^4$

Substitute the above values in the following equations.

$\sum \text{highQualityRate} = na + b \sum \text{year} + c \sum \text{year}^2$

$\sum \text{year} \times \text{highQualityRate} = a \sum \text{year} + b \sum \text{year}^2 + c \sum \text{year}^3$

```

 $\sum year^2 \times highQualityRate = a \sum year^2 + b \sum year^3 + c \sum year^4$ 
Gauss-Jordan elimination method is used to solve the
equations and to calculate a, b, and c
for each integer i from 0 to year[i].length do
  predictedFishQualityRate[i] = a + b  $\times$  year[i] +
  c  $\times$  year[i]2
end for
return
end

```

### 3. Results and discussion

The testing of the algorithms was done using the data collected from a Seafood company. The data span the period from January 2000 to December

2019. The results of Cephalopods and Crustaceans fish families are given below.

Fig. 2 depicts the success rate prediction of the Cephalopods fish family from a catching center (C1). The bestFitModelFinder program examines the input data and determines that a parabolic curve is the best model for fitting the data. The success rates of Cephalopods are plotted from 2000 to 2019, and the generated quadratic equation is used to estimate the future year's success rate.

Fig. 3 depicts the success rate prediction of the Crustaceans fish family from a catching center (C1). The best model for the input data is found to be linear.

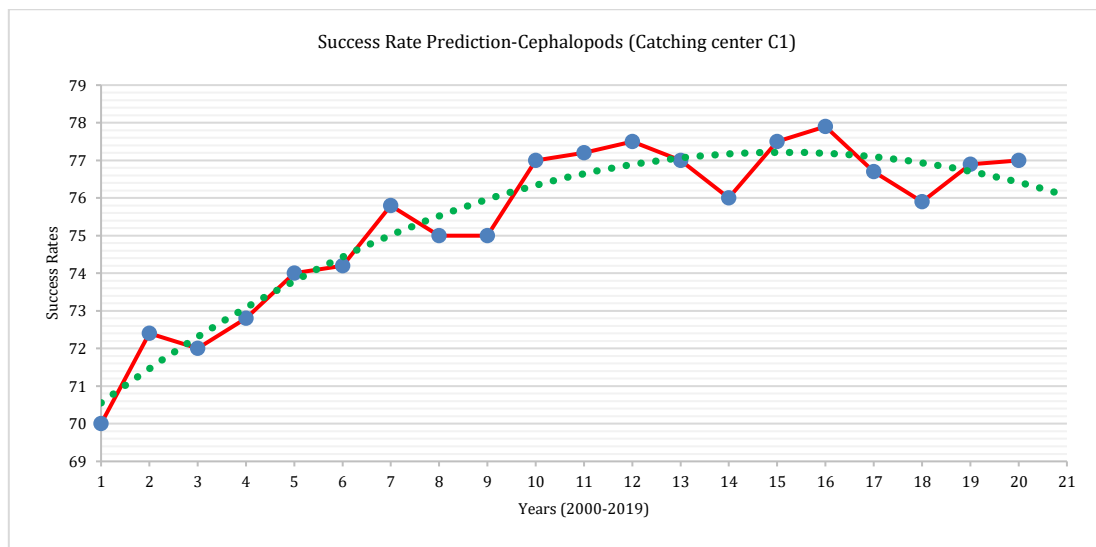


Fig. 2: Success rate prediction-cephalopods (Catching center C1)

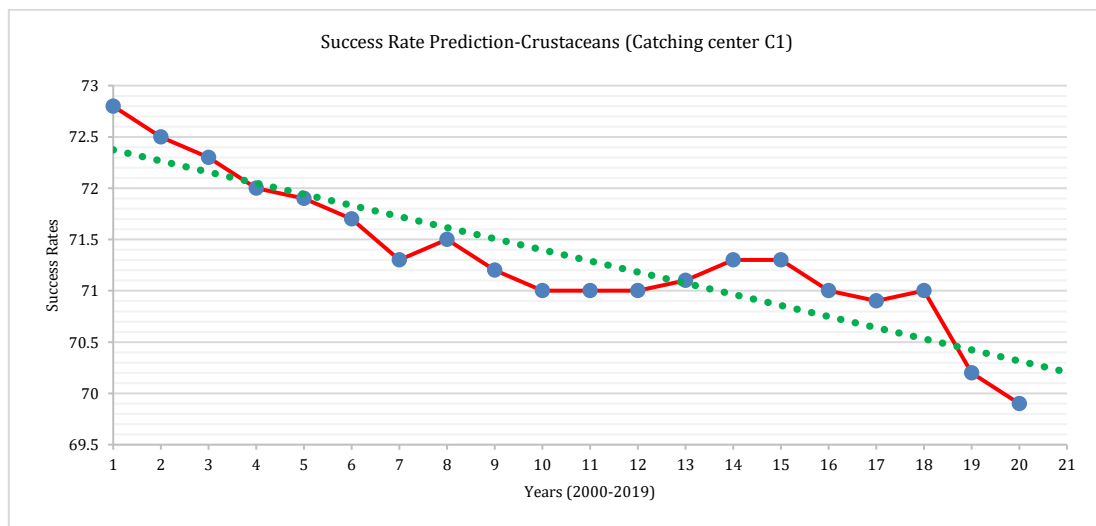
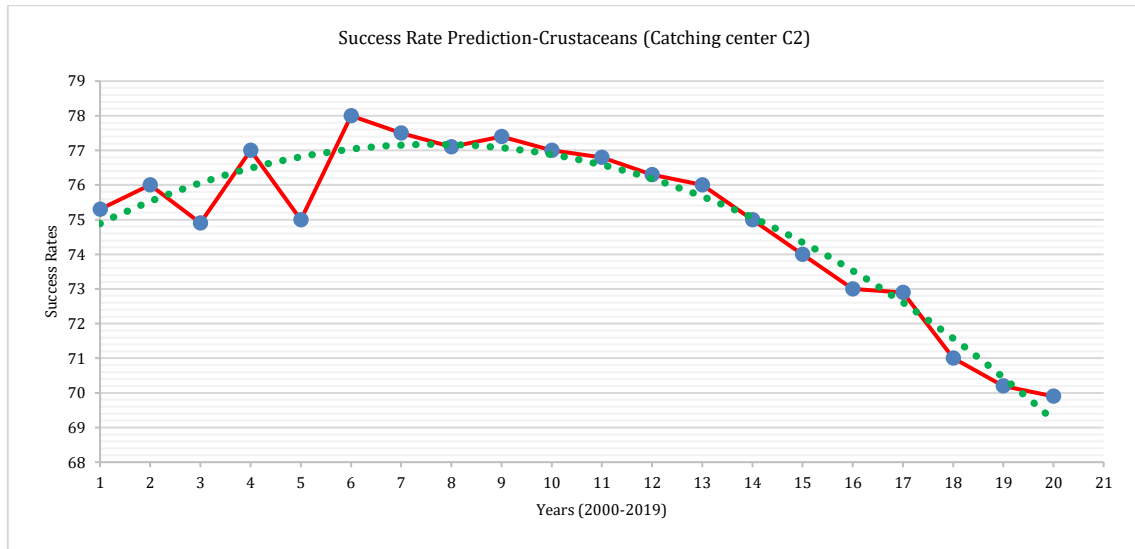


Fig. 3: Success rate prediction-crustaceans (Catching center C1)

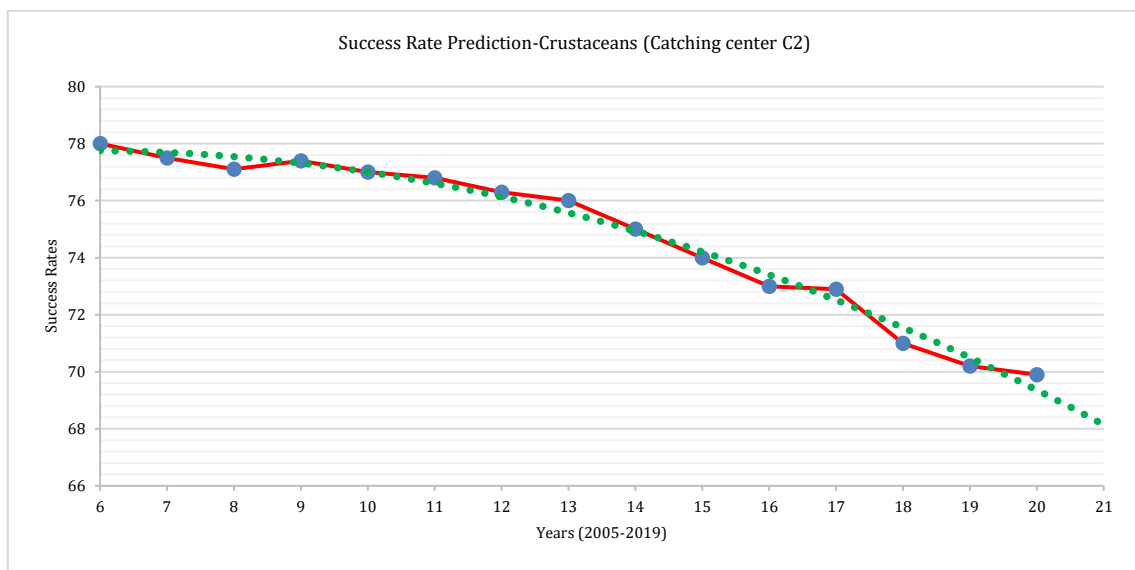
Fig. 4 depicts the success rate prediction of the Crustaceans fish family from a catching center (C2). The bestFitModelFinder program analyzes the data and determines that the square of the variations between the success rates of a parabola equation is the smallest and that the structure is uneven. The

standard error, however, is found to be outside of the permitted range. As a result, the method discards the first five data points before determining the best-fit curve (parabolic curve) within the allowable standard error. The updated model is given in Fig. 5.





**Fig. 4:** Success rate prediction-crustaceans (Catching center C2)



**Fig. 5:** Success rate prediction-crustaceans (Catching center C2) (Updated model)

#### 4. Performance analysis

The statistical measurements of mean absolute deviation (MAD), mean square error (MSE), root mean square error (RMSE) and mean absolute percentage error (MAPE) are used to assess the prediction accuracy of the presented algorithms (Myttenaere et al., 2016).

The mean absolute deviation is a typical method for calculating total prediction error. This value is calculated by dividing the sum of the absolute values of the individual prediction errors by the sample size. The mean absolute percentage error is the average absolute percent error for each prediction minus actuals divided by actual. The mean of the squared difference between the predicted and observed values is referred to as mean square error. The formulas of the measurements are given below:

$$MAD = \frac{\sum |e|}{n}$$

$$MSE = \frac{\sum e^2}{n}$$

$$RMSE = \sqrt{\frac{\sum e^2}{n}}$$

$$MAPE = \frac{\sum \left[ \frac{|e|}{Actual} \times 100 \right]}{n}$$

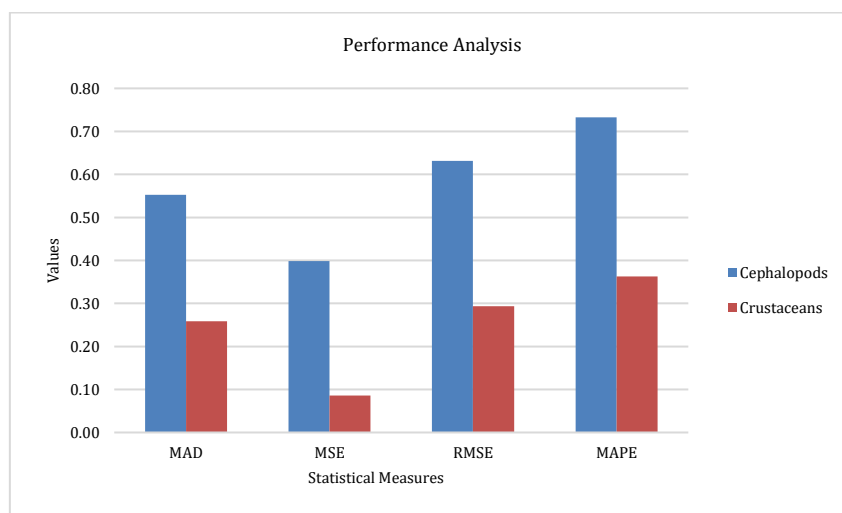
Fig. 6 displays the statistical measures of the two fish families. Compared to Cephalopods, Crustaceans resulted in lower values for all the measures.

#### 5. Conclusion

Food safety and quality concerns are becoming increasingly important in today's world. For people across the world, seafood is a key source of protein. Nonetheless, bacteria and other substances found in seafood represent a significant risk to people. The aim of this research is to propose prediction algorithms based on the Least-Square Method to predict the success rates of seafood based on different catching centers. The bestFitModelFinder algorithm is used to select the most appropriate model for the input data and then the predictionOfQuality algorithm predicts the success rate. Least-Squares-Based Methods are used to identify the best-fit curve. The prediction accuracy of the proposed algorithms was measured using

various statistical measurements and it resulted in lower error values. The algorithms provided can help seafood businesses determine the quality of seafood items sourced from diverse fishing areas.

Further research can be undertaken to investigate different methods to ensure the quality of export seafood, aquaculture, etc.



**Fig. 6:** Performance analysis of the proposed algorithms

## Funding

The research leading to these results has received funding from the Research Council (TRC) of the Sultanate of Oman under the Block Funding Program BFP/ RGP/ ICT/ 18/ 113.

## Compliance with ethical standards

## Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## References

- Freund RJ, Wilson WJ, and Sa P (2006). Regression analysis. Elsevier, Amsterdam, Netherlands.
- García MR, Cabo ML, Herrera JR, Ramilo-Fernández G, Alonso AA, and Balsa-Canto E (2017). Smart sensor to predict retail fresh fish quality under ice storage. *Journal of Food Engineering*, 197: 87-97. <https://doi.org/10.1016/j.jfoodeng.2016.11.006>
- Guest PG and Guest PG (2012). Numerical methods of curve fitting. Cambridge University Press, Cambridge, UK.
- Hardle W and Mammen E (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21: 1926-1947. <https://doi.org/10.1214/aos/1176349403>
- Howell JQ (1971). A least-square distance curve-fitting technique. National Aeronautics and Space Administration, Washington, USA.
- Huss HH (1994). Assurance of seafood quality. Food and Agriculture Organization, Rome, Italy.
- Khoshnoudi-Nia S and Moosavi-Nasab M (2019). Prediction of various freshness indicators in fish fillets by one multispectral imaging system. *Scientific Reports*, 9(1): 1-11.

<https://doi.org/10.1038/s41598-019-51264-z>  
**PMid:31605023 PMCID:PMC6789145**

- Kuhn M and Johnson K (2013). A short tour of the predictive modeling process. In: Kuhn M and Johnson K (Eds.), *Applied predictive modeling*: 19-26. Springer, New York, USA. <https://doi.org/10.1007/978-1-4614-6849-3>
- Motulsky HJ and Ransnas LA (1987). Fitting curves to data using nonlinear regression: A practical and nonmathematical review. *The FASEB Journal*, 1(5): 365-374. <https://doi.org/10.1096/fasebj.1.5.3315805> **PMid:3315805**
- Myttenaere DA, Golden B, Le Grand B, and Rossi F (2016). Mean absolute percentage error for regression models. *Neurocomputing*, 192: 38-48. <https://doi.org/10.1016/j.neucom.2015.12.114>
- Oliveira DCAF, Da Cruz AG, Tavolaro P, and Corassin CH (2016). Food safety: Good manufacturing practices (GMP), Sanitation standard operating procedures (SSOP), Hazard analysis and critical control point (HACCP). In: Barros-Velázquez J (Ed.), *Antimicrobial food packaging*: 129-139. Academic Press, Cambridge, USA. <https://doi.org/10.1016/B978-0-12-800723-5.00010-3>
- Pham H (2006). Springer handbook of engineering statistics. Springer Science and Business Media, Berlin, Germany. <https://doi.org/10.1007/978-1-84628-288-1>
- Picard RR and Cook RD (1984). Cross-validation of regression models. *Journal of the American Statistical Association*, 79(387): 575-583. <https://doi.org/10.1080/01621459.1984.10478083>
- Roos N, Mazharul Islam M, and Thilsted SH (2003). Small fish is an important dietary source of vitamin A and calcium in rural Bangladesh. *International Journal of Food Sciences and Nutrition*, 54(5): 329-339. <https://doi.org/10.1080/09637480120092125> **PMid:12907404**
- Steel RGD and Torrie JH (1960). Principles and procedures of statistics. McGraw-Hill Book Company Inc., New York, USA.
- WHO (2015). WHO estimates of the global burden of foodborne diseases. World Health Organization, Geneva, Switzerland.