

PAPER • OPEN ACCESS

Outliers Detection on Fisheries Commodity Transaction from Local Market in Tual City based on The x-means Clustering

To cite this article: Syahibul Kahfi Hamid *et al* 2019 *J. Phys.: Conf. Ser.* **1424** 012017

View the [article online](#) for updates and enhancements.

You may also like

- [Flood vulnerability assessment using artificial neural networks in Muar Region, Johor Malaysia](#)
A M Wahab and A N Muhamad Ludin
- [Pedagogical Content Knowledge Pre-Service Mathematics Teacher](#)
G Kadarisma, E Senjayawati and R Amelia
- [Modularization with Rule Based Algorithm in Three-Dimensional Design of Engine Room Layout Harbour Tug Diesel – LNG Dual Fuel Engine 2 x 2500 HP](#)
Aguk Zuhdi Muhammad Fathallah, Laurensius Abhisa Valasranggi Prasajo and Semin



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Outliers Detection on Fisheries Commodity Transaction from Local Market in Tual City based on The x-means Clustering

¹Syahibul Kahfi Hamid, ¹Wellem Anselmus Teniwut, ¹Roberto Mario Kabi Teniwut, ²Maimuna Renhoran

¹Fisheries Agribusiness Study Program, Tual State Fisheries Polytechnic, Langgur-Sathean Ringroad Km 6, Southeast Maluku District 97611, Indonesia

²Fishing Technology Study Program, Tual State Fisheries Polytechnic, Langgur-Sathean Ringroad Km 6, Southeast Maluku District 97611, Indonesia

Email: wateniwut@polikant.ac.id

Abstract. Correctly estimating factors that affect the income and profitability of coastal communities specifically catching fishermen is one way to increase the level of sustainable coastal community welfare. The purpose of this study is to detect outliers from fisheries commodity transactions in the fish market in Tual City, to have full view on which cases that may have certain traits that can improve the profitability of fishermen in the region. The research method used in this study is a survey method in which used as many as 1420 datasets obtained from interviews conducted starting from June 2017 to June 2018. The data mining approach is used in this study, in order to obtain the appropriate outliers then x-means clustering is used to cluster profitable transaction data, then data collected further being process with outlier detection. We found 25 high profitability transactions of 1420 datasets. The outlier analysis was carried out and found 1 outliers. We have figured that the species of fish and the location of the seller are the factors that influence the level of profit from the fishermen. This result is an input for policy makers and catch fishermen in conducting capture fisheries.

1. Introduction

Increasing the dynamics of development in an area is highly dependent on an effort to optimize the use of existing resources in the region for the prosperity of the people and for the benefit of all components in the region. Optimization of resource utilization in an area will depend on the knowledge of those who utilize the potential of natural resources In Maluku province there are three fisheries management areas (WPP), respectively: WPP-714 Banda Sea, WPP-715 Arafura Sea and WPP-716 Seram Sea, Maluku Sea and Tomini Bay. The Kei Islands is one of the regencies and cities in Maluku, which is between two WPPs that needs good capabilities and good knowledge. The potential of capture fisheries in the Kei Islands and Maluku Province in general according to [1] is still a source of income for fishermen in this region but refers to the existing conditions and trends related to behavior, fleets and techniques including illegal fishing so the tendency of the development of marine resources in particular fish will decrease in the next 8-10 years. The Kei Islands are included in the coral triangle waters region, so it is important to maintain the rich biodiversity of marine resources in the waters.

Based on studies by [2] showed that a total of over IDR 240 million was generated on average by each of the coastal villages in the Kei Islands, including the coastal villages in Tual City. This value is



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

still below an optimal number based on the daily needs of coastal communities and the population of coastal communities added with given the characteristics of coastal communities in this region which has a very high number of family dependents with an average low level of education [3]; [4]; [5]; [6]. This condition is compounded by the financial and capital capability to procure better fishing gear. This causes income from capture fishermen in this region to be very inconsistent, one of the causes which also one cliché problem of coastal communities in the Kei Islands is a result of the use of technology that is still limited and limited knowledge [5]. Another factor that is an obstacle to improving the welfare of coastal communities is the potential for conflicts that arise both between villages, in villages between families in the village due to the use of marine resources [6]; [7]. For this reason, research is needed that not only improves the efficiency of fish catching activities, but also avoids potential conflicts and increases in income of coastal communities. A data mining approach can be used to help resolve this problem.

Data mining is process where an intelligent method are applied to extract knowledge and often associated with term knowledge discovery [8]. The data mining approach has been widely used by various fields including in the fields of fisheries and marine affairs [9]; [10]; [11]; [12]. Utilization of this technique is very helpful in breaking down information and knowledge from the data obtained. To increase fish catch efficiency by fishermen and increase fishermen profits, it can be obtained by identifying fish species, catch time and fishing location. Thus, the purpose of this study is to detect outliers from consumer transaction datasets on fisheries commodities in the local fish market in Tual City, Kei Islands, which provide the highest profit value to obtain the empirical characteristics of these transactions to obtain a fishing profile that benefits fishermen in the area.

2. Methodology

The stages of this research can be seen in Fig. 1, where the data mining approach used in this study consists of two parts, namely cluster analysis using the x means clustering approach which then clusters with the highest profits will be used to be analyzed for outliers to identify transactions that are outside trend and has a unique profile, namely transactions that provide high profits with fish species, catch time and unique fishing locations. The data used in this study were 1420 datasets obtained from a survey for one year in May 2017-June 2018 on fish commodity transactions in the local fish market in Tual City.

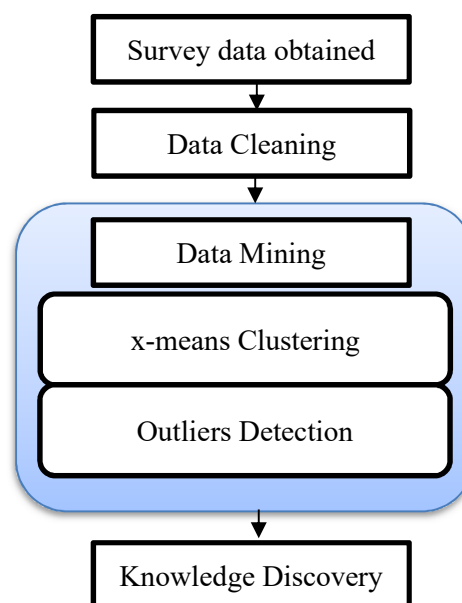


Figure 1. Research framework

2.1. X-means clustering

X-means clustering is a variation of K-means clustering, x-means clustering is a variation of K-means clustering. The x-means clustering algorithm is summarized as follows [13]:

1. Set initial number of clusters k_0 , then applied k-means algorithm into all data ($k_0 = k$) and repeat all procedure (setting $i = 1, 2, \dots, k_0$).
2. Apply k-means cluster with setting $k = 2$ for cluster $C_1, C_2 \dots, C_{k_0}$.
3. Assume p -dimensional distribute normal for data x in C_1 . Calculate BIC as, $BIC = -2 \log L(\hat{\theta}_i; x \in C_i) + q \log n_i$ and BIC' as, $BIC' = -2 \log L'(\hat{\theta}_i'; x \in C_i) + q' \log n_i$.
Where, $\hat{\theta}_i$ is the maximum likelihood estimation of p -dimensional, q is number of dimensions of the parameters and becomes $2p$, n_i is the number of elements contained C_1 and $L(.)$ is likelihood function, q' is given by $2 \times 2p = 4p$.
4. When $BIC > BIC'$ continue the division and if the vice versa, $BIC < BIC'$, divide further clusters and stop.
5. Number the cluster output identification from the center of each cluster, and the number of elements in each cluster.

2.2. Outliers detections

Anomaly detection identifies single records in datasets which significantly deviate from the normal data, further it is process of finding patterns in a given dataset which deviate from the characteristics of the majority. We can use the maximum likelihood method to estimate the parameters μ and σ by maximize the log-likelihood function, as follows [8]:

$$\ln \mathcal{L}(\mu, \sigma^2) = \sum_{i=1}^n \ln f(x_i | (\mu, \sigma^2)) = -\frac{n}{2}(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Where n is the total number of samples, for derivatives μ and σ^2 solving the first order condition leads to maximum likelihood estimates:

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

3. Result and Discussion

3.1. Clustering analysis

The results of data analysis using the x-means clustering approach found that there were 4 clusters, each of which was divided based on the main criteria, namely the level of income received from fishermen. Measurement of cluster distance using euclidean distance with the average distance of each cluster is 0.950 with Davies-Bouldin Index of 0.752. In Fig 2, it appears that cluster 3 is a cluster that is used to the next stage for detecting outliers. Obtained 25 transaction data from 1420 datasets which have a price of only 5.97% higher than other datasets, but have an earning rate of 604.59% greater than other datasets and a higher sales amount of 517.73%.

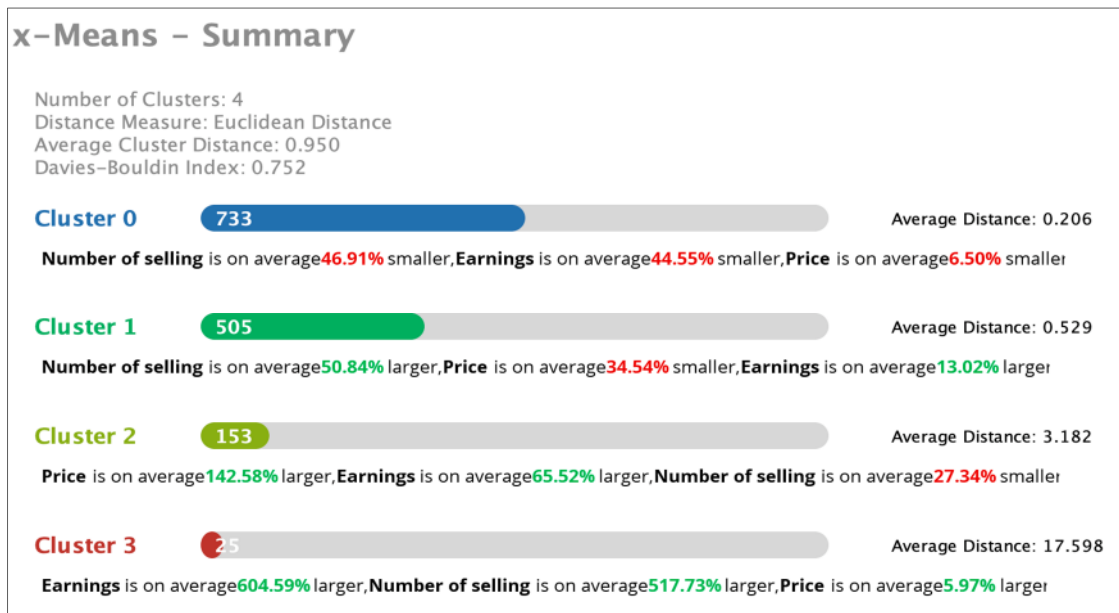


Figure 2. x-means clusters summary

Table 1. Cluster 3 datasets

No	Year	Period (month)	Seller Origin	Fish' Species	Number of selling	Price	Earnings	Fishing Ground
1	2017	June	Fair	Lethrinidae	143	50.000	7.150.000	Langgiar
2	2017	June	Warbal	Stolephorus	300	10.000	3.000.000	Debut
3	2017	July	Mangon	Sardinella maderensis	430	10.000	4.300.000	Kaimana
4	2017	July	Komp Banda Ely	Auxis rochei	270	30.000	8.100.000	Banda Ely
5	2017	July	Ngursit	Auxis rochei	145	50.000	7.250.000	Ngursit
6	2017	July	Sathean	Sardina pilchardus	290	20.000	5.800.000	Ohoililir
7	2017	July	Sathean	Sardina pilchardus	195	20.000	3.900.000	Ohoililir
8	2017	July	Banda Ely	Auxis rochei	150	25.000	3.750.000	Banda Ely
9	2017	July	Banda Ely	Auxis rochei	203	25.000	5.075.000	Banda Ely
10	2017	July	Banda Ely	Moolgarda seheli	165	25.000	4.125.000	Kaimana
11	2017	Jully	Wara	Auxis rochei	211	25.000	5.275.000	Ngursit
12	2017	July	Wara	Auxis rochei	213	25.000	5.325.000	Kei Besar
13	2017	July	Un	Sardina pilchardus	500	20.000	10.000.000	Pulau Ut
14	2017	July	Fair	Auxis rochei	39	30.000	1.170.000	Ngursit
15	2017	July	Banda Ely	Auxis rochei	150	25.000	3.750.000	Banda Ely
16	2017	August	Wara	Katsuwonus pelamis	70	50.000	3.500.000	Perairan
17	2017	August	Fair	Auxis rochei	315	20.000	6.300.000	Tan. Kei
18	2017	August	Mangon	Rastrelliger faughni	150	25.000	3.750.000	Ngursit
19	2017	August	Un	Stolephorus	60	10.000	600.000	Dobo
20	2017	August	Banda Eli	Auxis rochei	15	50.000	750.000	Lupus
21	2017	August	Kiom	Parexocoetus brachypterus	150	20.000	3.000.000	Ngursit
22	2017	August	Fidabot	Ocyurus chrysurus	20	50.000	1.000.000	Perairan
23	2017	August	Fidabot	Stolephorus	70	10.000	700.000	Dullah Laut
24	2017	August	Fair	Auxis rochei	200	20.000	4.000.000	Tam
25	2018	April	Kompleks Banda	Selar crumenophthalmus	60	20.000	1.200.000	Lupus
								Ngursit
								Sathean

3.2. Outliers detection

Based on the results of the x-means cluster analysis that has been done in the previous stage, then it has been used to detect outliers. Thus, it is expected to produce outliers that have unique characteristics and are out of trend to further serve as a reference for fishermen and fishery commodity sellers in Tual City in running related businesses. In Fig 3a, a comparison of fish species and fishing ground is seen that has unique characteristics. It is seen that on Langgiar fishing ground which is located in Kei Besar Selatan Barat sub-district, Southeast Maluku district with fish species that being capture was Lethrinidae. Then when seen in Fig. 3b, information on catch periods has been obtained in June which is included in The Indian summer monsoon. Thus, both fish species and period of catching information have been found, which it has the potential to provide more benefits to fishermen and sellers in terms of earning and profit.

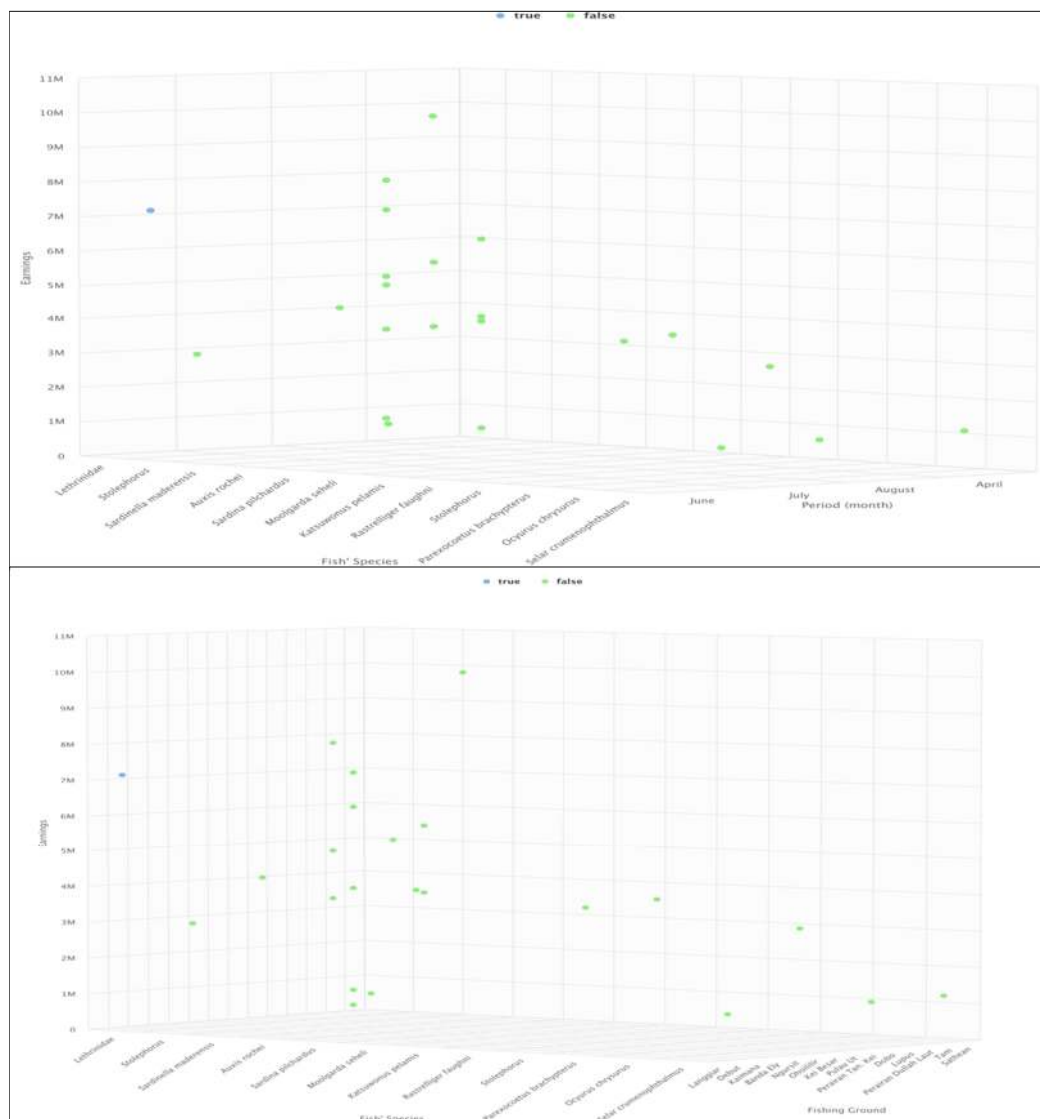


Figure 3. Outliers detection: (a) combination between period of catching and fish species; (b) combination between fishing ground and fish species

4. Conclusion

Based on the results of research that has been done, x-means clustering gained 25 datasets out of 1420 datasets of fishery commodity transactions in the local fish market in Tual City which has high profit characteristics. Further results were obtained that, there were 1 data outliers from 25 datasets included

in cluster 3. From the information of the outliers, it is obtained that the benefits obtained by the data are 57% higher than the average datasets in cluster 3, with a selling price of 53% higher than the average price in cluster 3. The results also found two factors that differentiate the fish species, Lethrinidae and the fishing period in June. Implication of research is that fishermen in this area should pay attention to fish species that are caught and sold because it is very influential on the income earned, this is related to taste of customer and distance of fish capture. In addition, fishers and seller also have to put more attention to period of fishing (monsoon), so it could be more efficient considering the natural and marine conditions in the Kei Islands, this is related to the demand and availability of fish in the market.

Acknowledgement

The authors thank Directorate of Research and Community Service, Ministry of Research and Technology the Higher Education Republic of Indonesia for funding this research under PTUPT (Penelitian Terapan Unggulan Perguruan Tinggi) scheme 2019-2020.

References

- [1] Teniwut, W. A 2016 For sustainable revenue of fisheries sector in small islands: evidence of Maluku, Indonesia. *Aquaculture, Aquarium, Conservation & Legislation*, 9(3), 722-732.
- [2] Picaulima, S., Teniwut, W. A., Kahfi, S., Teniwut, R. M., Susanti, I. I. D. A. R., Hungan, M., ... and Ngabalin, A. M 2017 Mapping marine resources utilization based on seascapes area: a study on gender comparison. In *IOP Conference Series: Earth and Environmental Science* (Vol. 89, No. 1, p. 012027). IOP Publishing.
- [3] Hamid, S. K., Teniwut, W. A., Teniwut, R. M., Rahantoknam, M. A., Hasyim, C. L., and Hungan, M 2017 The Support of MPA (Marine Protected Area) in Coral Triangle Area: Evidence from Kei Islands, Indonesia. In *IOP Conference Series: Earth and Environmental Science* (Vol. 89, No. 1, p. 012025). IOP Publishing.
- [4] Teniwut, R. M., Hasyim, C. L., and Teniwut, W. A 2017 Resource-Based Capability on Development Knowledge Management Capabilities of Coastal Community. In *IOP Conference Series: Earth and Environmental Science* (Vol. 89, No. 1, p. 012017). IOP Publishing.
- [5] Teniwut, W. A., and Teniwut, R. M 2018 Minimizing the instability of seaweed cultivation productivity on rural coastal area: a case study from Indonesia. *Aquaculture, Aquarium, Conservation & Legislation*, 11(1), 259-271.
- [6] Teniwut, W. A., Teniwut, Y. K., Teniwut, R. M., and Hasyim, C. L 2017 Family vs Village-Based: Intangible View on the Sustainable of Seaweed Farming. In *IOP Conference Series: Earth and Environmental Science* (Vol. 89, No. 1, p. 012021). IOP Publishing.
- [7] Yamazaki, S., Resosudarmo, B. P., Girsang, W., and Hoshino, E 2018 Intra-village and inter-village resource use conflict in Indonesia: The case of the Kei Islands. *Ocean & coastal management*, 155, 50-59.
- [8] Han, J., Pei, J., and Kamber, M 2011 *Data mining: concepts and techniques*. Elsevier.
- [9] Tsai, W. P., Huang, S. P., Cheng, S. T., Shao, K. T., and Chang, F. J 2017 A data-mining framework for exploring the multi-relation between fish species and water quality through self-organizing map. *Science of the Total Environment*, 579, 474-483.
- [10] Su, F., Zhou, C., Lyne, V., Du, Y., and Shi, W 2004 A data-mining approach to determine the spatio-temporal relationship between environmental factors and fish distribution. *Ecological Modelling*, 174(4), 421-431.
- [11] Shen, W., Guo, X., Wu, C., and Wu, D 2011 Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm. *Knowledge-Based Systems*, 24(3), 378-385.

- [12] Cheng, Y., Jiang, M., and Yuan, D 2009 Novel clustering algorithms based on improved artificial fish swarm algorithm. In *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery* (Vol. 3, pp. 141-145). IEEE.
- [13] T. Ishioka, T 2005 An Expansion of X-means for Automatically Determining the Optimal Number of Clusters, Proc. of The 4th IASTED International Conference on Computational Intelligence, pp.91-96.