Clustering Lakes in Kampar by Using Morphometry Data and Machine Learning Analysis

A Hendrizal\textsuperscript{1*}, P E Plaimo\textsuperscript{2}, Y Harjoyudanto\textsuperscript{1}, M Fauzi\textsuperscript{1}

\textsuperscript{1}Aquatic Resources Management, Universitas Riau
\textsuperscript{2}Faculty of Agriculture and Fishery, Universitas Tribuana Kalabahi

*andri.h@lecturer.unri.ac.id

Abstract. Lakes have an important role in hydrological and biochemical cycle. It also has some other crucial role such as domestic and industrial water use as well as irrigation. The monitoring and management of this aquatic resources is crucial. But with many numbers of lakes, it is very challenging to manage them all. Clustering lakes can provide the answer so the management of the same cluster lakes may be done efficiently. Within this study, morphometry data of 6 lakes in Kampar Regency, Riau Province, were analyzed by using one of artificial intelligence branch which is machine learning. Morphometrical data are collected by using information geographic system. These data then categorized by using python language. This categorization based on data mining categorization algorithm named K-means. Based on the K-means machine learning clustering, the optimum cluster based on Elbow method is \( k = 3 \). But there is a possibility to look around for \( k = 2 \). Based on \( k = 3 \), cluster 3 is defined as the lowest values of all attribute. Based on \( k = 2 \), the lowest value of morphometry data will be in the cluster 1. These data will not only provide basic data such as total area, shape, width and length, but also help to understand the large scale hydrological models.

1. Introduction
Lakes are one of the important ecosystems. It contains almost half of all water on the Earth’s surface. Lakes provide services such as drinking water, waste removal, fisheries, agricultural irrigation, industrial activity and recreation [1] [2]. A lot of organisms rely on freshwater for the survival. Human also depend on the freshwater. Freshwater are one of the primary needed for the ecosystem and human society [3]. For years, human civilization have concentrated around different types of water bodies [1]. It shows how important lakes to many anthropogenic activities.

The availability of freshwater has become big issues for many countries. Factors such as climate change, urbanization and population increase the rate of freshwater source degradation [4]. Since 1970, almost 35% of the aquatic resources has been declined. It almost three times of forest loss rate [3]. Therefore, the monitoring of lake and other inland resources is important.

Riau province has many aquatic resources such as river and lakes. One of regency in Riau is Kampar. There are many lakes in Kampar especially oxbow lake. Theses oxbow lakes need to be monitor frequently especially for conservational purpose.
Several techniques could be used to monitor lakes including physical, chemical, and biological aspects. Most of these methods are time-consuming and highly expensive. To overcome this issue, Geographical Information System (GIS) has been introduced to monitor lakes [5]. This method is now widely used than any other methods which is costly and time-consuming [6]. Especially for the change detection of water quality and surface morphometry of lakes. The continuous monitoring and evaluation of lake's water are very important keys for sustainable protection and management [7].

Lake surface morphometry is needed to obtain many information of the lake. Lake morphometry regulates nutrient load, primary production, and secondary production of zooplankton, zoobenthos, and fish [8]. In aquatic resources management, classification or clusterization is an important way to manage and monitor aquatic resources [9]. Most of classification or categorization was done by using biological and chemical aspects. Physical aspect clusterization is less done by researchers. Most of the lake classification or clustering by using physical aspect only focus on the total area (Ao) without considering another aspect such as SDI or length and width. This study tried to propose clusterization from morphometry data by using machine learning as a branch of artificial intelligence. Machine learning is a subset of Artificial Intelligence (AI) with which the machines learn from the data and are trained to make predictions or forecasts based on the past data. Machine learning has been widely used in remote sensing image classification for various applications [10]. Machine learning is a subset of Artificial Intelligence (AI) with which the machines learn from the data and are trained to make predictions or forecasts based on the past data. Machine learning has been widely used in remote sensing image classification for various applications.

2. Methods
Lake Morphometry
Lake morphometry data was collected by using geographical information system (GIS). 6 oxbow lakes satellite imaging was collected. 7 data information on morphometry aspect was collected such as maximum length (Lmax), effective maximum length (Le), width of the lake (Wmax), effective maximum width (W≡), total area (Ao), shore length (SL) and average of width (W) [2] [8]. The data was collected from satellite imaging record in July 2019. These imaging data then be processed by using geographical information system program (Quantum GIS).

Clustering by Machine Learning
Machine Learning (ML) is the branch of computer science that empowers the computer to learn like the way humans do. It is a subset of Artificial Intelligence (AI) and has found its way into solving most of the scientific problems of today [9]. Many machine learning algorithms are being used by the researchers that can be categorized primarily into regression and classification. Regression algorithms basically try to predict the output value based on the inputs whereas classification algorithms try to group the input data into different classes [10]. This study was used unsupervised classification. The algorithm used was K-Means clustering that doesn’t require the system to be trained before it can learn the data provided.

3. Result and Discussion
Lakes Morphometry
The results of the measurement of the morphometric aspects of 6 lakes in Kampar Regency can be seen in table 1. Based on the measurement results, all lakes have the same Lmax and Le. Likewise with the effective length and effective width. This is because there are no islands or land in the lake.
Among the effective factors considered in evaluating the annual water balance of the lake, surface area is of the most importance, since it directly influences parameters like evaporation. On the other hand, continuous falling trend in water level and progressive changes in coastlines during recent decade. Retreat of Urmia Lake from its original shoreline is not only a hydrological concern, but it also presents serious challenges for water quality, conservation, human health and economics. Dried coastal salt lands, which leads to salt marshes creation, has brought to attention the importance of our knowledge about the water surface area. Since direct measurement of the lake’s surface area is costly and time-consuming, therefore development or employment of new methods and tools for this purpose will be valuable [6].

The morphometry of a water body can be crucial for its sensitivity to pollution. This is because morphometry influences sediment and bottom dynamic conditions, but also because morphometry influences the theoretical water retention time. The latter is significant since the in and outflowing water either dilutes (‘purifies’) or pollutes the water body[11]. The larger the SDI value, the more irregular the shape of the lake is. The length of the perimeter and the SDI value will determine the amount of nutrients that enter. The longer the line around the lake and the greater the SDI value, the greater the input received by the lake. SDI can be used to describe the level of productivity of a waters. The higher the value, the more fertile the waters are [8]. The level of water productivity is closely related to the increasingly irregular shape of the lake. The more parts that are bayed and connected to the land, the greater the possibility of nutrients entering from the land.

### Lake Clustering
The continuous anthropogenic activities in bodies of water around the world, bodies of water are continually deteriorating and assisting in destroying the environment because of other chemicals that it contains. Thus, the classification of body water is one of the most important ways to manage and monitor the quality of water resources [9].

There are four popular techniques developed by researchers over the years to perform surface water delineation viz, single band density slicing, spectral index based, machine learning based classification and spectral unmixing based methods [10]. Among these techniques, machine learning based methods are become popular to be used because it could give accurate results. The clustering result of 6 lakes in Kampar can be found in Figure 2.

### Table 1. Lakes Morphometry Data

| No | Name of The Lake | Lmax  | Lp  | Wmax  | Wp  | Ao    | SL    | W    | SDI    |
|----|------------------|-------|-----|-------|-----|-------|-------|------|--------|
| 1  | Bakuok           | 1184,93 | 1184,93 | 143,62 | 143,62 | 132058,3 | 2540,22 | 111,4482 | 3,944787 |
| 2  | Baru             | 1121,53 | 1121,53 | 78,33  | 78,33  | 72349,29  | 2419,47  | 64,50946 | 5,076195   |
| 3  | Lubuk Siam       | 2499,2  | 2499,2  | 68,34  | 68,34  | 149265,8  | 5126,41  | 59,72542 | 7,488044   |
| 4  | Pinang Dalam     | 785,67  | 785,67  | 73,77  | 73,77  | 43823,42  | 1646,26  | 55,77841 | 4,43793    |
| 5  | Pinang Luar      | 2917,69 | 2917,69 | 110,21 | 110,21 | 202575,3  | 6057,8   | 69,43001 | 7,595509   |
| 6  | Raya             | 369,65  | 369,65  | 83,81  | 83,81  | 21332,01  | 861,32   | 57,70867 | 3,328004   |
Based on the K-means machine learning clustering, the optimum cluster based on Elbow method is k=3. But there is a possibility to look around for k=2. Based on K=3, cluster 3 is defined as the lowest values of all attribute. Based on k=2, the lowest value of morphometry data will be in the cluster 1.

An ecological approach to surface water assessment and management under WFD ensured a vast amount of ecological data obtained in freshwater monitoring programmes both at the national and European Union (EU) scale. However, the limitations of monitoring data, such as their extensiveness, variability, gaps and multiple sources of errors, can limit their effective use. The above characteristics of the data obtained in aquatic monitoring programmes allow them to be classified as big data. The use of big data in various fields of science, including freshwater research, has become common in recent years [12].
References

[1] R. Bhateria and D. Jain, *Sustain. Water Resour. Manag.*, vol. 2, no. 2, pp. 161–173, Jun. 2016, doi: 10.1007/s40899-015-0014-7.

[2] A. Ramaraju, p. 10, 2015.

[3] Z. Jiang, W. Jiang, Z. Ling, X. Wang, K. Peng, and C. Wang, *Water*, vol. 13, no. 2, p. 138, Jan. 2021, doi: 10.3390/w13020138.

[4] R. G. Abd Ellah, *Egypt. J. Aquat. Res.*, vol. 46, no. 1, pp. 1–12, Mar. 2020, doi: 10.1016/j.ejar.2020.03.001.

[5] *Am. J. Geogr. Inf. Syst.*, p. 7, 2019.

[6] J. Ahmadi, D. Kahforoushan, E. Fatehifar, K. Z. Benis, and M. Nadjafi, p. 5, 2015.

[7] H. A. El-Serehy, H. S. Abdallah, F. A. Al-Misned, S. A. Al-Farraj, and K. A. Al-Rasheid, *Saudi J. Biol. Sci.*, vol. 25, no. 7, pp. 1247–1256, Nov. 2018, doi: 10.1016/j.sjbs.2018.05.022.

[8] R. Ridoan, A. Muhtadi, and P. Patana, *DEPIK*, vol. 5, no. 2, Sep. 2016, doi: 10.13170/depik.5.2.4913.

[9] the College of Computer and Information Science at Malayan College Laguna, A Mapua School, Pulo-Dieyro Road, Cabuyao, Laguna, Philippines, J. G. Ebron, R. I. D. D. Leon, A. D. Alejandro, and B. A. Amoranto, *Int. J. Environ. Sci. Dev.*, vol. 11, no. 9, pp. 425–431, 2020, doi: 10.18178/ijesd.2020.11.9.1285.

[10] B. T.V., *J. Adv. Res. Dyn. Control Syst.*, vol. 12, no. 3, pp. 207–216, Mar. 2020, doi: 10.5373/JARDCS/V12I3/20201184.

[11] D. Lindgren, *Open Environ. Sci.*, vol. 5, no. 1, pp. 1–17, Jan. 2011, doi: 10.2174/1876325101105010001.

[12] D. Gebler, A. Kolada, A. Pasztaleniec, and K. Szoszkiewicz, *Environ. Sci. Pollut. Res.*, vol. 28, no. 5, pp. 5383–5397, Feb. 2021, doi: 10.1007/s11356-020-10731-1.