

# Application of Classification Tree Analysis (CTA) to Model the Potential Distribution of Harmful Algal Blooms (HAB's)

Aswin Nur Saputra<sup>1\*</sup>

<sup>1</sup>Geography Education, Faculty of Teacher and Education, Lambung Mangkurat University, Banjarmasin, Indonesia

\*Corresponding author. Email: [aswin.geografi@gmail.com](mailto:aswin.geografi@gmail.com)

## ABSTRACT

Remote sensing has a potential unconventional approach to observe the quality of lake as well as other waters land. Riam Kanan is a reservoir which has a water resource from Riam Kanan River with the wide of watershed is 1043 km<sup>2</sup>. The accumulation of nutrient simultaneously causes the condition of waters at reservoir is getting thriven. The thriven water condition can cause an increasingly growth of harm *micro algae* or *Harmful Algal Blooms* (HABs). This research tries to apply *Classification Tree Analysis* (CTA) method to model the potential distribution of HABs which uses image of satellite Landsat-8 OLI. Landsat 8 OLI image which was recorded on 14 August 2016 was used in this research based on value at surface reflectance. *Classification Tree Analysis* (CTA) method was used to model the potential distribution of HABs at Riam Kanan Reservoir. The result of CTA model then was used to analyse the parameter that affect the potential distribution of HABs. Based on the result of modelling with the total validation model 81,25 %, it is resulted that there are 4 potential classes, they are light, medium, heavy, and extremely heavy classes which the distribution of HABs in a high depth is dominated by medium class, whereas in shallower depth with area of waters that stick out into the water is included in heavy potential class. Potential of load pollution is obtained from outer part of the reservoir especially from dry land agriculture in width 303.811,95 Ha which is known has the amount of potential Nitrogen content of 20.507.306,6 kg and phosphorus with the total content as much as 4.557.179,25 kg.

**Keywords:** *Landsat-8 OLI, CTA Model, Harmful Algal Blooms, HABs*

## 1. INTRODUCTION

A rising of eutrophication at Riam Kanan Reservoir waters will cause another effect that is the growth of the potential of HABs (Harmful Algal Blooms) [1]. It is because of the rate of eutrophication, which keeps on rising; make the growth of the micro algae type of *Microcystis sp.* also increases [2]. Eutrophication level that always keeps rising every year does not only increase the potential of HABs, but also causes the negative effect on ecological condition of reservoir, such as massive death of millions fish because of ammonia poisoning. That case is not added yet by potential of danger when HABs is happened. The danger which will be emerged is not only to the ecology of the reservoir but also to the humans around the reservoir who use it for their daily life. Thus, it is important to know the distribution level of eutrophication that potentially happens to HABs at Riam Kanan Reservoir waters.

An increasing of nutrient in the reservoir comes from some sources. Generally, the resource of the nutrient comes from inlet area of the reservoir, while the source of water of the reservoir comes from inlet area that consists of some stream branches. The potential source on increasing the nutrients and pollution in the reservoir are waste from the population, industry, agricultural, and livestock.

The use of remote sensing image to observe waters quality especially for eutrophication study and *Harmful Algal Blooms* (HABs) to map the distribution has been widely used [3]–[6]. The previous research concerning of HABs was done more on using image with higher spectral resolution on lower spatial resolution. Observation of waters land quality needs remote sensing image with bigger spatial resolution. The use of Landsat-8 OLI is tended to fulfil the need.

According to some previous researches, it shows that CTA method is good in increasing acuration of classification [7], [8]. Besides that, the advantages of CTA are non-parametric method, simple, can handle the relationship between non-linear and noise between input features and label of class, and computation efficiency which enable to do spatial model on potential distribution of HABs. Based on that case, the use of Landsat-8 OLI image for waters quality study and the application on CTA method becomes one of the studies to be observed to know its ability to identify potential of HABs through extraction of information on remote sensing image.

## 2. STUDY AREA

This research was done at Riam Kanan Reservoir which has the wide 9.200 ha on flood water, and 1200 million m<sup>3</sup>

volume on normal water. Riam Kanan Reservoir is a reservoir which has a water resource from Riam Kanan River with the wide of watershed is 1043 km<sup>2</sup>. Riam Kanan Reservoir or Pangeran Mohammad Noor Reservoir is located in Aranio sub district, Banjar Regency, South Kalimantan. This reservoir is included into Barito watershed, and sub watershed of Riam Kanan. This reservoir was built from 1965, the last piling up of primary Dam was done in 1971, and started to be operated in 1972. The benefits of Riam Kanan Reservoir are to irrigate 6000 ha agriculture land, to use 30 MW for hydroelectric power, the source of drinking water of Banjar Regency, flood controller, fishery of floating net, and tourism [9]. Data which was used in this research is image of Landsat-8 OLI level 1T scene path 117 row 62, was recorded on 14 August 2016. Bathymetry measurement data was also used in this research which would be an input parameter in the model. Moreover, this research was also used the visual map of Indonesia which is used to get data of administrative boundary of research area [10].

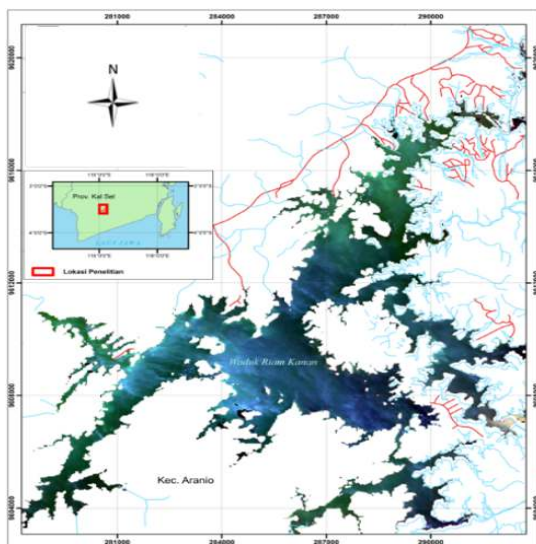


Figure 1 Research location

### 3. METODOLOGY

In this research, the data which was used was 1T level data that has been geometrically corrected automatically and the atmosphere disturbance also has been removed. Image conversion as the TOA reflectance value was done based on algorithm which was issued by United States Geological Survey [11], then was followed by relative atmospheric correction using the dark object subtraction (DOS) to generate image with surface reflectance value. The next process was image masking process which was limited to only cover land area and masking area based on the range of the pixel value on water and land as well by utilizing a range value of transformation of NDVI.

The result of image on masking then was being extracted by pre field primary data. Data extraction was done through the

processing of digital image using spectral transformation with band ratio method. Band ratio method was used algorithm that has been generated by previous research on each parameter to be measured in the field which included transparency of Secchi Disk [12], that is:

$$\begin{aligned} \ln(\text{SDT}) &= B1/B3(1.3083) + B1(-0.0631) + (-1.2937) \\ \text{Chlorophyll-a with algorithm [13]} \\ \ln(\text{Chl}) &= 9.82(\text{TM1} - \text{TM3})/\text{TM2} \\ \text{and algorithm for total of phosphor [14]} \\ \ln(\text{TP}) &= -21.45(\text{TM3}/\text{TM2}) - 14.42(\text{TM1}/\text{TM3}) + 42.99(\text{TM1}) + 27.1 \end{aligned}$$

Where:

$\ln(\text{SDT})$  = Natural logarithm from the depth of *Secchi Disk* transparency

$\ln(\text{Chl})$  = Natural logarithm from chlorophyll-a

$\ln(\text{TP})$  = Natural logarithm from total of phosphor

TM1 = Value of reflectance band 1 Landsat-5 TM

TM2 = Value of reflectance band 2 Landsat-5 TM

TM3 = Value of reflectance band 3 Landsat-5 TM

The result of transformation of band ratio would then be overlaid in which the result was used as a tentative map of the research and was used to determine the distribution of the sample. The sample was taken by stratified random sampling method based on classification of overlay result.

#### 3.1 Analysis statistic and interpolation of data

Empirical modelling was done by using statistical analysis between the pixel value of image on a tentative map parameter and the result of field measurement. Statistical analysis was used linear regression analysis. The result of sampling in the field then tested in the laboratory for the parameters of total phosphor and chlorophyll-a. The result of parameter in the laboratory and in the field then was converted into natural logarithm. The distribution for each sample was tested using normality test of sampling then was correlated between sample score and pixel value by using SPSS software.

In the correlation which was done to chlorophyll-a, the most important predictor was on band 3 ( $r = 0,46$ ) compared to alarithmetic on initial estimation. Whereas for SDT, there were three predictors which have the same important correlation ( $r = 0,28$ ); however, predictor alarithmetic 2 was taken as an initial alarithmetic in spectral transformation to estimate the initial distribution. On phosphate, the most influential predictor was on band 2 ( $r = 0,48$ ) compared to alarithmetic on initial estimation.

Then, on the result of linear regression statistical analysis of data showed that there was a very low determinacy coefficient between the result of field data for chlorophyll-a and reflectance score in the green band of Landsat-8 OLI image. This was possible since water of reservoir has already mixed with the rainwater. In addition, it was also possible since the peak of productivity of micro algae in the water reservoir has already passed.

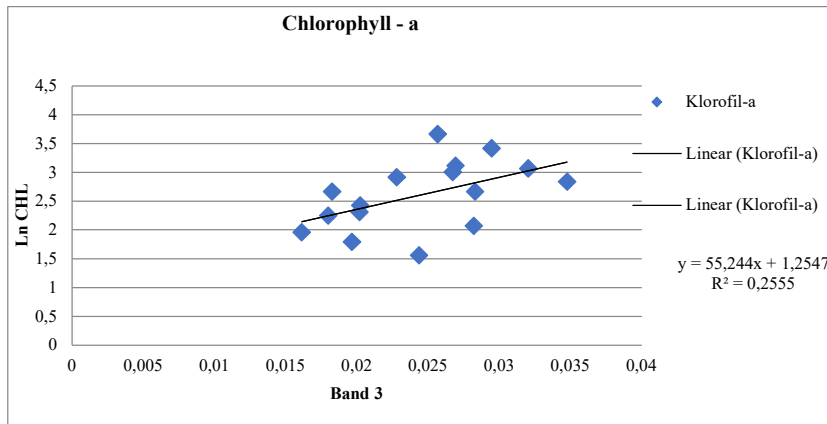


Figure 2 Graphic of linear regression of chlorophyll-a parameter

Regression and correlation on phosphate were compared to chlorophyll-a showed that there was an increasing of phosphorus content which was compared to a decreasing of chlorophyll-a. There was a possibility that the peak of micro algae productivity occurred between 14 August (image

recorded) and 15 September (field measurement) then the number of micro algae declined. This possibility was strengthening by the result of a high regression (R) and determinacy (R<sup>2</sup>).

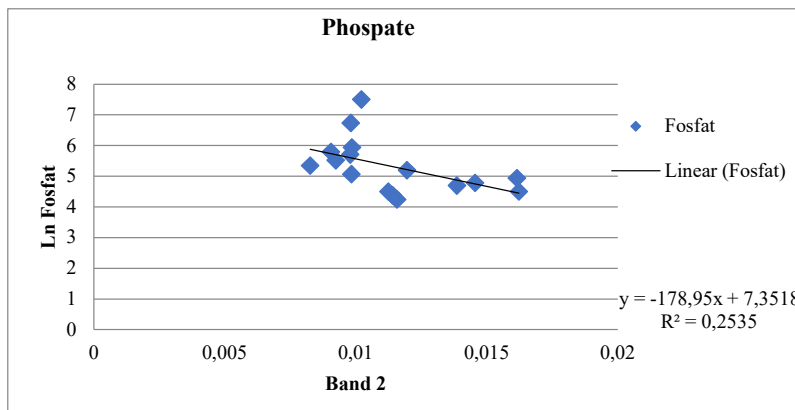


Figure 3 Graphic of linear regression of phosphate parameter

The result of regression on SDT as shown in figure 5 showed the level of translucency sun in water that relatively correlated with regression result and analysis of phosphate

regression. Therefore, there was a possibility of low connection of reservoir permeability and the level of total phosphor in the water of reservoir.

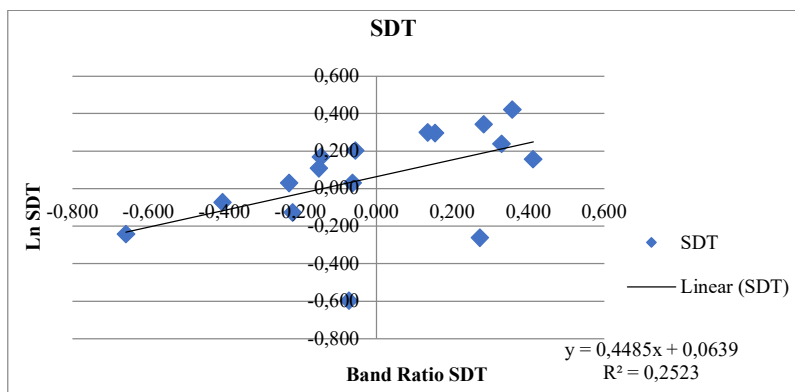


Figure 4 Graphic of linear regression of SDT parameter

4. RESULT AND DISCUSSION

4.1 Model of Classification Tree Analysis (CTA)

The parameters which would be used as input of CTA model for potential of HABs were chlorophyll-a, total of phosphor, and SDT, while the sample was data of chlorophyll-a. Chlorophyll-a was chosen as sample for model since chlorophyll-a represented the individual of micro algae. The result of transformation image based on the equality of the regression then was being an input for CTA modelling in IDRISI software. Data which would be the sample was classification of trophic status based on trophic index by Carlson [15]–[18]. The classes were 1-4 which 1 for light eutrophic level, 2 for medium eutrophic, 3 for heavy eutrophic, and 4 for hyper eutrophic. Modelling

was done by applying three kinds of algorithms; they are Entropy, Gini, and Ratio. These three algorithms then would be combined to level of pruning 1%-50%.

Based on the result of modelling, it is found that there are 150 rules from three kinds of algorithms and different kinds of pruning. The total of 150 rules is divided into 50 rules of Entropy method, 50 rules of Gini method, and 50 rules of Ratio method. The resulting rules is based on experiment of pruning on each percent. In the experiment of pruning, it is obtained a similar rule on certain range. From the result of the pruning, it is obtained the group range of pruning on each model. On resulted model, it is divided into entropy model with the pruning 1%-15%, 16%-23%, and 24%-50%. On gini alogarithm with the pruning 1%-7%, 8%-15%, 16%-33%, and 34%-50% whereas on ratio alogarithm is done on pruning 1%-33% and 34%-50%.

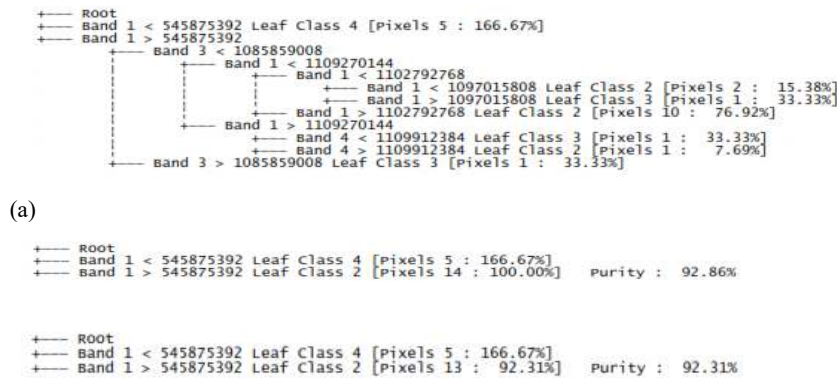


Figure 5 Rule with the longest tree on gini method with pruning 1%-7% (a) and rule with the shortest tree on gini and ratio methods with pruning 50% (b)

Overall, the tree on entropy alogarithm model is more influenced by score from bands 1, 2, and 3, whereas on gini alogarithm is more influenced by bands 1, 3, and 4. Ratio alogarithm is only influenced by 2 bands, they are bands 2 and 3. The depth of tree from the resulted model is also influenced by separability from training sample, such as on figure 5 when band 3 produces leaf for class 2 and 3.

4.2 Model Validation Test

The test of validation uses error matrix method which tests class from CTA modelling result and trophic status based on the result of field data that refers to the trophic index by Carlson [15]. Trophic class status by Carlson is taken on 4 bottom classes which potentially to emerge micro algae, they are light eutrophic, medium eutrophic, heavy eutrophic, and hyper eutrophic on the highest potential HABs. The best model is the model which has the highest validation compared to others, appropriate total classes for samples and or there is no class reduction as seen on table 1.

Table 1 The result of validation test of CTA classification result

Method	Validation Model (%)	Total Class
Entropy 1% - 15%	75	4
Entropy 16% - 23%	68,75	4
Entropy 24% - 50%	43,75	4
Gini 1% - 7%	68,75	4
Gini 8% - 15%	81,25	4
Gini 16% - 33%	81,25	4

Method	Validation Model (%)	Total Class
Gini 34% - 50%	100	3
Ratio 1% - 33%	75	4
Ratio 34% - 50%	100	3

The result of validation test on table 1 and graphic on figure 6 indicate that there is diverse of validation percentage. There are also some similarities on validation percentage whether on the same algorithm or different algorithm of pruning scale. Validation of that model produces classes based on the value of sample on CTA modelling. Resulted classes based on sample are 4 classes, but there are some algorithms with certain pruning are reduced in the number

of classes. As shown on table 1, it is seen that Gini method with pruning 34%-50% and Ratio 34%-50% have lower classes than model in another method. Even though they have the best model of validation, they cannot be used as the best model of validation since the resulted classes are reduced. Another quite high validation is shown on class of Gini method with pruning 8%-33% on 4 resulted classes.

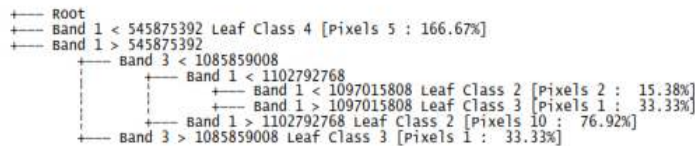


Figure 6 The best validation of tree on entropy model with pruning 10%

Through overall accuracy which is done on resulted models, then it is obtained that the highest validation is on model of Gini algorithms method with pruning 8%-15% and 81,25% accuracy, the tree of validation is shown on figure 6. The result of model on Gini method with pruning 16%-34% also produces the same score of validation, it is 81,25%, but the information of modelling result becomes more common since there is a bigger pruning which makes a lot of information combines into dominant class around it. Since the scale of score on each class is absolute, the reducing of the class makes the model is not accepted. The reducing of the class occurs because of the process of the majority to minority pixel generalization on another class. Since Gini and Ratio methods with pruning 34%-50% are not accepted, then the model with the highest accuracy and accepted total class is on model Gini method with pruning 8%-15%. Resulted tree on Gini method with pruning 8%-15% has 4 nodes with 4 leaf classes.

Through the elaboration of resulted tree, it is obtained that bathymetric data in band 1 has a great influence in the formation of spatial model for potential HABs besides chlorophyll-a as a parameter to the existence of micro algae in the waters. It is probably happened since the factor of the shallow and the depth of the waters in which will determine the level of dynamism of water movement on the surface of water or in the water column. The dynamic of water movement either on the surface or in the water column will make nutrients always be transported following the movement of water.

Based on the result of modelling which has done, it is showed on figure 7 that there is a distribution of water pollution with a high concentration of micro algae in some areas at the reservoir. Particularly in Tiwingan as the reservoir outlet, micro algae concentration keeps increasing especially in the area that intended to the land such as in Jungur and Belangian area; it is occurred since the water is relatively calm and is not too surging with shallow depth. In the depth of 23,3 m, it is found that it is possibly in emerging potential HAB's with medium class. Whereas the depth that potentially emerges quite high vulnerability of HAB's is below 13,9 m depth. Beside the factor of the depth, the factor of phosphate content as the food source of micro algae is also be the barrier at Riam Kanan reservoir. The score of 0,028 mg/L is being the barrier of phosphate content for medium and heavy vulnerable potential of HAB's. The reservoir area which has more phosphate content is categorized as a high potential area of the emergence of HAB's. According to Effendi [1] who states that phosphate degree is on 0,021-0,05 mg/L, then the result of the barrier score fits the waters fertility degree. The condition begins to support the development of micro algae and increased concentration of micro algae annually, in

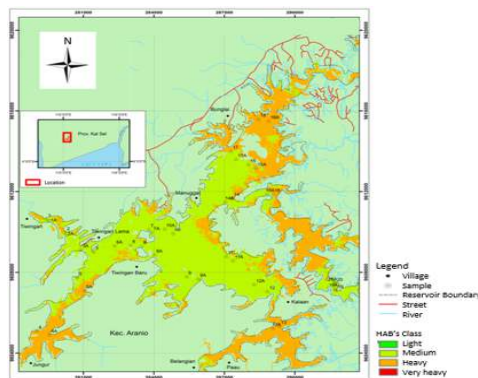


Figure 7 The potential distribution of HABs at Riam Kanan Reservoir

addition to the influence of the physiological condition of the reservoir.

The use of reservoir for agricultural also influences increased concentration of micro algae at the reservoir. Based on data of BPS in the year 2016, it is known that the total population who inhabits the settlement area of sub watersheds of Riam Kanan is 736.933 people, while the total population in Aranio sub district who inhabits the area around the reservoir is 8.899. Eko and Anong [19] in their research explained that there are 35 gr BOD content on every person's faces every day, 11,5 gr total Nitrogen (N) on every person everyday, and 0,8 gr total phosphate (P) on every person every day. Therefore, it can be assumed that from the total population in Aranio sub district it is found that there are 311,4 kg BOD content per day, 102,33 kg Nitrogen per day, and 7,2 kg phosphate per day in water. Those totals of content do not describe the content at the reservoir, but as the total potential content of the source of pollution to the reservoir.

Agricultural land also invests the increasing source of pollution at the reservoir. The increasing source of pollution is obtained from the use of agricultural fertilizer. Based on BPS data in 2008 [20], the use of fertilizers at Riam Kanan watershed in once harvest time were 150 kg/ha Urea, 75 kg/ha TSP, and 45% Nitrogen content and 20% phosphate in Urea. Then, if it was being totalled with 303.811,95 Ha width of agricultural dry land around the reservoir, the total potential of Nitrogen content was as much as 20.507.306,6 kg while the total potential of phosphate was 4.557.179,25 kg. The total of those contents has not reduced yet by the total contents which was absorbed by soil, plants, and dissolved by surface runoff. Nutrient elements that went into the reservoir were contained of residual fertilizer on the soil surface which was then dissolved by rainwater.

## 5. CONCLUSION

The result of modelling which has been done obtains 4 potential classes, they are light, medium, heavy, and extremely heavy classes which the distribution of HABs is more influenced by depth factor. Where in a high depth is dominated by medium class; whereas, in shallower depth with area of waters that stick out into the water is included in extremely heavy potential class.

The total population who is settled in Aranio sub district is 8.899, so the potential of reservoir pollution from human waste are consist of 311,4 kg per day BOD, 102,33 kg Nitrogen per day, and 7,2 kg phosphate per day; while, with 303.811,95 Ha width of dry land agriculture as of the total potential of Nitrogen content is as much as 20.507.306,6 kg and the total potential of phosphore is 4.557.179,25 kg. As a result, there is a significant relationship between the increasing level of reservoir fertility, human activity, and the changes in the surrounding area.

## REFERENCES

- [1] H. Effendi, *Telaah Kualitas Air: Bagi Pengelolaan Sumber Daya Lingkungan Perairan*. Jakarta: Kanisius, 2003.

- [2] S. S. Brahmana and Y. Summarriani, "Kualitas Air dan Eutrofikasi Waduk Riam Kanan di Kalimantan Selatan," pp. 24–46, 2010.
- [3] M. W. Matthews, S. Bernard, and K. Winter, "Remote sensing of cyanobacteria-dominant algal blooms and water quality parameters in Zeekoevlei, a small hypertrophic lake, using MERIS," *Remote Sens. Environ.*, vol. 114, no. 9, pp. 2070–2087, 2010, doi: 10.1016/j.rse.2010.04.013.
- [4] R. P. Stumpf and M. C. Tomlinson, "Remote sensing of harmful algal blooms," *Remote Sens. Digit. Image Process.*, vol. 7, pp. 277–296, 2005, doi: 10.1007/1-4020-3100-9\_12.
- [5] M. Kahru, O. P. Savchuk, and R. Elmgren, "Satellite measurements of cyanobacterial bloom frequency in the Baltic Sea: Interannual and spatial variability," *Mar. Ecol. Prog. Ser.*, vol. 343, pp. 15–23, 2007, doi: 10.3354/meps06943.
- [6] V. Barale, J. M. Jaquet, and M. Ndiaye, "Algal blooming patterns and anomalies in the Mediterranean Sea as derived from the SeaWiFS data set (1998-2003)," *Remote Sens. Environ.*, vol. 112, no. 8, pp. 3300–3313, 2008, doi: 10.1016/j.rse.2007.10.014.
- [7] R. L. Lawrence and A. Wright, "Rule-based classification systems using classification and regression tree (CART) analysis," *Photogramm. Eng. Remote Sensing*, vol. 67, no. 10, pp. 1137–1142, 2001.
- [8] M. Hansen, R. Dubayah, and R. Defries, "Classification trees: An alternative to traditional land cover classifiers," *Int. J. Remote Sens.*, vol. 17, no. 5, pp. 1075–1081, 1996, doi: 10.1080/01431169608949069.
- [9] BPPI, "Waduk Riam Kanan Menyumbang 30 MW untuk PT. PLN Persero," 2015. .
- [10] R. S. DeFries, M. Hansen, J. R. G. Townshend, and R. Sohlberg, "Global land cover classifications at 8 km spatial resolution: The use of training data derived from Landsat imagery in decision tree classifiers," *Int. J. Remote Sens.*, vol. 19, no. 16, pp. 3141–3168, 1998, doi: 10.1080/014311698214235.
- [11] U.S. Geological Survey, *LANDSAT 8 data users handbook version 1.0*. 2015.
- [12] S. M. Kloiber, P. L. Brezonik, L. G. Olmanson, and M. E. Bauer, "A procedure for regional lake water clarity assessment using Landsat multispectral data," *Remote Sens. Environ.*, vol. 82, no. 1, pp. 38–47, 2002, doi: 10.1016/S0034-4257(02)00022-6.
- [13] P. A. Brivio, C. Giardino, and E. Zilioli, "Determination of chlorophyll concentration changes in Lake Garda using an image-based radiative transfer code for Landsat TM images," *Int. J. Remote Sens.*, vol. 22, no. 2–3, 2001.

- [14] C. Wu *et al.*, “Empirical estimation of total phosphorus concentration in the mainstream of the Qiantang River in China using Landsat TM data,” *Int. J. Remote Sens.*, vol. 31, no. 10, pp. 2309–2324, 2010, doi: 10.1080/01431160902973873.
- [15] R. E. Carlson, “A trophic state index for lakes,” *Limnol. Oceanogr.*, vol. 22, no. 2, pp. 361–369, 1977, doi: 10.4319/lo.1977.22.2.0361.
- [16] A. M. Sheela, J. Letha, S. Joseph, K. K. Ramachandran, and S. P. Sanalkumar, “Trophic state index of a lake system using IRS (P6-LISS III) satellite imagery,” *Environ. Monit. Assess.*, vol. 177, no. 1–4, pp. 575–592, 2011.
- [17] A. Bekteshi and A. Cupi, “Use of trophic state index (Carlson, 1977) for assessment of trophic status of the shkodra lake,” *J. Environ. Prot. Ecol.*, vol. 15, no. 1, pp. 359–365, 2014.
- [18] A. G. Devi Prasad and P. Siddaraju, “Carlson’s Trophic State Index for the assessment of trophic status of two Lakes in Mandya district,” *Adv. Appl. Sci. Res.*, vol. 5, pp. 2992–2996, 2012.
- [19] I. W. Eko and S. Anong, “Karakteristik beban pencemaran limbah penduduk di Bandung dan Yogyakarta,” *Bull. Pus Air, Media Kegiat. Penelit. Keair.*, vol. 5, no. 21, pp. 15–35, 1996.
- [20] BPS, *Kabupaten Banjar Dalam Angka*. Kabupaten Banjar: Badan Pusat Statistik, 2008.