CLASSIFICATION OF BURNED PEATLAND USING PROBABILISTIC NEURAL NETWORK ALGORITHM BASED ON HIGH TEMPORAL DATA

Neneng Rachmalia Feta

Information Systems and Technology
Bank Rakyat Indonesia Institute of Technology and Business
http://bri-institute.ac.id/
nenengrachmaliafeta@gmail.com

Abstrak

Kebakaran lahan di Indonesia terjadi di lahan kering dan juga di lahan gambut. Kebakaran di lahan gambut lebih berbahaya dan lebih menantang untuk diatasi daripada kebakaran di lahan non-gambut, dan akibat kebakaran lahan gambut yang terjadi sangat merugikan masyarakat. Salah satu solusi yang ditawarkan dalam menilai kebakaran hutan dan lahan gambut adalah teknologi penginderaan jarak jauh. Citra satelit yang diperoleh dari teknologi penginderaan jarak jauh biasanya diklasifikasikan untuk analisis lebih lanjut. Tujuan utama dari penelitian ini adalah untuk mengembangkan model klasifikasi menggunakan Probabilistic Neural Network (PNN) untuk mengklasifikasikan area di lahan gambut sebelum, selama, dan setelah terbakar pada citra satelit Landsat 7 ETM+. Selanjutnya model tersebut digunakan untuk mendapatkan pola trajectory area yang terbakar menggunakan algoritma DBScan. Daerah penelitian adalah Kabupaten Ogan Komering Ilir Provinsi Sumatera Selatan, citra Landsat 7 ETM+ diambil dari Januari 2015 – Desember 2015.

Kata kunci: Lahan Gambut, Klasifikasi, PNN, DBScan, Landsat 7 ETM + citra

Abstract

Land fires in Indonesia occur on dry land as well as on peatlands. Fires on peatlands are more dangerous and more challenging to tackle than fires on non-peatlands, and the consequences of peatland fires that occur are very detrimental to communities. One of the solutions offered in assessing forest and peatland fires is remote sensing technology. Satellite images obtained from remote sensing technology are usually classified for further analysis. The main objective of this study is to develop a classification model using the Probabilistic Neural Network (PNN) to classify areas in peatlands before, during, and after burning on Landsat 7 ETM+ satellite imagery. Furthermore, the model is used to obtain the trajectory pattern of the burned area using the DBScan algorithm. The research area is Ogan Komering Ilir Regency; South Sumatra Province Landsat 7 ETM+ images were taken from January 2015 – December 2015.

Keywords: Peatland, Classification, PNN, DBScan, Landsat 7 ETM + imagery

INTRODUCTION

Peatlands in Indonesia are spread over several islands such as Sumatra, Kalimantan, and Papua. Peatland consists of many organic materials derived from the dead and decaying plant remains (W. W. International, 2015). The ability of peat to absorb water is relatively high; therefore, natural peatlands are not flammable. However, ecological balance can be disrupted by land conversion or canal making. Peatland conditions will dry to a certain depth in the dry season and cause flammable land. Peatland fires are much more challenging to handle than fires in the highlands as the flames spread below the surface (Adinugroho et al., 2005).

Global Forest Watch analysis shows that 75% of hot spot warnings occur in areas of peatlands that are mainly composed of decomposing organic materials (Sizer et al., 2014).

Meanwhile, in this decade, peatland fires have averaged 32.1% in Sumatra and 25.1% in Kalimantan (WWF, 2015). In September 2015, the Agency for Disaster Management (BPBD) of Dry Ogan Regency detected 234 hot spots spread over ten sub-districts (R. D, 2015). According to Greenpeace analysis (Greenpeace, 2014), the frequency of hotspots is five times as much on peatland compared to mineral soil (dryland).

Detecting peatland fires can be done by taking advantage of remote sensing technology. The

processing of satellite imagery generated from remote sensing is beneficial for relevant stakeholders in providing information on the spatial distribution of areas experiencing forest and land fires (burning areas), especially information on the area of burned land (Suwarsono et al., 2013). From the satellite, image data can be used for to classification process. Classification is one method in data mining used to create models that describe data classes and predict data classes in new data.

Antropov et al. (Antropov et al., 2014) evaluate the performance of Fully Polarimetric SAR (PolSAR) data in several land cover mapping studies in boreal forest environments, taking advantage of the high canopy penetration capability in the L-band. The research included mapping multiclass land cover, forest-non-forest delineation, and classification of soil types under vegetation using the Probabilistic Neural Network (PNN). The study results obtained an accuracy of up to 82.6% on land cover mapping of five classes. More than 90% of forest-non-forest mapping in wall-to-wall validation shows the suitability of PolSAR data for mapping large areas of land and forest cover. Boulila et al. (Boulila et al., 2010) researched a high-level approach for modeling Spatio-temporal knowledge from satellite images and proposed multi-approach segmentation involving several segmentation methods that help improve image modeling interpretation. The experiments made on LANDSAT scenes show that the approach outperforms classical methods in image segmentation and can predict Spatio-temporal changes of satellite images.

This research will classify burned peatland in Ogan Komering Ilir District using the PNN algorithm. Landsat 7 image satellite is used as data set in this research. Extracted image features as input for the classifier are image bands. Moreover, the result of the PNN classification model is used to get the trajectory pattern of a burned area using the DBScan algorithm.

RESEARCH METHODS

This research consists of four phases: image pra-processing, building an image classification model, trajectory pattern, and evaluation. Image pre-processing is done using Quantum GIS, and this phase eliminates all areas except peatland in Ogan Komering Ilir and determines class labels. The image classification model is built using PNN, trajectory pattern using DBScan algorithm, and the last phase is evaluation. Every phase in this research is shown in Figure 1.
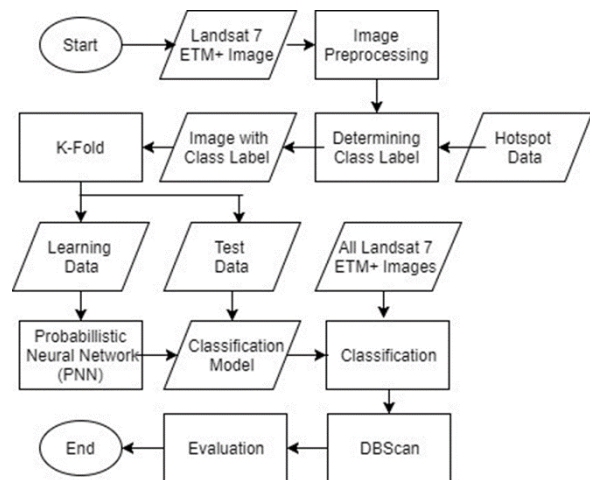

Figure 1. Research Methodology

Study Area

The study area in this research is peatland in Ogan Komering Ilir District, Province South Sumatera, which has 19,023.47 km². This district is located at $104^0.20'$ to $106^0.00'$ east longitude and $2^0.30'$ to $4^0.15'$ south latitude and geographically bordered by Ogan Ilir District, Banyuasin District and Palembang City (north), Lampung Province (south), Ogan Ilir District, and Ogan Komering Ulu Timur (west), Bangka Strait and Java Sea (east).

Data

Data sets used in this research are image data from satellite, peatland map, and hotspot data. Image data Ogan Komering Ilir District taken from http://earthexplorer.usgs.gov, this image data is Landsat 7 ETM+ with some path/row is 124/62 and has resolution 30x30, it is mean 1 pixel on the image has 30 meters in real. This image data was taken from Satelite in 2015.

Peatland map data was taken from http://data.globalforestwatch.org on 27 May 2017, and this map consists of Indonesian Peatland. This data is an SHP file representing peatland area in polygon and is used to clip Ogan Komering Ilir District Landsat Image. Hotspot data is used from range date 1 – 6 September 2015, and it is taken from FIRMS MODIS Fire/Hotspot, NASA/University of Maryland. Hotspot data determines class labels in the building classification model process.

Image Pre-processing

Landsat 7 ETM+ image data is pre-processed with the following processes. The first process is filling the gap in the image. The raw data Satellite image consists of 8 bands, and all image has a hole. This gap is caused by Landsat 7 satellite has

failure at Scan Line Corrector (SLN) sensor. This process is done by filling the image with a mask included in the downloaded data set file.

The second process is the compositing band. Compositing is a method for combining multiple rounds to 1 layer, and bands are used for compositing are band7 (red), band4 (green), and band2 (blue). The output of the compositing band is an image with RGB color, and each pixel has three digital numbers: red, green, and blue in range 0-255 for each.

The third process is clipping images with polygon peatland boundary. Area study focuses on peatland, area out of peatland eliminated with clipping process. Peatland Indonesian and Ogan Komering Ilir District Map are two map polygon data for clipping.

The fourth process is determining class labels. Class label is defined to the area in peatland manually, seen in display monitor and helped with additional data set: hotspot data. Hotspot data is used to determine which peatland location is burning, burned, and not burned. At the end of pre-processing, we collected a sample area with has 4 class labels: burning, burned, not burned, and cloud

K-Fold

K-Fold Cross Validation is a method for partitioning data. Data sets with a class label result in pre-processing data divided to be 2 data types, learning data, and test data. The number of K in this research is 10; input for the classification process is one partition as test data and nine partitions as learning data. Index of partition test data is incremental from 1 to 10, and every iteration of the others nine partitions learning information is all data except test data partition. This process is done using the Caret package in R.

PNN

PNN is Artificial Neural Network (ANN) that uses classical probability theorem (Bayes Classification); this classifier is supervised training (Specht, 1990). PNN is composed of 4 layers: the input layer, the pattern layer, the summation layers, and the output layer (Sa'adah & Pratiwi, 2020). This layer is shown in Figure 2.

The input layer is x object which consists of some k lengths features vector and classified to n class (Guan et al., 2021). Processes after input layer executed are:

Pattern Layer

The pattern layer uses one node for every learning data, and this node has number t data for each class. Every pattern node is dot-matrix multiplication from x- x_{ij}, then divide with bias σ, all

included in radial basis function $radbas = exp(−n)$. The combined formula for the pattern layer is:

$$f(x) = exp(-\frac{(x-x_{ij})^T - (x-x_{ij})}{2\sigma^2})$$ .... (1)

Summation Layer

This layer receives input from pattern layer nodes, and output is input for the output layer. The summation layer is summing t nodes for each class in the pattern layer.
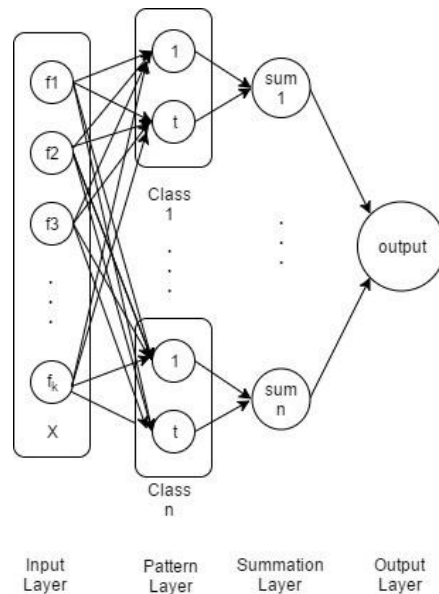

Figure 2. PNN Layers

$$p(x) = \frac{1}{2(\pi)^{\frac{k}{2}}\sigma^k t} \frac{1}{N_i} \sum_{i=1}^{t} exp(-\frac{(x-x_{ij})^T - (x-x_{ij})}{2\sigma^2}) .$$
.... (2)

Output Layer

This layer determines class from a given input. Input x classified as class Y, if $p_y(x)$ is more significant than other classes.

Evaluation Classification Model

The evaluation classification model is done by calculating the classification accuracy, and the classification model is tested to whole image data sets.

$$accuracy = \frac{\Sigma\ testing\ data\ correctly\ classified}{\Sigma\ testing\ data} \ x\ 100\%$$

DBScan

Density-Based Spatial Clustering of Applications with Noise (DBScan) is a clustering

method that groups dots based on data density in a region (Deng, 2020). The DBScan algorithm is designed to search for clusters and outliers in spatial data (Ester M, Kriegel H P, 1996). The DBScan algorithm looks for a central object that is an object that has a dense neighborhood. The DBScan algorithm connects the object center with its neighborhood to form 1 region as a cluster (Chen et al., 2021). The DBScan algorithm requires two input parameters: epsilon distance (Eps) and a minimum number of points (MinPts) (Ohadi et al., 2020). Epsilon is the distance between points indicating the object's density, while the MinPts is the minimum number of points in The center of a cluster object (Mustakim et al., 2021). The neighboring neighborhood that meets the epsilon distance is called the e-neighborhood. The following DBScan algorithm (Han et al., 2012):

1) Select a starting point (p) at random.
2) Get the e-neighborhood point from point p.
3) If the number of points from step 2 meets the MinPts value, the point p is the core point, forming a cluster.
4) If the point number of step 2 does not meet the MinPts value, then point p is the border point and select the next point.
5) Perform steps 2 through 4 until all the points have been processed and no dots can be added to the cluster.

RESULTS AND DISCUSSION

Image Pre-processing

SLC failure happened on 31 Mei 2003 and caused a gap in the image captured by Landsat 7. A gap in the image makes the data lose some information and can make low accuracy. Filling the gap is done using QGis; every image band is masked with file masks. NASA provides mask files. Figure 3 is shown an example image gap filling.
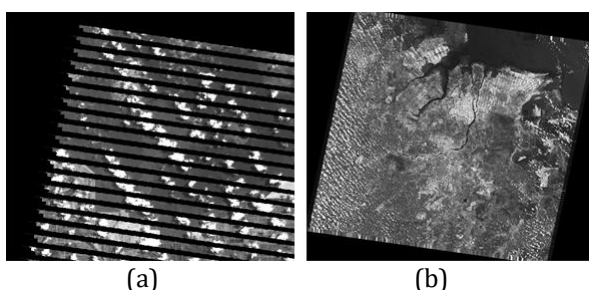

(a)   (b)

Figure 3. Pre-process (a) Satelite image band7 with a gap (b) Satelite image band7 after gap filled

Before compositing, all images are in brown; every pixel in the image has a one-digit number. In this process, image band7, band4, and band two are combined to become one layer. Image band7 is represented red color intensity, image band4 is represented green color intensity, and image band2 is represented blue color intensity. Every brand has a different digital number. Afterimage is composited, every pixel in the image consists of 3 digital numbers representing value for band7, band4, and band2. Visually after the composite is done, the image in brown color becomes an image RGB color.

Compositing image band is done using QGis, below is shown in Figure 4 process compositing image.
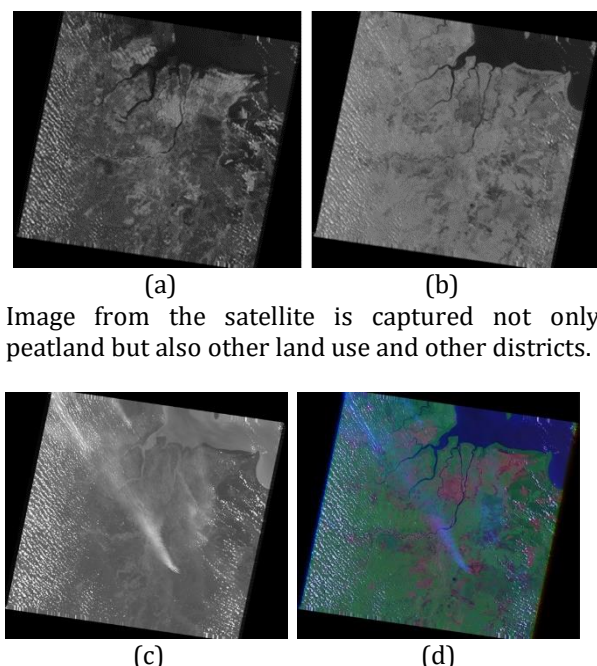

(a)   (b)

Image from the satellite is captured not only peatland but also other land use and other districts.

(c)   (d)

Figure 4. Process Compositing Band (a) Satelite image band7 before compositing (b) Satelite image band4 before compositing (c) Satelite image band2 before compositing (d) Composited Image

Overlay and clipping are used to extract peatland from other land use. The overlay shows peatland polygon on top image satellite, and clipping is used to take and cut peatland area only in the imaging satellite. This process is done using QGis and is shown in Figure 5.
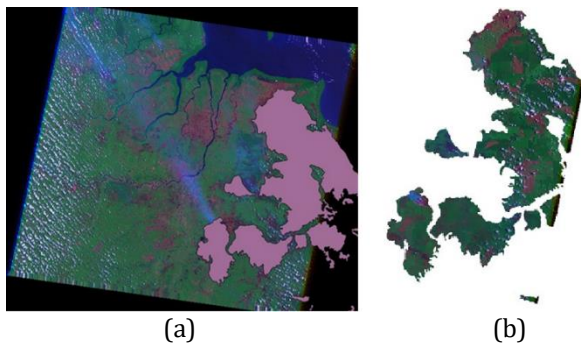
Figure 5. Extracting Peatland (a) Overlaying image satellite with peatland polygon (b) Peatland image satellite extracted

Image satellite still has no class label in this process, and supervised learning in the classification phase needs a class label. Therefore, the following process is determining class labels. To choose a class label, hotspot data is used to make a class label. Hotspot data represented in points and SHP format file. Hotspot data can be overlayed to a peatland image satellite. The class label used in this research is 4: burning, burned, not burned, and cloud. A sample is taken from image satellite data to represent each class; the example of each class is shown in Figure 6.
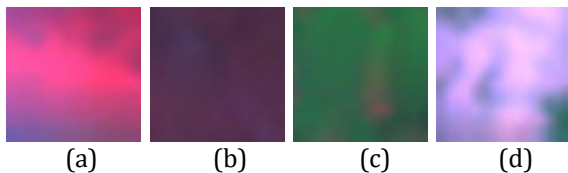

Figure 6. Example Landuse (a) Burning Class (b) Burned Class (c) Not Burned Class (d) Cloud Class

Build Classification Model

A digital number is extracted from the sample pixel. Using package Caret in R, these digital numbers are divided into learning and testing data using the K-Fold method. Data sets are divided into 10 group experiments with a proportion of 9/10 used as training data and 1/10 used as testing data. Every experiment iteration uses different learning and testing data; therefore, every iteration has different accuracy.

The classification was conducted using PNN with given data sets in each iteration. The smoothing parameters are used in this research are in the range 0.8 - 1. In a study conducted by (Liu, Wang, & Cheng, 2011), smoothing parameters can be adjusted based on learning data. Figure 7 below shows the result of each fold classification accuracies using PNN.
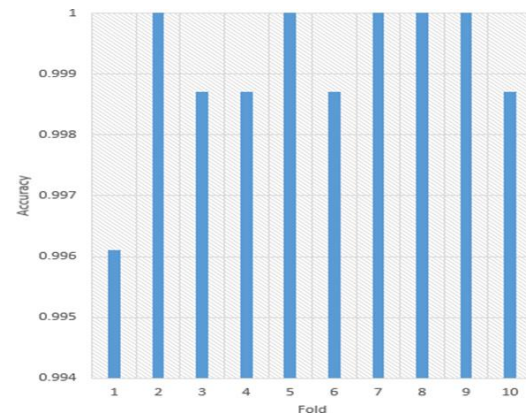

Figure 7. Result PNN Classification Accuracies

Figure 7 above shows the accuracy results in each fold. In some folds such as 2, 5, 7, 8, and 9, the accuracy reaches 100%, while the lowest is on fold one, which only reaches 0.996%. The overall accuracy of classification using the PNN algorithm results very well. PNN classification model is then used to classify 12 other Landsat images taken per month for one year in 2015. One of the classification results using the model made can be seen in Figure 8. The pixels included in the burned area are colored with a burning red, then the burnt area is dark red, the green color is applied to the unburned area, and the cloud area is given a white color.

Trajectory Pattern

After obtaining the classification model with PNN, the next step was to classify all Landsat images in 2015. The entire image is pre-processed before classification. One by one, imagery is classified at the pixel level. Once everything is successfully classified, the next step is to search the trajectory pattern of the burning land area. The algorithm used is DBScan.

In 2015 the Landsat image of the Ogan Komering Ilir area had a lot of fog.
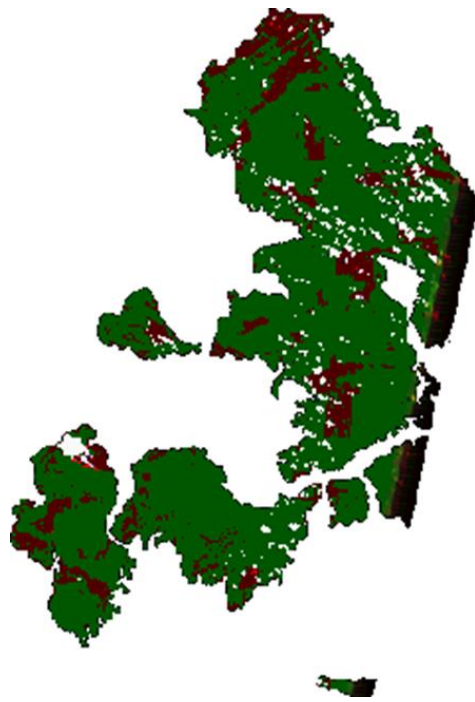
Figure 8. Result of Classification Landsat ETM + Image

The classification process on the whole image has been done, but the image data required DBScan algorithm is not sufficient to make a trajectory pattern. In Figure 9, we can see the result of the image preview.
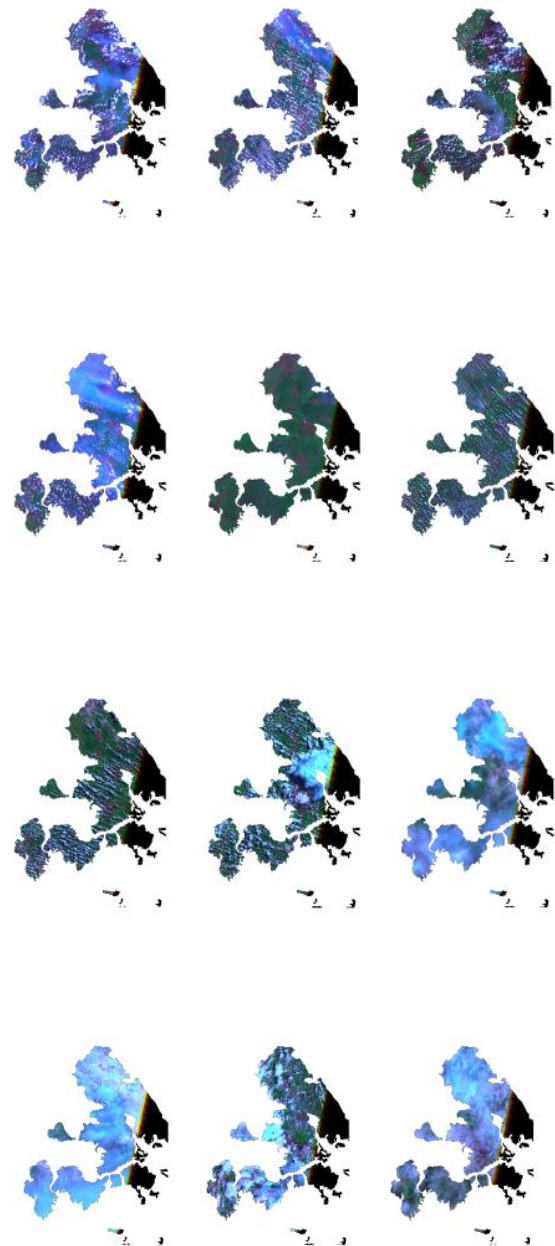

Figure 9. All Landsat ETM + Image in 2015 at Ogan Komering Ilir

CONCLUSIONS AND SUGGESTIONS

Conclusion

This research succeeded in implementing the PNN algorithm to classify the burning peatland area in Ogan Komering Ilir district, South Sumatra. The average accuracy of algorithms is 99%. The result of the image classification of the PNN algorithm shows a similarity of pixel between land after burning with the burning field. Because the

burnt land has a brownish-red color, and the red color is identical to the burning class.

The model generated by the PNN algorithm was successfully used to classify peatland areas by 2015. But the result is that many clouds block peat so that the cloud class dominates the area. This causes trajectory pattern on this data can not be done.

Suggestion

Suggestions that can later be made for further research include implementing another comparison algorithm besides the PNN and DBScan algorithms. Then use updated data and a more extended period, for example, more than one year, maybe two years or more, so that the research results are more representative.

REFERENCES

Adinugroho, wahyu catur, Suryadiputra, I. N. N., Saharjo, B. H., & Siboro, L. (2005). *Manual for the control of fire in peatlands and peatland forest.* (B. H. Saharjo (ed.); 1st ed.). Wetlands International – Indonesia Programme.

Antropov, O., Rauste, Y., Astola, H., Praks, J., Hame, T., & Hallikainen, M. T. (2014). Land cover and soil type mapping from spaceborne polsar data at l-band with probabilistic neural network. *IEEE Transactions on Geoscience and Remote Sensing*. https://doi.org/10.1109/TGRS.2013.2287712

Boulila, W., Farah, I. R., Ettabaa, K. S., Solaiman, B., & Ghézala, H. Ben. (2010). Spatio-temporal modeling for knowledge discovery in satellite image databases. *CORIA 2010: Actes de La COnference En Recherche d'Information et Applications - Proceedings of the Conference on Information Retrieval and Applications*.

Chen, X., Liu, D., Wang, X., Chen, Y., & Cheng, S. (2021). Improved DBSCAN Radar Signal Sorting Algorithm Based on Rough Set. *2021 2nd International Conference on Big Data and Informatization Education (ICBDIE)*, 398–401. https://doi.org/10.1109/ICBDIE52740.2021.00096

Deng, D. (2020). DBSCAN Clustering Algorithm Based on Density. *2020 7th International Forum on Electrical Engineering and Automation (IFEEA)*, 949–953. https://doi.org/10.1109/IFEEA51475.2020.00199

Ester M, Kriegel H P, S. J. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise, in: Proceedings of Second International Conference on Knowledge Discovery and Data Mining. *Kdd*.

Greenpeace. (2014). *Sumatra: will be covered with smoke*. Pers Greenpeace, Tech. Rep. http://www.greenpeace.org/seasia/id/PageFiles/616273/Kabut Asap Sumatera.pdf

Guan, S., Fang, Q., & Guan, T. (2021). Application of a Novel PNN Evaluation Algorithm to a Greenhouse Monitoring System. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–12. https://doi.org/10.1109/TIM.2021.3079558

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and TechniquesHan, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. San Francisco, CA, itd: Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-381479-1.00001-0. In *San Francisco, CA, itd: Morgan Kaufmann*.

Mustakim, Rahmi, E., Mundzir, M. R., Rizaldi, S. T., Okfalisa, & Maita, I. (2021). Comparison of DBSCAN and PCA-DBSCAN Algorithm for Grouping Earthquake Area. *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*, 1–5. https://doi.org/10.1109/ICOTEN52080.2021.9493497

Ohadi, N., Kamandi, A., Shabankhah, M., Fatemi, S. M., Hosseini, S. M., & Mahmoudi, A. (2020). SW-DBSCAN: A Grid-based DBSCAN Algorithm for Large Datasets. *2020 6th International Conference on Web Research (ICWR)*, 139–145. https://doi.org/10.1109/ICWR49608.2020.9122313

R. D. (2015). *Hotspot in Oki occurred 234 points*. Internet, Tech. Rep. http://daerah.sindonews.com/read/1044643/190/hotspotdi-oki-tercatat-234-titik-1442252742

Sa'adah, S., & Pratiwi, M. S. (2020). Classification of Customer Actions on Digital Money Transactions on PaySim Mobile Money Simulator using Probabilistic Neural Network (PNN) Algorithm. *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 677–681. https://doi.org/10.1109/ISRITI51436.2020.9315344

Sizer, N., Leach, A., Minnemeyer, S., Higgins, M., Stolle, F., Anderson, J., & Lawalata, J. (2014). *Preventing forest fires in Indonesia: focus on Riau Province, peatland, and illegal burning*. World Resources Institute.

Specht, D. F. (1990). Probabilistic neural networks. *Neural Networks*.

https://doi.org/10.1016/0893-6080(90)90049-Q

Suwarsono, Rokhmatuloh, & Waryono, T. (2013). Pengembangan Model Identifikasi Daerah Bekas Kebakaran Hutan Dan Lahan (Burned Area) Menggunakan Citra Modis Di Kalimantan. *Jurnal Penginderaan Jauh*.

W. W. International. (2015). *"Peatland,"*. Internet, Tech. Rep. http://www.wetlands.org/Whatarewetlands /Peatlands/tabid/2737/Default.aspx

WWF. (2015). *World wide fund for nature 2015 could the decline in hotspots is reached?* Internet, Tech. Rep. http://www.wwf.or.id/tentangwwf/upayaka mi/iklimdan_nergi/solusikami/