Register 6 (2) 2020 pp. 109-118
Jurnal Ilmiah Teknologi Sistem Informasi (Scientific Journal of Information System Technology)

ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)

Available online to www.journal.unipdu.ac.id

Unipdu

S2-Accredited – SK No. 34/E/KPT/2018
Journal page is available to www.journal.unipdu.ac.id:8080/index.php/register

Implementation DBSCAN algorithm to clustering satellite surface temperature data in Indonesia

Ariel Kristianto [a], Eko Sediyono [b], Kristoko Dwi Hartomo [c]

[a,b,c] Department of Information Systems, Universitas Kristen Satya Wacana, Salatiga, Indonesia

email: [a] kristiantoariel@gmail.com , [b] eko@uksw.edu , [c] kristoko@uksw.edu

ARTICLE INFO

Article history:
Received 5 April 2020
Revised 19 May 2020
Accepted 19 May 2020
Published 2 July 2020

Keywords:
cluster
DBSCAN
forest fire
Landsat-8
silhouette

IEEE style in citing this article:
A. Kristianto, E. Sediyono and K. D. Hartomo, "Implementation DBSCAN algorithm to clustering satellite surface temperature data in Indonesia," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 6, no. 2, pp. 109-118, 2020.

ABSTRACT

Forest and land fires are national and international problems. The frequency of fires in one of Indonesia's provinces, Riau, is a significant problem. Knowing where to repair the burn is essential to prevent more massive fires. Fires occur because of a fire triangle, namely fuel, oxygen, and heat. The third factor can be seen through remote sensing. Using the Landsat-8 satellite, named the Enhanced Vegetation Index (EVI) variable, Normalized Burn Area (NBR), Normal Difference Humidity Index (NDMI), Normal Difference Difference Vegetation Index (NDVI), Soil Adapted Vegetation Index (SAVI), and Soil Surface Temperature (LST). DBSCAN, as a grouping algorithm that can group the data into several groups based on data density. This is used because of the density of existing fire data, according to the character of this algorithm. The selected cluster is the best cluster that uses Silhouette Coefficients, eps, and minutes value extracted from each variable, so there is no noise in the resulting cluster. The result is more than 0, and the highest is the best cluster result. There are 5 clusters formed by clustering, each of which has its members. This cluster is formed enough to represent the real conditions, cluster which has a high LST value or has an NBR value. A high LST value indicates an increase in the area's temperature; a high NBR value indicates a fire has occurred in the area. The combination of LST and NBR values indicates the area has experienced forest and land fires. This study shows that DBSCAN clustered fire and surface temperature data following data from the Central Statistics Agency of Riau Province. Proven DBSCAN can cluster satellite imagery data in Riau province into several clusters that have a high incidence of land fires.

1. Introduction

The matter of concern on a national and international scale is forest and land fires. Refer to Minister of Forestry Regulation Number: P.12/Menhut-II/2009, forest, and land fires are one of the environmental problems that cause economic, ecological, and social issues [1]. When a fire occurs, it is the result of a chain reaction between a fire triangle, fuel or combustible material, oxygen, and heat [2].

Remote sensing technology can be used to observe forest and land fires, one of which is the Landsat 8 OLI satellite. Landsat 8 OLI is a continuation of the Landsat 7 project with several improvements [3]. Fire triangles that are a source of the fire can be detected by Landsat-8 satellites, fuels, or combustible materials and oxygen using EVI, NDVI, NDMI, SAVI, and heat using LST and NBR [4, 5, 6, 7, 8]. Fire can be anticipated if the location of the group that has the potential to burn is known. The clustering process can be carried out by the clustering method [1].

Algorithms that can be used in the clustering process vary widely, including algorithms that can be used to complete supervised learning such as Parallelepiped, Minimum Distance, Mahalanobis Distance, Maximum Likelihood, Naive Bayesian, k-Nearest Neighbor. Algorithms that can be used to

ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)

complete unsupervised clustering include linear regression, Isodata, k-Means, Improved Split and Merge Classification (ISMC), Adaptive Clustering, and DBSCAN [9, 10, 11]. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a good algorithm in grouping data. This algorithm clustered data based on data density (density-based clustering) in each cluster that is formed has a minimum amount of data. DBSCAN is able to work well and is not disturbed by existing noise and produces clusters with good structure [12, 13, 14]. DBSCAN is an algorithm that works well with density, very suitable with the high frequency of forest fires, which is the reason why Riau was chosen as the data to be examined, especially during the dry season, which is during the transition (transition) around August, September, and October.

The case study in this study was carried in one of the provinces in Indonesia, namely Riau. Riau Province was chosen because the region has a high rate of forest fires. National Disaster Management Authority states that the area of forest and land fires most affected is Riau. Riau Province is located at 02° 25′ LU-01° 15° S and 100° 03′-104° 00′ East, this province has an area of 87,023.66 Km^2, or around 8.7 million hectares in which 7.1 million hectares are forests and 3.9 million hectares are peatlands. According to the "SiPongi" system belonging to the Ministry of Environment and Forestry, Riau experienced an increase in the area of fires from 2017 to 2019. These fires can be prevented by knowing the clusters where potential fires can occur if the location of potential forest and land burning is known, firefighters can be alerted so that if a fire occurs, it can be prevented from spreading.

Therefore, in this study, a DBSCAN application is needed to request cluster locations that often occur in fires, the selection of the best cluster results using Silhouette Coefficient, the variable uses not only surface heat temperature but also other fire supporting factors such as the fire triangle. After the cluster data is obtained, compare the results of DBSCAN calculations with actual data, whether DBSCAN is able to group according to reality.

2. State of the Art

Forest fires that occur every year become a non-stop problem. Fires occur because of a chain reaction between three main causes, fuel or combustible material, oxygen, and heat, or the so-called triangle of fire **[2]**. These three materials can be identified using the Landsat-8 satellite by remote sensing. Landsat-8 was chosen because it has several improvements from its predecessor Landsat-7, research comparing the value of NDVI (Normalized Difference Vegetation Index) between Landsat 7 and Landsat 8 satellite imagery, gives the result that there is a positive and strong correlation, if the NDVI value in Landsat 7 rises, the NDVI value in Landsat 8 also rises, this study also provides results that the overall accuracy of Landsat 8 is better than Landsat 7. Dimitris revealed in his research that Landsat 8 has accurate results in its mapping, the same thing also expressed by Mancino, Landsat 8 OLI derived vegetation indices represent an effective tool to study the ecosystem variability, characterized by very high land use and land cover variability, due to the combination of both land management practices and physical characteristics mainly related to high climatic variability, also Acharya, Landsat-8 has improved its features as well as introduced new bands for more accurate studies in broader fields. It offers more combinations with a more narrow classification. Yet, more has to be explored about Landsat-8. Scientific research, as well as whole mankind, will be benefited from its improved imagery archives [3, 15, 16, 17].

NBR can be used as an aid in mapping burnt areas. According to Zubaidah et al. [18], in his research, the mapping of Landsat-8 shows that the burned area has 96% similarity with the burned area as a result of direct field observation. Deasy revealed in his research, the NBR and NDVI methods can be an alternative way to monitor forest burning. The pre-fire NDVI values were relatively higher than the post-fire NDVI values [5]. Thermal Infrared Sensor (TIRS) as one of the bands in Landsat-8 is able to map land surface temperature quite well, previous research states that TIRS is able to provide a good depiction of the surface temperature of the earth and can be used for various studies. A study suggests the same thing, the surface temperature of the earth (LST) can be described properly, also provides suggestions for researchers to combine it with other sensors such as NDVI [19, 20]. NDVI has a relationship with LST, the smaller the NDVI value, the greater the air temperature, meaning the less land cover or the less plants, the higher the value the more plants there that have the potential to become fuel when a forest fire occurs [21, 22]. NDMI is a method of measuring water content in a particular area canopy, NDMI is also sensitive to the level of humidity in vegetation, which is in areas prone to fire relative to the availability of fuel (fuel) fires [23, 24]. Gowri in his research said that SAVI and NDMI are suited for finding soil background information. The NDMI is the best index to identify the soil and moisture content [25]. Mancino said that EVI represents an improvement of NDVI, showing a reduced saturation in high vegetation cover regions, a reduction in atmospheric influences and a decoupling of the canopy background signal [16].

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a grouping method based on density-connected connections. DBSCAN has the advantage of having a simple interface for fast grouping algorithms and cluster extraction, expandable data structures, and methods for visualization and density-based grouping representations. DBSCAN has the ability to distinguish noise from large spatial databases. Each cluster has a minimum amount of data, objects that are not included in the cluster are considered as noise [12, 13, 14, 26]. According to Ahmed and Razak, DBSCAN is able to form clusters with various shapes and handle noise effectively [14]. Id [13] said that the DBSCAN algorithm showed that the resulting cluster had a good structure and was not sensitive to noise in the data. Hahsler et al. [26] said that DBSCAN offers a set of scalable, robust, and complete implementation of popular density-based clustering algorithms. The main features of DBSCAN are a simple interface to fast clustering and cluster extraction algorithms, extensible data structures and methods for both density-based clustering visualization and representation [26].

Not only using DBSCAN to find clusters, Silhouette Coefficient is used to find the best cluster and extract the data and match it with the data in the actual situation. Some of the previous studies that used separate variable, in this study, there were several variable combined to look for clusters that affected the cluster of forest and land fires. Based on this research, this research will use EVI, NBR, NDMI, NDVI, SAVI and LST variables to find the best cluster using DBSCAN in the case of fire in Riau province.

3. Method

The research method uses literature studies to find references in this research, the stages of research can be seen in Figure 1. The stages of research can be described in the following order according to Figure 1:

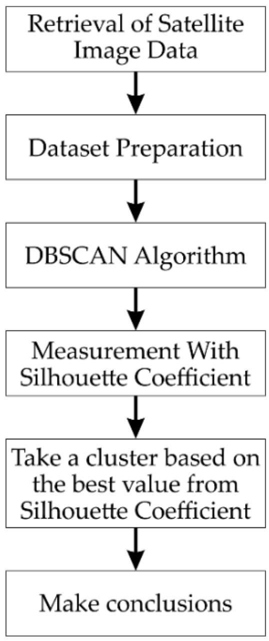

Figure 1. Research method

3.1. Retrieval of satellite image data

The data to be used comes from the United States NASA satellite, Landsat-8 satellite. The data used is satellite imagery data for 2018-2019. Data is downloaded through the site http://earthexplorer.usgs.gov. The data used in this study is Riau province data that includes 10 regencies and 2 cities, 166 districts. The Landsat-8 satellite image data was obtained from http://earthexplorer.usgs.gov, the data covered 4 scenes namely path = 126 row = 59, path = 126 row = 60, path = 127 row = 59, and path = 127 row = 60. Data is taken from 1 January 2018 to 22 March 2019.

ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)

3.2. Dataset preparation

The data obtained is processed by divided for each district and city, up to the district level in the Riau province. The data is then processed by getting several variables, namely EVI, NBR, NDMI, NDVI, SAVI, and LST. These variables are processed using the existing equation, and it will be extracted to get the mean value of each variable. Table 1 is a partial sample of data from the value extraction results for each EVI, NBR, NDMI, NDVI, SAVI, and LST index per district. NDVI values can be calculated using Equation 1 [6, 7, 8, 27, 28, 29], where NIR is band 5 and Red is band 4.

Table 1. Sample of data processing

District	EVI	NBR	NDMI	NDVI	SAVI	LST
BANTAN	-0.181806782	0.427912029	0.451602214	0.026698154	0.040046904	-73.49629466
BANGKO	-5.96797	0.502092	0.662728	0.441827	0.662728	-32.5536
BUKIT BATU	-0.709302201	0.249809566	0.18142461	0.140990243	0.153720474	-64.0509107
PINGGIR	-0.031999816	0.250635925	0.186388184	0.157160357	0.00815928	-129.3534074
KERINCI KANAN	-5.57286	0.457035	0.278652	0.471589	0.551399	-67.3431
MERBAU	-1.182526044	0.268172323	0.211594612	0.155906563	0.233856721	-60.66719004
PULAU MERBAU	-0.764301555	0.27424062	0.244900525	0.097507051	0.146258942	-69.05686308
TAMBUSAI	-10.0406	0.43399	0.243751	0.466471	0.672586	-31.5954
RANGSANG BARAT	-1.027765123	0.33570639	0.299184861	0.139185109	0.208774936	-61.11613237
RANGSANG PESISIR	-0.778433245	0.351302719	0.32874146	0.111513192	0.167267733	-61.94164145
TASIK PUTRI PUYU	-0.545406228	0.337500477	0.316778327	0.095504885	0.14325553	-64.20995192
MINAS	-0.00035	-0.11445	-0.17244	0.130257	0.000134	-140.259
TEBINGTINGGI TIMUR	-1.364496603	0.330661883	0.255321656	0.200313341	0.300464559	-42.66635836
TELUK MERANTI	-0.412715866	0.343775172	0.258137393	0.126444291	0.087393555	-52.75704682
BUNGARAYA	-1.131543637	0.35144546	0.253995627	0.138153511	0.207227822	-54.80725282
DAYUN	-0.329301918	0.357409481	0.289866179	0.098332038	0.070607311	-59.88124014
KOTO GASIB	-0.683622321	0.275469746	0.203915845	0.138437533	0.140528709	-59.23063119
MEMPURA	-1.08948309	0.322739698	0.246512636	0.154283601	0.231422487	-56.76043944
PUSAKO	-1.40527892	0.334646782	0.238199317	0.178172845	0.267255763	-52.89845723
SABAAUAH	-1.817030607	0.312816817	0.206756566	0.189295268	0.283939174	-49.28645468
SIAK	-0.815406778	0.330768146	0.249011521	0.118331793	0.177495643	-55.50714719
SUNGAI APIT	-1.244628673	0.305738435	0.231022546	0.164109493	0.246160902	-55.76091128
SUNGAI MANDAU	-0.985276541	0.301249639	0.218408769	0.165690006	0.22033377	-59.39461533
TUALANG	-0.211262818	0.286434545	0.221574087	0.138827616	0.047950258	-70.08900512
MINAS	-0.07059394	0.409889127	0.265488405	0.438163233	0.015699941	-140.2830193
BATANG TUAKA	-0.881884972	0.216642003	0.121476479	0.141809634	0.124451127	-51.33256896
ENOK	-0.526251554	0.207875892	0.126941747	0.177279393	0.073701041	-63.98883991
GAUNG	-1.001110659	0.267739247	0.155387316	0.16938661	0.215497117	-38.20402464
GAUNG ANAK SERKA	-0.946824463	0.252305363	0.149295297	0.183734718	0.17769604	-42.5575534
KEMPAS	-2.302771382	0.28735008	0.169133544	0.236117073	0.354170008	-34.10122661
KUALA KAMPAR	0.4682648	0.3047607	0.1874117	0.1727535	0.2591263	-36.623709
KEMUNING	-1.410990794	0.281359467	0.15500683	0.172200789	0.215601808	-35.05376274
LIMAPULUH	0.538354	0.193462	0.085759	0.179092	0.268632	-30.2319
KERITANG	-2.008083063	0.298979886	0.176007612	0.210537179	0.31338034	-37.60180176
MANDAH	-0.365959904	0.225990462	0.119103127	0.067878461	0.052607073	-52.19487191
PELANGIRAN	-0.593837886	0.231260944	0.12690937	0.070900612	0.08475444	-41.832089
PULAU BURUNG	-0.545892256	0.346579188	0.208050238	0.259803499	0.091499577	-99.51478996
RETEH	-0.034935365	0.246169962	0.159308308	0.210671604	0.005084458	-140.2690669
SUNGAI BATANG	-0.148493383	0.199045871	0.120913523	0.174372314	0.01801933	-124.8547778
TELUK BELENGKONG	-0.565268817	0.250241645	0.139471128	0.081167367	0.080786045	-55.51939492
TEMBILAHAN HULU	-0.999354813	0.204062588	0.120940319	0.145292941	0.131953485	-82.20011562

$$\mathbf{NDVI} = \left(\frac{\mathbf{NIR - Red}}{\mathbf{NIR + Red}}\right) \quad (1)$$

EVI is a method that has been developed to be able to measure the level of greenness and biomass of land or area. The EVI value can be calculated using Equation 2 *[7, 27, 28, 29, 30]*, where G is 2.5, a is 6, b is 7.5, L is 1, $\boldsymbol{\rho}$NIR is band 5, $\boldsymbol{\rho}$R is band 4 $\boldsymbol{\rho}$B is band 2.

$$\boldsymbol{EVI = G \times \frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + a \times \rho_R - b \times \rho_B + L}} \quad (2)$$

NBR can be used as an aid in mapping burnt areas; NBR values can be calculated using Equation 3 [29, 31, 32], where NIR is band 5 and SWIR 2 is band 7.

$$\boldsymbol{NBR = \left(\frac{NIR - SWIR\ 2}{NIR + SWIR\ 2}\right)} \quad (3)$$

SAVI is a development of the NDVI algorithm that uses the equation of vegetation with the same density and different soil background (vegetation isoline). The SAVI value can be calculated using Equation 4 [7, 8, 27, 29, 31], where ρNIR is reflectant value from Near-Infrared Band, ρRed is reflectant value from Red Band, L is correction of ground background enlightenment.

$$\boldsymbol{SAVI} = (\mathbf{1} + \boldsymbol{L}) \times \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red} + L} \quad (4)$$

NDMI is a method of measuring water content in a particular area canopy, NDMI is also sensitive to the level of humidity in vegetation, which is in areas prone to fire relative to the availability of fuel (fuel) fires [23, 24]. NDMI values can be calculated using Equation 5 [24, 25, 33, 34], where Red is band 4 and NIR is band 5.

$$\boldsymbol{NDMI} = \left(\frac{Red - NIR}{Red + NIR}\right) \quad (5)$$

The value for each EVI, NBR, NDMI, NDVI, SAVI, and LST can be explained:

- The value range for EVI is -1 to 1, and for healthy vegetation, it varies between 0.2 and 0.8.
- The value range for NBR can be divided like.
  - NBR = < -0.25 represent High post-fire regrowth
  - NBR = -0.25 to -0.1 represent Low post-fire regrowth
  - NBR = -0.1 to +0.1 represent Unburned
  - NBR = 0.1 to 0.27 represent Low-severity burn
  - NBR = 0.27 to 0.44 represent Moderate-low severity burn
  - NBR = 0.44 to 0.66 represent Moderate-high severity burn
  - NBR = > 0.66 represent High-severity burn
- The value range for NDMI can be divided like:
  - NDMI <=1 represent barren rock, sand or snow.
  - NDMI 0.2 to 0.5 represent shrubs or agriculture.
  - NDMI 0.6 to 0.9 represent Dense vegetation (Forest), to get the best silhouette value.
- The value range for SAVI is from -1 to 1. The lower the value, the lower the amount/cover of green vegetation.
- The value range for NDVI can be divided like:
  - NDVI = -1 to 0 represent Water bodies.
  - NDVI = -0.1 to 0.1 represent Barren rocks, sand, or snow.
  - NDVI = 0.2 to 0.5 represent Shrubs and grasslands or senescing crops.
  - NDVI = 0.6 to 1.0 represent Dense vegetation or tropical rainforest.
- The range of values for LST varies from positive to negative, the higher the value, the warmer the surface temperature of the soil.

3.3. DBSCAN algorithm

Processing data using the DBSCAN algorithm, the data is processed to produce clusters that indicate a particular group. DBSCAN has the ability to distinguish noise from large spatial databases. Each cluster has a minimum amount of data; objects that are not included in the cluster are considered as noise. Computing (DBSCAN) can be described as follows [12, 13, 14, 26]:

1. Initialize parameters minpts (minimum points), eps (radius).
2. Specify the starting point or p randomly.
3. Calculate the distance of a point that is density reachable to p using Equation 6, where x is first point and y is second point.
   $$\boldsymbol{E}(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=0}^{n} (\boldsymbol{X}_i - \boldsymbol{Y}_i)^2} \quad (6)$$
4. If the point formed in the eps radius is more than the same as minpts, then the p point is the core point, and the cluster is formed .
5. If p is a border point and there is no point that is density reachable to p, then the process continues to another point.
6. Repeat steps 3-5 until all points have been processed.

3.4. Measurement with silhouette coefficient

Measurement using Silhouette Coefficient will show the best parameter in showing the most optimal cluster, what is the value for "minpts" and what is the value for "eps". Silhouette Coefficient is a combination of two methods: a method that functions to measure how close the relationship between objects in a cluster is called cohesion, and a method that functions to measure how far a cluster is separated from another cluster called separation. The stages in calculating the silhouette coefficient [1].

ISSN 2502-3357 (Online)
A. Kristianto et al./Register 6 (2) 2020 pp. 109-118
ISSN 2503-0477 (Print)

1. Calculate the average distance of objects with all other objects in a cluster with Equation 7, where j is another cluster in one cluster *A* dan $\boldsymbol{d(i,j)}$ is the distance between cluster *i* with *j*.

$$\boldsymbol{a(i)} = \frac{\mathbf{1}}{[\boldsymbol{A}]-\mathbf{1}} \sum \boldsymbol{j} \in_{\boldsymbol{A},\boldsymbol{j}\neq\boldsymbol{i}} \ \boldsymbol{d(i,j)} \tag{7}$$

2. Calculate the average distance of an object with all other objects in another cluster, then take the minimum value with Equation 8, where $\boldsymbol{d(i,C)}$ is average distance of cluster *i* to all objects in another cluster *C* where $\boldsymbol{A} \neq \boldsymbol{C}$.

$$\boldsymbol{d(i,C)} = \frac{\mathbf{1}}{[\boldsymbol{A}]} \sum \boldsymbol{j} \in \boldsymbol{C}\ \boldsymbol{d(i,j)} \tag{8}$$

3. Calculate the Silhouette Coefficient value with the Equation 9 [35], where $\boldsymbol{a(i)}$ is the average distance between $\boldsymbol{xi}$ and each data point in the same cluster and $\boldsymbol{b(i)}$ is minimum average distance between $\boldsymbol{xi}$ and each data point in another cluster.

$$\boldsymbol{s(i)} = \frac{\boldsymbol{b(i)-a(i)}}{\boldsymbol{max}\ (\boldsymbol{a(i)},\boldsymbol{b(i)})} \tag{9}$$

4. The value of the silhouette coefficient results in the range of values -1 to 1, the better the grouping of data in one cluster, then the results are close to value 1. Conversely, if the worse the grouping of data in one cluster, the results are close to the value of -1.
5. Take a cluster based on the best value from Silhouette Coefficient: After knowing the optimal value of "minpts" and "eps", a cluster can be taken to compare it to the actual situation.
6. Draw the conclusions: After getting the results, it can be concluded whether DBSCAN can do optimal clustering.

4. Experiment and Results

The data that has been obtained is processed using R, and data preparation need to be done before processing further data using the clustering algorithm. The preparation includes separating the data of each object variable, eliminating missing data, then processing the data of each variable with the DBSCAN algorithm. After the data is processed, the silhouette coefficient is used to measure the success of each cluster, and the data will be obtained, as shown in Figure 2.

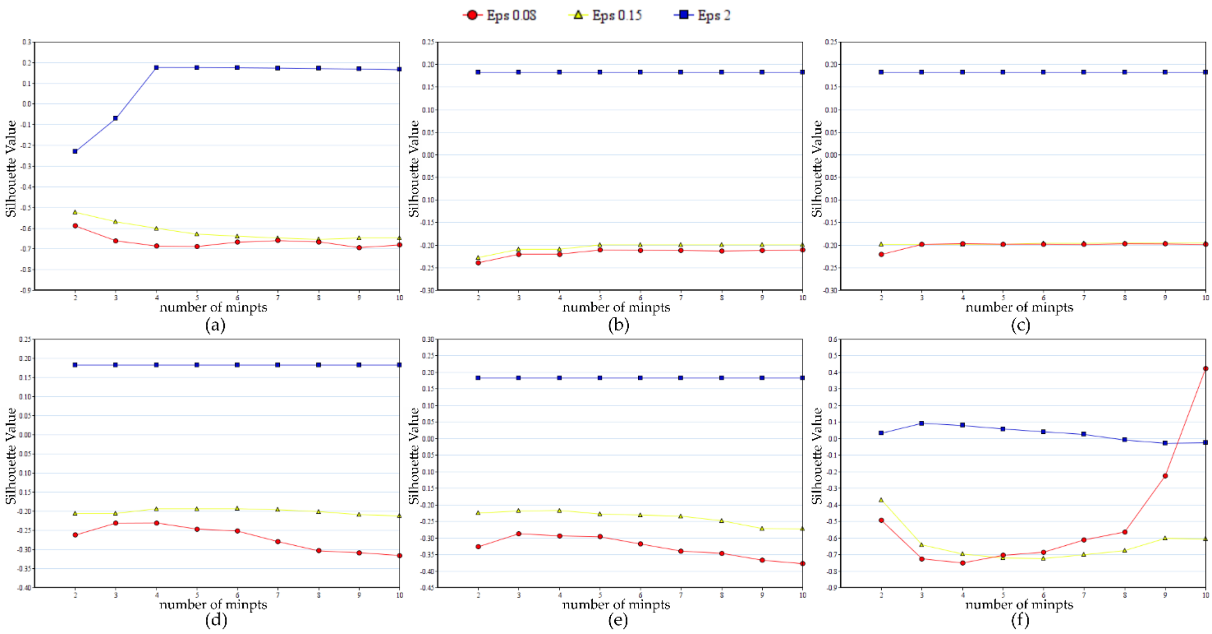

Figure 2. DBSCAN Silhouette Value for: (a) EVI; (b) NBR; (c) NDMI; (d) NDVI; (e) SAVI, and (f) LST

Variables are individualized to find uniform eps and minpts values, where all variables get positive values to get results that are in accordance with reality, so there is no data that can be biased or noise that can interfere with the clustering process. Each variable shows different results, which are calculated in Figure 2, are some interesting Silhouette Coefficient values for assessment, in DBSCAN, the values "eps" and "minpts" are very useful in the grouping, different values will produce different Silhouette Coefficients. Silhouette Coefficient values look very volatile, from negative to positive or positive to negative, from this value can be seen as the most optimal cluster to be taken as a result of clustering. This study tries to test the eps values from 0.01 up to the maximum value, with increments

ISSN 2502-3357 (Online)
A. Kristianto et al./Register 6 (2) 2020 pp. 109-118 ISSN 2503-0477 (Print)

of 0.01 for each test. This study also tested the minpts values ranging from 2 to the maximum value.

EVI results are shown in Figure 2, high average results if the value of eps = 2, at eps = 2, the highest value is owned when minpts = 4 to minpts = 10, which indicates a better cluster. Likewise, for NBR, eps = 2 dominates with a positive value, the same as other variables such as NDMI, NDVI, and SAVI. LST has different things when eps = 0.08 has the highest result when minpts = 10, although other minpts have negative results, While eps = 2 has an average positive result, the average positive value obtained by eps = 2 forms a good cluster, so it is used for cluster formation of research results.

After getting the Silhouette Coefficient value, the research continues on the process of clustering, clustering using DBSCAN using the maximum eps and minpts values of the Silhouette Coefficient, which will get an optimal cluster. Clustering using variables EVI, NBR, NDMI, NDVI, SAVI, and LST produces results that indicate several areas with the potential for forest fires, the cluster is shown in Table 2, data from Landsat-8 will be processed and clustered with DBSCAN will produce several clusters as follows. Cluster data has members like Table 2.

Table 2. The cluster formed by DBSCAN

Cluster	Region/City
0	Rokan Hilir, Pelalawan, Kampar, Rokan Hulu, Bengkalis, Indragiri Hulu, Indragiri Hilir, Kuantan Singingi, Pekanbaru, Siak, Dumai, Kepulauan Meranti
1	Kepulauan Meranti, Kampar, Siak, Rokan Hilir, Bengkalis
2	Rokan Hulu, Indragiri Hulu
3	Kuantan Singingi , Indragiri Hilir, Pelalawan
4	Pekanbaru, Dumai

There are 5 clusters formed by clustering, each of which has its own members. This cluster is formed enough to represent the real conditions. Cluster 0 has a high LST value or has an NBR value or has a burnt area, cluster 1 is a cluster that has a high NBR value or a high burnt area, cluster 2 has a medium NBR value or a medium burnt area but has an LST value that is high or has a high surface temperature, cluster 3 has the same character as cluster 2 only has a higher LST value, and cluster 4 has a lower LST value than the other clusters. A high LST value indicates that an increase in the temperature of the area, a high NBR value indicates a fire has occurred in the area. The combination of LST and NBR values indicates the area has experienced forest and land fires.

Table 3. Data from the Riau Province Central statistics agency on forest and land fires

Regency/City	Forest & Land Fire		
	Hotspot	Area Burnt	Number of Events
Pekanbaru	0	12.7	6
Dumai	395	122.75	25
Kepulauan Meranti	515	236.11	19
Kampar	548	83.25	35
Siak	682	76.5	9
Indragiri Hilir	1378	82	5
Bengkalis	1826	64	11
Rokan Hulu	1869	68	10
Indragiri Hulu	1919	45.3	8
Kuantan Singingi	2170	24.5	3
Rokan Hilir	3198	392	16
Pelalawan	3296	162.16	41

According to data released by the Central Statistics Agency of Riau Province in Table 3, all regions in Riau province have experienced forest and land fires. This study shows that DBSCAN clustered fire and surface temperature data in accordance with reality, as shown in Table 2. The results of this study are in accordance with the Central Statistics Agency data from Riau Province, cluster 0, which has a high LST, and NBR values have a burning experience data as in Table 3. Cluster 4, which has a low LST value, Table 3 also explains that the cluster has the lowest number of hotspots. This shows that the DBSCAN clustering research on forest and land fire data was successful and provided appropriate data.

5. Conclusions

This research proves that DBSCAN has a good ability in clustering, evidenced by the results of groupings that are in accordance with reality, where districts/cities grouped in groups are areas that have experienced forest and land fires. To determine the best cluster used, Silhouette Coefficient, this will tell the optimal cluster when the value is close to 1. As in cluster 0, Rokan Hilir, Pelalawan, Kampar,

ISSN 2502-3357 (Online)
A. Kristianto et al./Register 6 (2) 2020 pp. 109-118
ISSN 2503-0477 (Print)

Rokan Hulu, Bengkalis, Indragiri Hulu, Indragiri Hilir, Kuantan Singingi, Pekanbaru, Siak, Dumai, Meranti Islands, are areas that have experienced forest and land fires. In cluster 3, Kuantan Singingi, Indragiri Hilir, Pelalawan are areas that have high surface temperatures, and many hotspots are found in the area. The DBSCAN algorithm has a clustering result that is very dependent on the eps or radius values, DBSCAN also depends on the minimum point in the grouping process, so an assessment is needed to find the best cluster, one of the ways is with Silhouette Coefficient. Recommendations for further research, for data that use high density (density), it is better to choose DBSCAN, because data that has distortion will be considered as noise so that it does not improve clustering performance. Suggestions for future researchers can use data that has a variety of characters, such as maximum values, minimum values, standard deviations that are more diverse so that it looks different for each grouping for the DBSCAN algorithm.

6. References

[1] D. F. Pramesti, M. T. Furqon and C. Dewi, "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* vol. 1, no. 9, pp. 723-732, 2017.

[2] H. K. Muchtar, H. Ibrahim and S. Raodhah, "Analisis Efisiensi Dan Efektivitas Penerapan Fire Safety Management Dalam Upaya Pencegahan Kebakaran di PT. Consolidaetd Electric Power Asia (Cepa) Kabupaten Wajo," *Higiene: Jurnal Kesehatan Lingkungan,* vol. 2, no. 2, pp. 91-98, 2016.

[3] D. R. Amliana, Y. Prasetyo and A. Sukmono, "Analisis Perbandingan Nilai Ndvi Landsat7 dan Landsat8 Padakelas Tutupan Lahan (Studi Kasus: Kota Semarang, Jawa tengah)," *Jurnal Geodesi Undip,* vol. 5, no. 1, pp. 264-274, 2016.

[4] M. T. Mustafa, K. I. Hassoon, H. M. Hussain and M. H. Abd, "Using Water Indices (NDWI, MNDWI, NDMI, WRI and AWEI) to Detect Physical and Chemical Parameters by Apply Remote Sensing and Gis Techniques," *International Journal of Research - Granthaalayah,* vol. 5, no. 10, pp. 117-128, 2017.

[5] D. Arisanty, S. Adyatma, M. Muhaimin and A. Nursaputra, "Landsat 8 OLI TIRS Imagery Ability for Monitoring Post Forest Fire Changes," *Pertanika J. Sci. & Technol,* vol. 27, no. 3, pp. 1105-1120, 2019.

[6] M. A. Ganie and A. Nusrath, "Determining the Vegetation Indices (NDVI) from Landsat 8 Satellite Data," *International Journal of Advanced Research (IJAR),* vol. 4, no. 8, pp. 1459-1463, 2016.

[7] E. K. Andana, "Pengembangan Data Citra Satelit Landsat-8 untuk Pemetaan Area Tanaman Hortikultura dengan Berbagai Metode Algoritma Indeks Vegetasi (Studi Kasus: Kabupaten Malang dan Sekitarnya)," in *Prosiding Seminar Nasional Manajemen Teknologi XXI*, Surabaya, 2015.

[8] S. M. C. Nogueira, M. A. Moreira and M. M. L. Volpato, "Relationship Between Coffee Crop Productivity and Vegetation Indexes Derived from Oli/Landsat-8 Sensor Data with and without Topographic Correction," *Engenharia Agrícola,* vol. 38, no. 3, pp. 387-394, 2018.

[9] N. R. Apriyanti, R. A. Nugroho and O. Soesanto, "Algoritma k-Means Clustering dalam Pengolahan Citra Digital Landsat," *KLIK (Kumpulan jurnaL Ilmu Komputer),* vol. 2, no. 2, 2015.

[10] P. P. Adikara, R. C. Wihandika, F. Utaminingrum, Y. A. Sari, M. A. Fauzi, D. Syauqy and R. Maulana, "Regresi linier berbasis clustering untuk deteksi dan estimasi halangan pada smart wheelchair," *Register: Jurnal Ilmiah Teknologi Sistem Informasi,* vol. 3, no. 1, pp. 11-16, 2017.

[11] M. Mustaqim, B. Warsito and B. Surarso, "Kombinasi Synthetic Minority Oversampling Technique (SMOTE) dan Neural Network Backpropagation untuk menangani data tidak seimbang pada prediksi pemakaian alat kontrasepsi implan," *Register: Jurnal Ilmiah Teknologi Sistem Informasi,* vol. 5, no. 2, pp. 116-127, 2019.

[12] N. M. A. S. Devi, I. K. G. D. Putra and I. M. Sukarsa, "Implementasi Metode Clustering DBSCAN pada Proses Pengambilan Keputusan," *Lontar Komputer,* vol. 6, no. 3, pp. 185-191, 2015.

[13] I. D. Id, "Modifikasi DBSCAN (Density-Based Spatial Clustering with Noise) Pada Objek 3 Dimensi," *Jurnal Komputer Terapan,* vol. 3, no. 1, pp. 41-52, 2017.

[14] K. N. Ahmed and T. A. Razak, "An Overview of Various Improvements of DBSCAN Algorithm in Clustering Spatial Databases," *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 5, no. 2, pp. 360-363, 2016.

[15] D. Poursanidis, N. Chrysoulakis and Z. Mitraka, "Landsat 8 vs. Landsat 5: A comparison based on urban and peri-urban land cover mapping," *International Journal of Applied Earth Observation and Geoinformation,* vol. 35, pp. 259-269, 2015.

[16] G. Mancino, A. Ferrara, A. Padula and A. Nolè, "Cross-Comparison between Landsat 8 (OLI) and Landsat 7 (ETM+) Derived Vegetation Indices in a Mediterranean Environment," *Remote Sens,* vol. 12, no. 2, p. 291, 2020.

[17] T. D. Acharya and I. Yang, "Exploring Landsat 8," *International Journal of IT, Engineering and Applied Sciences Research (IJIEASR),* vol. 4, no. 4, pp. 4-10, 2015.

[18] A. Zubaidah, S. Sulma, Suwarsono, Y. Vetrita, M. Priyatna and K. A. D. Sukowati, "Akurasi Luas Areal Kebakarandari Data Landsat-8 Oli di Wilayah Kalimantan," *Majalah Ilmiah Globë,* vol. 19, no. 1, pp. 21-32, 2017.

[19] U. Avdan and G. Jovanovska, "Algorithm for Automated Mapping of Land Surface Temperature Using LANDSAT 8 Satellite Data," *Journal of Sensors,* vol. 2016, 2016.

[20] D. Jeevalakshmi, S. N. Reddy and B. Manikiam, "Land Surface Temperature Retrieval from LANDSAT data using Emissivity Estimation," *International Journal of Applied Engineering Research,* vol. 12, no. 20, pp. 9679-9687, 2017.

[21] D. K. Sunaryo and M. Z. Iqmi, "Pemanfaatan Citra Landsat 8 untuk Pendeteksian dan Mengetahui Hubungan Kerapatan Vegetasi Terhadap Suhu Permukaan (Studi Kasus: Kota Bandar Lampung, Provinsi Lampung)," *Spectra,* vol. 25, no. 13, pp. 55-72, 2015.

[22] D. P. Wiguna, "Identifikasi Suhu Permukaan Tanah dengan Metode Konversi Digital Number Menggunakan Teknik Penginderaan Jauh dan Sistem Informasi Geografi," *Jurnal Teknologi Informasi dan Komunikasi,* vol. 6, no. 2, pp. 59 - 69, 2017.

[23] S. Rahman and V. Mesev, "Change Vector Analysis, Tasseled Cap, and NDVI-NDMI for Measuring Land Use/Cover Changes Caused by a Sudden Short-Term Severe Drought: 2011 Texas Event," *Remote Sensing,* vol. 11, no. 19, 2019.

[24] T. Long, Z. Zhang, G. He, W. Jiao, C. Tang, B. Wu, X. Zhang, G. Wang and R. Yin, "30 m Resolution Global Annual Burned Area Mapping Based on Landsat Images and Google Earth Engine," *Remote Sensing,* vol. 11, no. 5, 2019.

[25] L. Gowri and K. R. Manjula, "Evaluation of Various Vegetation Indices for Multispectral Satellite Images," *International Journal of Innovative Technology and Exploring Engineering (IJITEE),* vol. 8, no. 10, pp. 3494-3500, 2019.

[26] M. Hahsler, M. Piekenbrock and D. Doran, "dbscan: Fast Density-Based Clustering with R," *Journal of Statistical Software,* vol. 91, no. 1, 2019.

[27] X. She, L. Zhang, Y. Cen, T. Wu, C. Huang and M. H. A. Baig, "Comparison of the Continuity of Vegetation Indices Derived from Landsat 8 OLI and Landsat 7 ETM+ Data among Different Vegetation Types," *Remote Sens,* vol. 7, pp. 13485-13506, 2015.

[28] M. Arekhi, C. Goksel, F. B. Sanli and G. Senel, "Comparative Evaluation of the Spectral and Spatial Consistency of Sentinel-2 and Landsat-8 OLI Data for Igneada Longos Forest," *ISPRS Int. J. Geo-Inf,* vol. 8, no. 2, 2019.

[29] Z. Wu, B. Middleton, R. Hetzler, J. Vogel and D. Dye, "Vegetation Burn Severity Mapping Using Landsat-8 and WorldView-2," *Photogrammetric Engineering & Remote Sensing,* vol. 81, no. 2, p. 143–154, 2015.

A. Kristianto et al./Register 6 (2) 2020 pp. 109-118
ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)

[30] U. C. Nugroho and D. D. Domiri, "Identification of Land Surface Temperature Distribution of Geothermal Area in Ungaran Mount by using Landsat 8 Imagery," *International Journal of Remote Sensing and Earth Sciences,* vol. 12, no. 2, p. 143–150, 2015.

[31] S. Wang, M. H. A. Baig, S. Liu, H. Wan, T. Wu and Y. Yang, "Estimating the area burned by agricultural fires from Landsat 8 Data using the Vegetation Difference Index and Burn Scar Index," *International Journal of Wildland Fire,* vol. 27, p. 217–227, 2018.

[32] T. A. Warner, N. S. Skowronski and M. R. Gallagher, "High spatial resolution burn severity mapping of the New Jersey Pine Barrens with WorldView-3 near-infrared and shortwave infrared imagery," *International Journal of Remote Sensing,* vol. 38, no. 2, pp. 598-616, 2017.

[33] C. E. Blinn, M. N. House, R. H. Wynne, V. A. Thomas, T. R. Fox and M. Sumnall, "Landsat 8 based leaf area index estimation in loblolly pine plantations," *Forests,* vol. 10, no. 3, 2019.

[34] P. Stych, J. Lastovicka, R. Hladky and D. Paluba, "Evaluation of the Influence of Disturbances on Forest Vegetation Using the Time Series of Landsat Data: A Comparison Study of the Low Tatras and Sumava National Parks," *ISPRS Int. J. Geo-Inf.,* vol. 8, no. 2, 2019.

[35] N. Kaoungku, K. Suksut, R. Chanklan, K. Kerdprasop and N. Kerdrasop, "The silhouette width criterion for clustering and association mining to select image features," *International journal of machine learning and computing,* vol. 8, no. 1, pp. 1-5, 2018.