

Spatial Temporal Density-Based Spatial Clustering Applications With Noise (ST-DBSCAN) Untuk Pengelompokan Titik Kebakaran Hutan dan Lahan di Provinsi Riau Pada Tahun 2021

Spatial Temporal Density Based Spatial Clustering Applications With Noise (ST-DBSCAN) For Classification of Forest and Land Fire Points in Riau Province in 2021.

Nilamelinda Putri¹, Tiani Wahyu Utami², Rochdi Wasono³

^{1,2,3} Universitas Muhammadiyah Semarang, Kota Semarang

Corresponding author: nilamelindaputri@gmail.com

Abstrak

Kebakaran hutan dan lahan sering terjadi pada masa kemarau yang menimbulkan dampak buruk bagi lingkungan. Kejadian tersebut umumnya terjadi di Indonesia salah satunya di Provinsi Riau. Untuk mengatasi hal tersebut, perlu adanya pengelompokan titik panas untuk mengetahui daerah yang rawan terjadi kebakaran hutan dan lahan menggunakan teknik *clustering*. Metode *clustering* yang biasa digunakan yaitu *Spatial Temporal Density-Based Spatial Clustering Of Applications With Noise (ST-DBSCAN)*. Data yang digunakan berupa data titik panas pada periode Januari 2021 hingga Desember 2021 di Provinsi Riau. Penelitian ini menggunakan parameter jarak *spasial* ($Eps1 = 0,025$), parameter jarak *temporal* ($Eps2 = 30$) dan jumlah minimal anggota *cluster* ($MinPts = 2$), menghasilkan total *cluster* 484 dengan 6 *cluster* besar dan 866 *noise*. Pusat titik kebakaran hutan banyak ditemukan di Kabupaten Pelalawan dan Kabupaten Bengkalis. Tipe pola *satio-temporal* yang paling banyak ditemukan adalah tipe *irregular*. Nilai *silhouette coefficient spasial* sebesar 0,1282 dan nilai *silhouette coefficient temporal* sebesar -0,8298.

Kata Kunci: *Clustering, Kebakaran Hutan dan Lahan, Silhouette Coefficient, Spatial Temporal Density-Based Spatial Clustering Of Applications With Noise (ST-DBSCAN)*

Abstract

Forest and land fires often occur during the dry season, causing adverse impacts on the environment. These incidents commonly happen in Indonesia, including in Riau Province. To address this issue, it is necessary to group hotspots to identify areas prone to forest and land fires using clustering techniques. One of the clustering methods used is the Spatial Temporal Density-Based Spatial Clustering Of Applications With Noise (ST-DBSCAN). The data used consists of hotspot data from January 2021 to December 2021 in Riau Province. This study used spatial distance parameter ($Eps1 = 0.025$), temporal distance parameter ($Eps2 = 30$), and minimum cluster member count ($MinPts = 2$), resulting in a total of 484 clusters with 6 major clusters and 866 noise points. The centers of forest fire hotspots were frequently found in Pelalawan Regency and Bengkalis Regency. The most commonly found spatio-temporal pattern was of irregular type. The spatial silhouette coefficient value was 0.1282, and the temporal silhouette coefficient value was -0.8298.

Keywords: *Clustering, Forest and land fires, Silhouette Coefficient, Spatial Temporal Density-Based Spatial Clustering Of Applications With Noise (ST-DBSCAN)*

PENDAHULUAN

Kebakaran hutan dan lahan merupakan suatu kejadian yang umum terjadi di Indonesia, terutama saat masa kemarau. Hal ini dapat menimbulkan berbagai dampak buruk bagi lingkungan seperti pencemaran udara, depresiasi ekonomi hutan dan kualitas tanah,

peningkatan pelepasan karbon serta dampak terhadap keanekaragaman hayati. Kebakaran hutan terjadi di lahan kering maupun lahan basah, termasuk lahan/hutan gambut dan paling sering terjadi di masa kemarau ketika lahan basah tersebut mengalami kekeringan (Adinugroho, 2005).

Provinsi Riau memiliki hutan yang cukup luas. Menurut Badan Penanggulangan Bencana Daerah (BPBD) Provinsi Riau, akan ada sekitar 2.730 titik panas yang terdeteksi di wilayah Riau per 21 Oktober 2022. Kemarau yang berkepanjangan serta wilayah hutan yang gersang menentukan resiko wilayah yang telah dinyatakan rawan kebakaran hutan. Tindakan utama yang dapat dilakukan adalah mengidentifikasi daerah-daerah yang paling rentan terhadap kebakaran hutan untuk mengurangi risikonya. Proses ini dapat disederhanakan melalui pengelompokan atau *clustering*.

ST-DBSCAN (*Spatial Temporal Density-Based Spatial Clustering Applications With Noise*) merupakan salah satu teknik pengelompokan yang dapat diterapkan. Algoritma ST-DBSCAN adalah algoritma *unsupervised learning* yang bersifat nonparametric, oleh karena itu dalam implementasinya tidak bergantung pada asumsi tertentu. Metode ini merupakan salah satu teknik terunggul untuk mengidentifikasi *cluster* pada basis data spasial yang besar (Gaonkar, 2013).

Adapun penelitian terkait metode *Spatial Temporal Density-Based Spatial Clustering Applications With Noise* (ST-DBSCAN) pernah dilakukan oleh (Trisnaningtyas, 2014), dengan $Eps_1 = 0,4$, $Eps_2 = 0$, $\Delta \epsilon = 2$, dan $MinPts = 5$ menghasilkan 4 *cluster* dan 18 *noise* serta *Silhouette Coefficient* senilai 0,020.

Berlandaskan pada pemaparan tadi, tujuan dari penelitian ini adalah mengimplementasikan algoritma ST-DBSCAN (*Spatio Temporal-Density Based Spatial Clustering Applications with Noise*) untuk mengelompokkan data titik *hotspot* di Provinsi Riau tahun 2021 dan mengevaluasi kualitas dari kelompok tersebut menggunakan *Silhouette Coefficient*.

METODE

a. Sumber data

Penelitian ini menggunakan data sekunder mengenai titik kebakaran hutan dan lahan di Provinsi Riau dari tanggal 1 Januari 2021 hingga 31 Desember 2021. Data *hotspot* yang digunakan diperoleh dari Badan Riset dan Inovasi Nasional (BRIN) dan disediakan oleh Lembaga Penerbangan dan Antariksa Nasional (LAPAN). Jumlah data penelitian yang digunakan adalah sebanyak 3244 titik *hotspot*.

b. Variabel Penelitian dan Struktur Data

Variabel yang dianalisis dalam penelitian ini meliputi *longitude* dan *latitude* sebagai aspek spasial, serta tanggal munculnya titik *hotspot* sebagai aspek temporal

Table 1 Struktur Data

| No. | Longitude (X) | Latitude (Y) | Date (T) |
|-----|---------------|--------------|----------------|
| 1 | x_1 | y_1 | 1 Januari 2021 |
| 2 | x_2 | y_2 | 2 Januari 2021 |
| 3 | x_3 | y_3 | 3 Januari 2021 |
| ⋮ | ⋮ | ⋮ | ⋮ |

| | | | |
|------|------------|------------|------------------|
| ⋮ | ⋮ | ⋮ | ⋮ |
| 3244 | x_{3244} | y_{3244} | 31 Desember 2021 |

c. Langkah Penelitian

1. Pengumpulan data
2. *Preprocessing* data
 - a. Perubahan format tanggal menjadi *number*
 - b. Pemilihan variabel
3. Menggunakan grafik *k-distance* untuk penentuan nilai *Epsilon* (Eps) dan *Minimal Points* (MinPts)
4. Menghitung jarak Euclidean berdasarkan aspek *spasial* (*longitude* dan *latitude*) dan aspek temporal (tanggal *hotspot*)

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{in} - x_{jn})^2}$$

5. Melakukan penelitian kluster menggunakan metode algoritma ST-DBSCAN.
6. Mengevaluasi kualitas kluster dengan *Silhouette Coefficient*

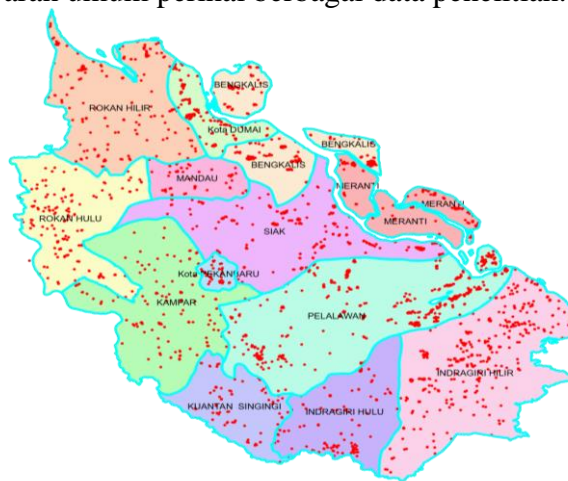
$$s(i) = \frac{b(i) - \alpha(i)}{\max \alpha(i), b(i)}$$

7. Melakukan analisis pola *spatio-temporal*
8. Melakukan visualisasi dan interpretasi dari pembentukan *cluster* hasil dari penggunaan metode ST-DBSCAN.

HASIL DAN PEMBAHASAN

a. Analisis Deskriptif

Data yang digunakan pada penelitian ini adalah data titik kebakaran hutan di Provinsi Riau tahun 2021. Data titik kebakara hutan yang diambil sebanyak 3244 titik panas dengan atribut yaitu tanggal, *longitude*, *latitude*. Sebelum melakukan analisis *cluster*, maka dilaksanakan analisis deskriptif terlebih dahulu agar bisa memperoleh sebuah gambaran umum perihal berbagai data penelitian.



Gambar 1 Persebaran Titik Hotspot di Provinsi Riau

Gambar 1 merupakan peta persebaran titik *hotspot* di Provinsi Riau pada tahun 2021. Dapat dilihat bahwa persebaran titik *hotspot* hampir menyebar pada segala penjuru Kabupaten yang ditemui pada Provinsi Riau. Jumlah titik *hotspot* yang menyebar berjumlah sekitar 3244 titik *hotspot*, hal ini mengakibatkan Provinsi Riau menduduki peringkat pertama dalam kasus kebakaran hutan dan lahan. Jumlah titik *hotspot* tersebut yang nantinya akan di kelompokkan menggunakan algoritma ST-DBSCAN.

b. Preprocessing Data

Preprocessing data adalah tahap data diubah supaya sesuai dengan proses mining. Berikut adalah langkah-langkah *preprocessing* data yang dilakukan:

i. Perubahan format pada kolom tanggal

Penyesuaian format dilakukan dengan tujuan agar memudahkan dalam proses penghitungan jarak temporal menggunakan matrik Euclidean.

Tabel 1 Transformasi format tanggal menjadi Number

| No | Tanggal | Tanggal (<i>number</i>) |
|------|------------|---------------------------|
| 1 | 2021-01-01 | 44197 |
| 2 | 2021-01-05 | 44201 |
| 3 | 2021-01-05 | 44201 |
| 4 | 2021-01-05 | 44201 |
| 5 | 2021-01-06 | 44202 |
| ⋮ | ⋮ | ⋮ |
| 3244 | 2021-12-30 | 44560 |

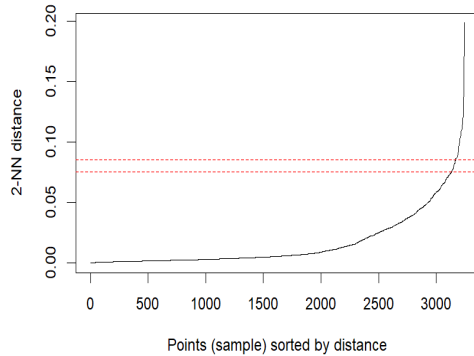
Berdasarkan tabel 1. diperoleh hasil perubahan format tanggal 1 Januari 2021 menjadi format number 44197. Perubahan format tanggal ini diperlukan agar memudahkan dalam proses perhitungan jarak temporal menggunakan rumus jarak Euclidean.

ii. Pemilihan variabel

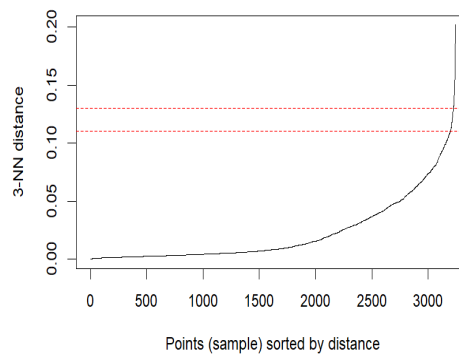
Pada tahapan ini dilakukan pemilihan variabel yang akan diperlukan dalam penelitian. Variabel awal yaitu *longitude*, *latitude*, tingkat kepercayaan dan radius kemungkinan. Pada tahap ini dilakukann pemilihan variabel yang akan diperlukan dalam penelitian yaitu *longitude*, *latitude* dan tanggal terdeteksinya titik kebakaran hutan oleh satelit.

c. Penentuan Nilai *Epsilon* (Eps) dan *Minimal Points* (MinPts)

Penentuan nilai *Epsilon* (Eps) dan *Minimal Points* (MinPts) yang optimal menggunakan grafik *k-distance*. Pada penelitian kali ini, nilai MinPts yang akan diteliti yaitu 2 dan 3.



Gambar 2 k-dist dengan k=2



Gambar 3 k-dist dengan k=3

Bisa ditinjau melalui gambar 2 bahwa plot *k-distance* k=2 terjadi perubahan tajam di antara titik 0,025 sampai 0,075. Pada gambar 3 bisa ditinjau jika plot *k-distance* k=3 terjadi perubahan tajam di antara titik 0,045 sampai 0,13. Artinya, jika MinPts=2, rentang optimal untuk Eps1 adalah dari 0,025 sampai dengan 0,075. Namun jika MinPts=3, rentang optimal untuk Eps1 adalah dari 0,045 sampai dengan 0,13. Total pasangan parameter Eps1, Eps2 dan MinPts pada penelitian ini berjumlah 24 pasangan dan nilai Eps1 bisa ditinjau melalui tabel 2 dibawah:

Tabel 2 Eps yang diteliti

| MinPts | Eps1 |
|--------|-------|
| 2 | 0,025 |
| | 0,035 |
| | 0,045 |
| 3 | 0,045 |
| | 0,055 |
| | 0,065 |

Berlandaskan tabel 2 bisa diketahui jika penelitian ini menggunakan pasangan parameter MinPts= 2 dengan pasangan parameter Eps1 = 0,025; 0,035; 0,045 dan MinPts= 3 dengan pasangan parameter Eps2= 0,045; 0,055; 0,065. Sedangkan untuk Eps2 menggunakan perbandingan mingguan yakni 7 hari, 14 hari, 21 hari serta pada 30 hari. Pasangan untuk parameter tersebut ditentukan oleh peneliti berdasarkan nilai yang optimal dari grafik *k-distance* diatas.

d. Jarak Euclidean

Dalam proses pembentukan kluster, dibutuhkan pengukuran jarak antara setiap objek. Jarak antara tiap objek diukur dengan metode Euclidean. Perhitungan manual jarak Euclidean berlandaskan dalam koordinat spasial (*longitude* dan *latitude*) serta unsur aspek temporal (tanggal terjadinya kejadian kebakaran) bisa ditinjau melalui tabel berikut:

Tabel 3 Perhitungan Jarak Euclidean

| (i,j) | d(i,j) spasial | d(i,j) temporal |
|-------------|----------------|-----------------|
| (1,1) | 0 | 0 |
| (1,2) | 2,13058 | 4 |
| (1,3) | 1,50522 | 4 |
| (1,4) | 0,66085 | 4 |
| (1,5) | 0,27664 | 5 |
| ⋮ | ⋮ | ⋮ |
| (3244,3244) | 0 | 0 |

Tabel 3 diatas merupakan tabel yang berisikan jarak antar semua objek pengamatan berdasarkan aspek spasial (*longitude* dan *latitude*) dan aspek temporal (tanggal terjadinya kebakaran). Sebagai contoh pada baris kedua dan kolom pertama menunjukkan bahwa jarak antar objek sekitar 2,13058. Hal tersebut berarti jarak antar objek pertama dengan objek kedua sekitar 2,13058.

e. Pembentukan *Cluster*

Cluster hasil dari 24 pasangan parameter Eps1, Eps2, serta MinPts dalam suatu data titik kebakaran hutan di Provinsi Riau telah dibuat menggunakan metode ST-DBSCAN dengan bantuan perangkat lunak R. Kemudian, dari hasil *cluster* tersebut, dipilih parameter dengan nilai *Silhouette Coefficient* tertinggi sebagaimana tercantum di Tabel 4.

Tabel 4 Parameter Terbaik

| MinPts | Eps1 | Eps2 | Jumlah Cluster | Jumlah Noise | <i>Silhouette Spasial</i> | <i>Silhouette Temporal</i> |
|--------|-------|------|----------------|--------------|---------------------------|----------------------------|
| 2 | 0,025 | 30 | 484 | 866 | 0,1282 | -0,8298 |
| 2 | 0,035 | 30 | 485 | 745 | 0,0967 | -0,827 |
| 2 | 0,045 | 30 | 463 | 652 | 0,0734 | -0,8192 |
| 3 | 0,045 | 30 | 267 | 1043 | 0,0149 | -0,8080 |

Pada tabel 4 dapat dilihat bahwa parameter MinPts=2; Eps1=0,025; Eps2=30; mendapatkan nilai *Silhouette Coefficient* tertinggi yaitu *Silhouette Spasial* sebesar 0,1282, dengan nilai *Silhouette Temporal* sebesar -0,8298 serta jumlah *cluster* 484 dan noise sebanyak 866 titik. Pada penelitian ini menggunakan pasangan parameter MinPts=2; Eps1=0,025; Eps2=30 yang selanjutnya akan dianalisis pola *spatial temporal*.

f. Analisis Pola *Spatio-Temporal*

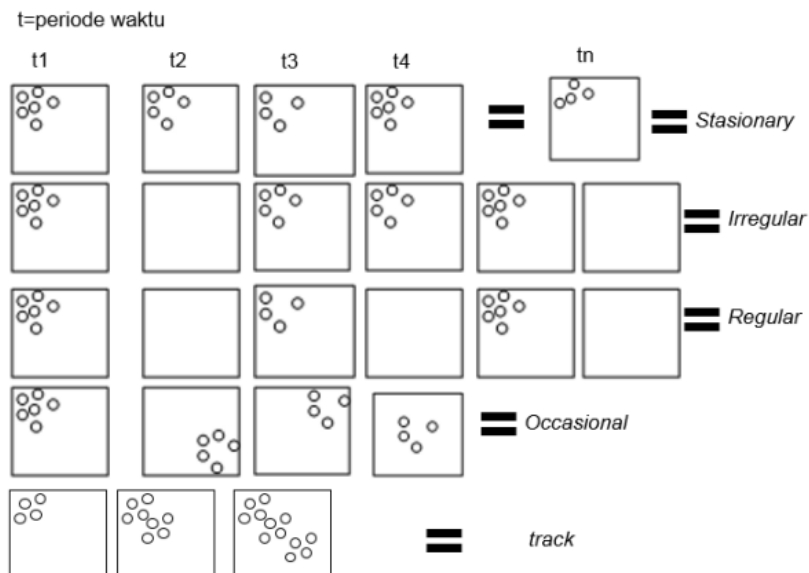
Pengelompokan pada pola yang berupa *spatio-temporal* bisa dibagi sebagai 4 bagian yakni *stasioneri*, *reappearing*, lalu ada *oscasional*, serta *track*. Diterangkan oleh Pöelitz serta Andrienko (2010), jika suatu *cluster* yang signifikan untuk analisis yakni yang mempunyai total titik dengan minimalnya kisaran 30 titik. Sementara itu, *cluster* dengan total titik di bawah 30 dianggap cukup kecil serta pada polanya tidak bisa untuk teridentifikasi. Dengan menggunakan parameter paling akurat

Eps1=0,025; lalu Eps2=30; serta dengan MinPts=2, ditemukan enam *cluster* besar dimana mempunyai sejumlah titik yang lebih dari 30 titik (Lihat Tabel 5).

Tabel 5 Cluster Besar

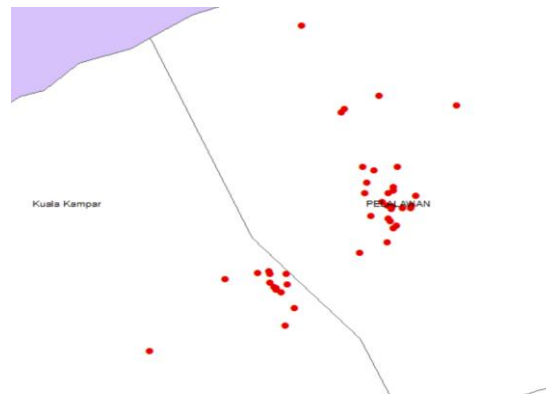
| <i>Cluster</i> ke- | Jumlah titik <i>hotspot</i> |
|--------------------|--------------------------------|
| 42 | 42 |
| 65 | 54 |
| 69 | 50 |
| 74 | 68 |
| 134 | 52 |
| 301 | 44 |

Tabel 5 merupakan 6 *cluster* besar dengan jumlah titik lebih dari 30 titik *hotspot*. *Cluster* tersebut meliputi *cluster* ke 42 dengan 42 titik *hotspot*, *cluster* ke 65 dengan 54 titik *hotspot*, *cluster* ke 69 dengan 50 titik *hotspot*, *cluster* 74 dengan 68 titik *hotspot*, *cluster* 134 dengan 52 titik *hotspot*, *cluster* 301 dengan 44 titik *hotspot*. *Cluster* tersebut yang nantinya akan dianalisis polanya, gambaran pola dapat dilihat pada gambar 4



Gambar 4 Ilustrasi Pola Spatio-Temporal (Purwanto, 2012)

Cluster dibagi berdasarkan setiap periode, mengingat parameter paling baik dalam penelitian yakni berupa Eps2=30, sehingga setiap periode berlangsung selama 30 hari untuk membandingkan distribusi titik *hotspot* dalam satu periode. Sebagai contoh, pada *cluster* 42 terdapat pola *irregular* (lihat Gambar 5).



Gambar 5 Cluster ke-42

Gambar 5 merupakan visualisasi titik *hotspot cluster* ke 42. Titik *hotspot* berpusat di Kabupaten Pelalawan. *Cluster 42* memiliki pola *spatio-temporal* yaitu pola *irregular*, yang berarti pengelompokan *cluster* terjadi pada tempat yang sama yaitu di Kabupaten Pelalawan tetapi muncul pada periode waktu yang berbeda yaitu pada bulan Februari – April 2021.

Setelah melakukan interpretasi 6 *cluster* besar maka dapat dilihat pada tabel daerah yang menjadi pusat dari lokasi titik *hotspot*

Tabel 6 Daerah yang menjadi Pusat dari Lokasi Titik *Hotspot*

| <i>Cluster ke-</i> | Jumlah <i>Hotspot</i> | Pusat Titik <i>Hotspot</i> | Waktu | Pola <i>spatial temporal</i> |
|--------------------|-----------------------|----------------------------|-----------------------|------------------------------|
| 42 | 42 | Kabupaten Pelalawan | Februari - April 2021 | <i>Irregular</i> |
| 65 | 54 | Kabupaten Pelalawan | Februari - Maret 2021 | <i>Irregular</i> |
| 69 | 50 | Kabupaten Bengkalis | Maret - November 2021 | <i>Irregular</i> |
| 74 | 68 | Kota Dumai | Maret 2021 | <i>Track</i> |
| 134 | 52 | Kabupaten Bengkalis | Februari - Maret 2021 | <i>Irregular</i> |
| 301 | 44 | Kabupaten Rokan Hilir | Juli 2021 | <i>Regular</i> |

Pada tabel 6. dapat dilihat bahwa Kabupaten Bengkalis merupakan wilayah dengan konsentrasi titik panas tertinggi di Provinsi Riau dan mengalami penggerombolan dengan pola yang tidak teratur. Fenomena penggerombolan ini terjadi dari bulan Februari hingga November 2021, dengan total 102 titik panas. Melalui analisis ini, diharapkan pemerintah daerah dapat mengambil langkah-langkah yang sesuai untuk mengurangi dampak kerugian, baik secara materiil maupun terhadap nyawa manusia, di wilayah yang rentan terhadap kebakaran hutan.



KESIMPULAN

Pada penelitian ini yang menghasilkan 484 *cluster*, *noise* sebanyak 866 dan 6 *cluster* besar yang dianalisis dengan parameter paling baik yakni $Eps_1=0,025$; $Eps_2=30$ serta $MinPts=2$. Terdapat 6 *cluster* besar tersebut meliputi *cluster* 42 pusat titik *hotspot* di Kabupaten Pelalawan dengan jumlah *hotspot* 42 titik, *cluster* 65 pusat titik *hotspot* di Kabupaten Pelalawan dengan jumlah *hotspot* 54 titik, *cluster* 69 pusat titik *hotspot* di Kabupaten Bengkalis dengan jumlah *hotspot* 50 titik, *cluster* 74 pusat titik *hotspot* di Kota Dumai dengan jumlah *hotspot* 68 titik, *cluster* 134 pusat titik *hotspot* di Kabupaten Bengkalis dengan jumlah *hotspot* 54 titik, *cluster* 301 pusat titik *hotspot* di Kabupaten Rokan Hilir dengan jumlah *hotspot* 44 titik. Parameter tersebut menghasilkan *Silhouette Coefficient Spasial* yaitu 0,1282 dengan *Silhouette Coefficient Temporal* yaitu -0,8298. Titik pusat kebakaran hutan sering kali terjadi di Kabupaten Pelalawan dan Kabupaten Bengkalis. Pola *spasial-temporal* yang dominan adalah pola *irregular*.

DAFTAR PUSTAKA

- Adinugroho, W. C., I. Nyoman N. Suryadiputra, & and Bambang H. Saharjo. (2005). Panduan Pengendalian Kebakaran Hutan dan Lahan Gambut.
- Birant, D., & Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data and Knowledge Engineering*, 60(1), 208–221.
- Collier Bernard Carey Ellen Grusy Curt Marjaniemi Donald Sautter, K. (1998). *A Perspective on Data Mining*.
- Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM*, 49(9), 76-82.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). *Knowledge Discovery and Data Mining: Towards a Unifying Framework*.
- Furqon, M. T., & Muflikhah, L. (2016). Clustering The Potential Risk Of Tsunami Using Density-Based Spatial Clustering Of Application With Noise (DBSCAN). *Journal of Environmental Engineering & Sustainable Technology JEEST*, 03(01), 1–8.
- Gaonkar, M. N. & Sawant, K., 2013. AutoEPsDBSCAN. DBSCAN with Eps Automatic for Large, pp. 2:2319-2526.
- Giglio, L., Descloitres, J., Justice, C. O., & Kaufman, Y. J. (2003). An enhanced contextual fire detection algorithm for MODIS. *Remote Sensing of Environment*, 87(2–3), 273–282.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining concepts and techniques third edition*. University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University.
- Pölitiz, C., Andrienko, G., & Andrienko, N. (2010). Finding arbitrary shaped clusters with related extents in space and time.
- Ramadhan, A., Efendi, Z., & Mustakim, M. (2017). Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling. In *Seminar Nasional Teknologi Informasi Komunikasi dan Industri* (pp. 219-226).
- Purwanto, U. Y. (2012). *Spatial Hotspots Clustering of Forest and Land Fires using DBSCAN and ST-DBSCAN*.



- Rowell, A., & Moore, P. F. (2000). Global review of forest fires (pp. 66-66). Forests for Life Programme Unit, WWF International.
- Swindiaro, V. T. P., Sarno, R., & Novitasari, D. C. (2018, September). Integration of fuzzy C-means clustering and TOPSIS (FCM-TOPSIS) with silhouette analysis for multi criteria parameter data. In 2018 International Seminar on Application for Technology of Information and Communication (pp. 463-468). IEEE.
- Syaufina, L. (2008). Kebakaran hutan dan lahan di Indonesia: perilaku api, penyebab, dan dampak kebakaran. Bayumedia Pub.
- Tan, P. N., Steinbach, M., & Kumar, V. (2006). Data mining introduction. People's Posts and Telecommunications Publishing House, Beijing.
- Trisnaningtyas, A. (2014). Pengelompokan Data Indeks Pembangunan Manusia di Pulau Jawa dengan Algoritme ST-DBSCAN dan Bahasa Pemrograman Python.
- Zaki, M. J., Meira Jr, W., & Meira, W. (2014). Data mining and analysis: fundamental concepts and algorithms. Cambridge University Press.