



**ANALISIS CLUSTERING K-MEANS PADA PENGELOMPOKKAN TITIK PANAS
KEBAKARAN HUTAN DAN LAHAN**

Putri Yuli Utami*

Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Muhammadiyah
Pontianak, Jalan A. Yani No. 111, Pontianak, Kalimantan Barat 78124

*email: putriyuli@unmuhpnk.ac.id

Received: 2023-05-20 Accepted: 2023-06-20 Published: 2023-06-30

Abstrak

Kebakaran hutan dan lahan adalah masalah besar yang dapat menyebabkan kerusakan ekosistem hutan dan lingkungan. Salah satu indikator penting untuk mencegah kebakaran hutan dan lahan adalah dengan menyediakan informasi tentang titik panas dan karakteristik wilayah yang berpotensi munculnya titik panas. Tanda-tanda kebakaran hutan dan lahan dapat dideteksi dengan mengidentifikasi titik api di lokasi tertentu. Penelitian ini mengimplementasikan metode K-Means menggunakan RapidMiner, aplikasi ini freeware dan menyediakan berbagai metode pengolahan data yang mudah digunakan. Metode K-Means digunakan untuk mengumpulkan titik panas kebakaran hutan dan lahan. Proses pengumpulan data, pra-pemrosesan data, penerapan metode, evaluasi model, dan analisis hasil adalah bagian dari penelitian ini. Data titik panas dari tahun 2019 hingga 2021 digunakan untuk penelitian ini. Hasil uji validasi *cluster* menggunakan nilai koefisien silhouette algoritma K-Means menunjukkan nilai koefisien silhouette terbaik sebesar 0,756 termasuk pada kategori *strong structure* dengan nilai $k = 2$.

Kata kunci: *clustering, data mining, K-Means, lahan gambut, RapidMiner, titik panas*

Abstract

Forest and land fires are a major problem that can cause damage to forest ecosystems and the environment. One of the important indicators to prevent forest and land fires is to provide information about hotspots and the characteristics of areas with potential hotspots. Indications of forest and land fires can be known through hotspots detected in a particular location. This research implements the K-Means method using RapidMiner, this application is freeware and provides various data processing methods that are easy to use. The K-Means method is used to collect hotspots for forest and land fires. The processes of data collection, data pre-processing, method application, model evaluation, and result analysis are part of this research. Hotspot data from 2017 to 2021 was used for this study. The results of the cluster validation test using the silhouette coefficient value of the K-Means algorithm show the best silhouette coefficient value of 0.756 that's include strong structure with a value of $k = 2$.

Keywords: *clustering, data mining, hotspots, K-Means, peatlands, RapidMiner*

How to cite (in APA style): Utami, P. Y. (2023). Analisis clustering k-means pada pengelompokan titik panas kebakaran hutan dan lahan. *Jurnal Pendidikan Informatika dan Sains*, 12(1), 165-172.

Copyright (c) 2023 Putri Yuli Utami
DOI: 10.31571/saintek.v12i1.6001



PENDAHULUAN

Letak Geografis Indonesia berada dipertemuan empat lempeng tektonik, yaitu Eurasia, Indo-Australia, Filipina, dan Pasifik hal ini membuat Indonesia rawan bencana. Indonesia juga rawan terhadap banjir, tanah longsor, banjir bandang, kekeringan dan kondisi cuaca ekstrim. Cuaca ekstrim menyebabkan kebakaran hutan dan lahan (D. Oleh, B, 2018). Kalimantan Barat terletak diantara 2°08 LU dan 3°05 LS dan 108°0 dan 114°10 BT. Kalimantan Barat dilintasi oleh garis khatulistiwa yang berada Kota Pontianak dan beriklim tropis. Kalimantan Barat memiliki suhu udara dan kelembaban yang tinggi (I. Sumariyanto & J. Hutauruk, 2021).

Informasi luas kebakaran hutan dan lahan di Kalimantan Barat meningkat antara tahun 2014 dan 2019, yaitu. H. 3.556,10 ha dan 151.070,00 ha pada tahun 2019 (Kementerian Lingkungan Hidup dan Kehutanan Republik Indonesia, 2019). Potensi kebakaran hutan dan lahan di desa-desa di Kalimantan Barat sangat besar, yakni menyebar di 322 desa diantaranya di Kabupaten Kubu Raya memiliki *hotspot* di 18 desa. Penyebaran data *hotspot* masih terbatas, yaitu informasi yang diberikan dalam bentuk laporan kebakaran hutan dan lahan tidak secara akurat menunjukkan wilayah atau daerah mana yang memiliki titik panas. *Clustering* adalah teknik data mining untuk mengelompokkan objek yang memiliki kemiripan karakteristik. Penelitian *clustering* dilakukan untuk mengelompokkan kawasan yang memiliki potensi kebakaran hutan (Hermawati & Sitanggang, 2016), (Gunawan, 2021).

Penelitian menggunakan metode K-Means, untuk mengklasifikasikan area kebakaran hutan dengan karakteristik serupa dilakukan dengan mengelompokkan data kedalam *cluster*. Setiap *cluster* memiliki *centroid* (tengah) yang mewakili *cluster* tersebut. Penelitian *clustering* bencana menggunakan algoritma K-Means, di Jawa Barat tahun 2022 memberikan nilai *centroid* untuk tiga cluster yaitu C1 (low level cluster), C2 (cluster tingkat menengah), C3 (kluster tingkat tinggi) (Athifaturrofifah et al., 2019) (M. Firman, A. Halik, & L. Septiana, 2022). Algoritma K-Means adalah metode pengelompokan data yang membagi data kedalam beberapa kelompok berdasarkan ciri masing - masing data (Handayani et al., 2021).

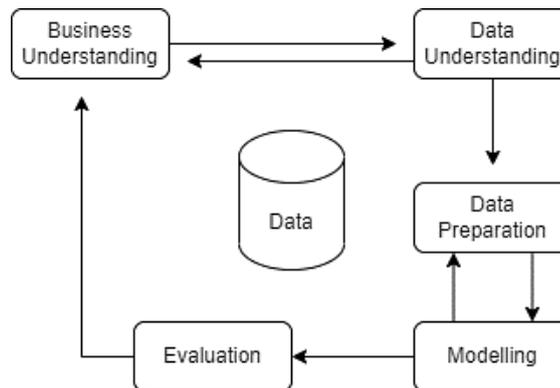
Penelitian lain yang menggunakan K-Means di Provinsi Riau pada *hotspot*. Dataset dibagi menjadi tiga klaster, yaitu 133 titik yang termasuk dalam area High Risk Areas, 101 titik yang termasuk dalam klaster area rawan kubus dan 77 indikasi. kelompok daerah yang tidak rawan *hotspot*. Nilai DBI (Davies Bouldin Index) sebesar 0,361 menunjukkan bahwa 3 klaster K-Means sudah optimal (Id & Angraini, 2018).

Sebuah studi yang dilakukan dengan RapidMiner mencoba mengklasifikasikan data penjualan aksesoris di Rafadel Acc Store. Penelitian ini mengelompokkan data penjualan untuk mengetahui produk paling laris terjual, yaitu yang paling banyak terjual (Mardalius, 2018). Penerapan RapidMiner menggunakan K-Means digunakan untuk mengklasifikasikan kasus vaksin campak pada anak kecil. Penelitian ini dijadikan masukan oleh pemerintah khususnya provinsi untuk lebih memperhatikan provinsi dengan kasus campak yang tinggi dan meningkatkan sosialisasi vaksinasi campak pada anak balita (R.W. Sari, et al., 2018). Penerapan RapidMiner dengan metode K-Means juga diimplementasikan dalam studi klasifikasi wilayah kasus demam berdarah dengue (DBD). Penelitian ini memberikan informasi provinsi-provinsi banyak terkena penyakit demam berdarah dengue (DBD). Informasi ini digunakan oleh pemerintah untuk penanggulangan penyakit DBD di masyarakat (K. Fatmawati & A. P. Windarto, 2018).

Penelitian ini mengimplementasikan metode K-Means *clustering* titik panas kebakaran hutan dan lahan menggunakan aplikasi RapidMiner. Analisis *clustering* dilakukan untuk melihat performa metode K-Means dalam mengelompokkan titik panas yakni menggunakan pengujian *silhouette coefficient*.

METODE

Pada penelitian ini menggunakan metode CRISP-DM, tahapan CRISP-DM ditunjukkan pada Gambar 1.



Gambar 1 Tahapan CRISP-DM

Business Understanding

Tahap ini dilakukan pencarian informasi terkait kebakaran hutan dan lahan di Kalimantan Barat, yakni Algoritma K-Means dan menganalisis penerapan Algoritma K-Means untuk mengelompokkan potensi daerah rawan bencana kebakaran hutan dan lahan.

Data Understanding

Data penelitian merupakan data kebakaran hutan dan lahan di Kabupaten Kubu Raya Kalimantan Barat. Dataset tersebut terdiri dari 7084 records yang berisi atribut No, Desa, Latitude, Brightness, FRP, Confidence dan Year.

Data Preparation

Dataset penelitian ini diperoleh dari <https://hotspot.brin.go.id/> data dari tahun 2019-2021 lokasi di Kabupaten Kubu Raya. Normalisasi data menggunakan Microsoft Excel 16.4 kemudian mengimplementasikan algoritma K-Means menggunakan software RapidMiner versi 10. Persiapan data sebanyak 7084 dari 3684 hotspot tahun 2019, 370 data hotspot tahun 2020, dan 3030 data hotspot tahun 2021. Pada tahap ini biasanya dilakukan proses pembersihan data, salah satunya *missing value*.

Modelling

Penelitian ini mengimplementasikan Algoritma K-Means menggunakan *tools RapidMiner*. K-Means merupakan algoritma data mining dengan membagi data menjadi beberapa kelompok berdasarkan kemiripan dari setiap data (Pramana S., et al., 2018) (Talakua, M.W., et al., 2017). Pengelompokan data menggunakan K-Means dilakukan sebagai berikut (Yuan C & Yang H., 2019)

- Tentukan banyaknya kelompok.
- Data secara acak dibagi menjadi beberapa kelompok.
- Hitung pusat cluster dari data di setiap cluster. Posisi pusat setiap kelompok ditentukan oleh nilai rata-rata semua data untuk setiap fungsi. Jika M adalah jumlah data dalam suatu grup, i adalah jumlah grup, dan p adalah dimensi data, persamaan tersebut menggunakan persamaan (1) tersebut untuk menghitung pusat massa fitur ke-i.

$$K^i = \frac{1}{M} \sum_{j=1}^M X_j$$

Keterangan :

K^i = centroid ke-i

M = jumlah data dalam kelompok

X_j = titik ke-j

- d. Sejajarkan setiap datum dengan centroid terdekat. Hitung jarak data ke pusat cluster menggunakan jarak Euclidean yang dinyatakan dalam persamaan (2).

$$d = \sqrt{(x_1 - x_2)^2 + (c_1 - c_2)^2} \quad (2)$$

Keterangan :

d = jarak *Euclidean*

C_j = *centroid* ke-j

X_j = titik ke-j

- e. Jika terdapat perubahan nilai *centroid* di atas nilai ambang yang ditentukan, selanjutnya dilakukan kembali penghitungan pada langkah C dan seterusnya.

Evaluation

Evaluasi model menggunakan teknik *Silhouette Coefficient* untuk memvalidasi model dan mengukur kualitas *cluster* yang terbentuk. Evaluasi *Silhouette Coefficient* menggunakan persamaan (3) (Maulana, A., Akbar, K. N., & Nurahman, N, 2021):

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

Keterangan :

$S(i)$: Nilai *Silhouette Coefficient* data ke-i

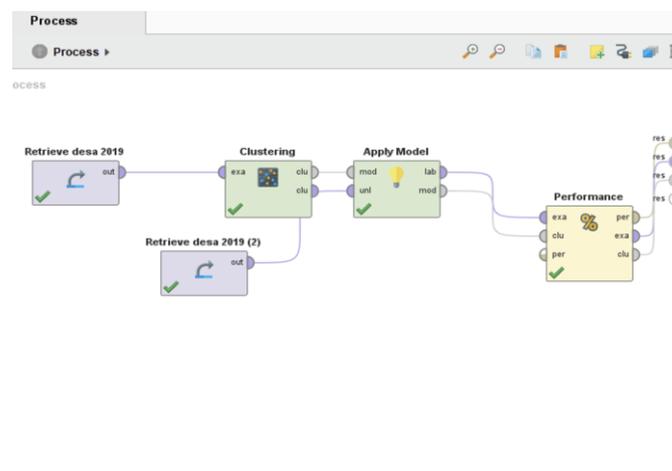
$a(i)$: Rata-rata jarak suatu data ke-i

$b(i)$: Rata-rata jarak suatu data ke-i dengan semua objek

HASIL DAN PEMBAHASAN

Perancangan K-Means Clustering pada RapidMiner

Perancangan model K-Means dilakukan dengan RapidMiner diperlihatkan pada Gambar 2 berikut ini.



Gambar 2 Rancangan K-Means pada RapidMiner

Pada tahap ini dilakukan perancangan dan implementasi data. Data yang telah ditransformasikan selanjutnya dibuat rancangan dataset menggunakan RapidMiner. Rancangan dataset ini menggunakan algoritma K-Means. Tahapan yang dilakukan pada Rapidminer pertama *retrieve* data atau input data proses K-Means, selanjutnya menghubungkan port-port dari operator

clustering, apply model, dan performance. Rancangan model clustering K-Means dari tahun 2019 sampai tahun 2021 terlihat pada Gambar 3.

Cluster Model

Cluster 0: 2229 items
 Cluster 1: 1455 items
 Total number of items: 3684

(a)

Cluster Model

Cluster 0: 208 items
 Cluster 1: 162 items
 Total number of items: 370

(b)

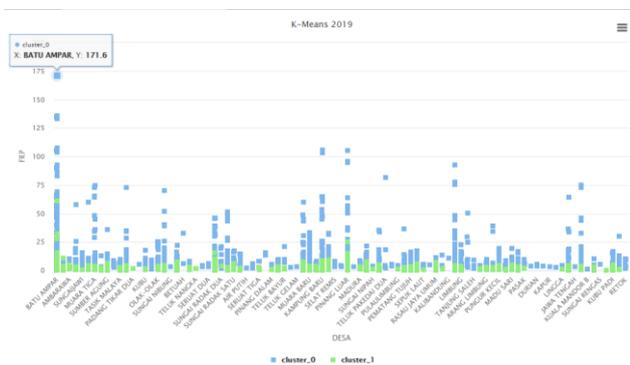
Cluster Model

Cluster 0: 1924 items
 Cluster 1: 1106 items
 Total number of items: 3030

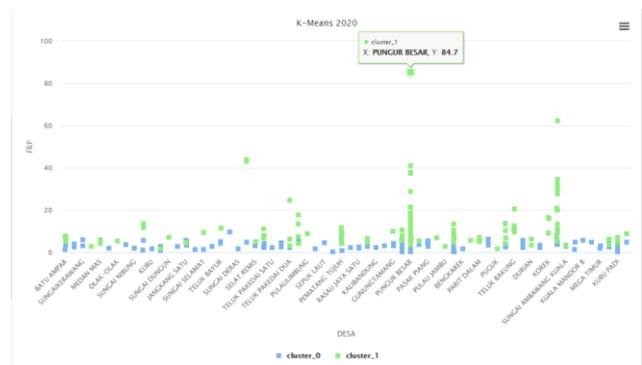
(c)

Gambar 3 Hasil Cluster Model K-Means

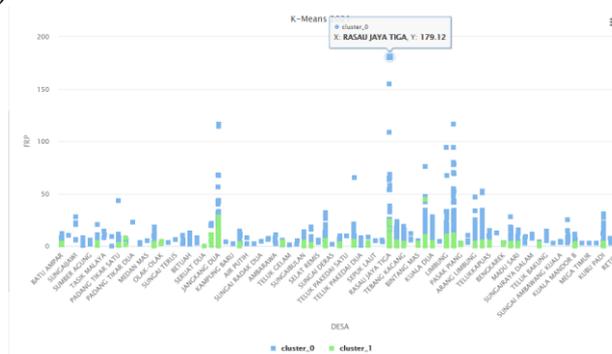
Hasil *cluster* model dari data tahun 2019, setelah dilakukan training pada data 3684 items terbentuk 2 cluster yaitu cluster 0 sebanyak 2229 items dan cluster 1 sebanyak 1455 items. Tahun 2020, setelah dilakukan training pada data 370 items terbentuk 2 cluster yaitu cluster 0 sebanyak 208 items dan cluster 1 sebanyak 162 items. Tahun 2021, setelah dilakukan training pada data 3030 items terbentuk 2 cluster yaitu cluster 0 sebanyak 1924 items dan cluster 1 sebanyak 1106 items.



(a)



(b)



(c)

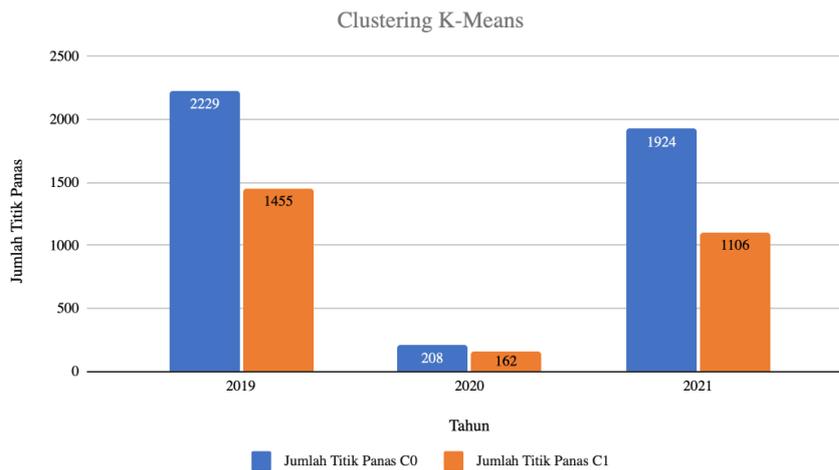
Gambar 1 (a) Visualisasi clustering tahun 2019 (b) Visualisasi clustering tahun 2020
 (c) Visualisasi clustering tahun 2021

Pada gambar diatas memperlihatkan data *clustering* titik panas berdasarkan nama desa dan FRP. Hasil *clustering* dibagi menjadi *clustering* 0 berwarna biru mengindikasikan desa sangat rawan dan *cluster* 1 berwarna hijau mengindikasikan desa rawan kebakaran hutan dan lahan. Gambar 4 (a) menunjukkan visualisasi *clustering* tahun 2019 daerah dengan frekuensi FRP tertinggi adalah Batu

Ampar pada cluster 0 yaitu termasuk areal rawan dengan tingkat frekuensi titik panas sangat tinggi. Gambar 4 (b) menunjukkan visualisasi *clustering* tahun 2020 daerah dengan frekuensi FRP tertinggi adalah Punggur Besar pada cluster 1 yaitu termasuk areal rawan dengan frekuensi titik panas tinggi. Gambar 4 (c) menunjukkan visualisasi *clustering* tahun 2021 daerah dengan frekuensi FRP tertinggi adalah Rasau Jaya Tiga pada *cluster* 0 yaitu termasuk areal rawan dengan tingkat frekuensi titik panas sangat tinggi.

Analisis Hasil *Clustering* metode K-Means

Berdasarkan *clustering* menggunakan metode *K-Means* ditentukan dengan menetapkan parameter $k = 2$ sebagai kelompok *clustering*. Terdapat 2 *cluster* dalam penelitian yaitu *cluster* 0 adalah areal rawan dengan tingkat frekuensi titik panas sangat tinggi dan *cluster* 1 adalah areal rawan dengan frekuensi titik panas tinggi. Jumlah titik panas berdasarkan tahun ditunjukkan pada Gambar 5.



Gambar 5 Jumlah titik panas per tahun

Berdasarkan grafik diatas terlihat tahun 2019 C0 terdeteksi 2229 titik panas dan C1 terdeteksi 1455 titik panas. Tahun 2020 C0 terdeteksi 208 titik panas dan C1 terdeteksi 162 titik panas. Tahun 2021 C0 terdeteksi 1924 titik panas dan C1 terdeteksi 1106 titik panas. Jumlah titik panas terbanyak terdeteksi pada tahun 2019 yaitu 3684 titik panas dan jumlah titik panas paling sedikit terdeteksi pada tahun 2020 yaitu 370 titik panas. Hal ini terlihat bahwa kebakaran hutan dan lahan paling banyak terjadi pada tahun 2019 dan areal rawan paling banyak.

Hasil evaluasi *clustering* menggunakan metode *K-Means* menggunakan nilai *Silhouette coefficient* ditunjukkan pada Tabel 1.

Tabel 1 Nilai *Silhouette Coefficient* menggunakan K-Means

Tahun	Cluster Ke-	Jumlah Titik Panas	<i>Silhouette coefficient</i>
2019	C0	2229	0.570
	C1	1455	0.580
2020	C0	208	0.588
	C1	162	0.501
2021	C0	1924	0.762
	C1	1106	0.750

Berdasarkan hasil tabel diatas, hasil *clustering* menggunakan metode K-Means tahun 2019 nilai *Silhouette coefficient* C1 yaitu 0,570 dan C2 yaitu 0,580 termasuk pada kategori *strong structure*. Tahun 2020 nilai *Silhouette coefficient* C1 yaitu 0,588 dan C2 yaitu 0,501 termasuk pada kategori *medium structure*. Tahun 2021 nilai *Silhouette coefficient* C1 yaitu 0,762 dan C2 yaitu 0,750 termasuk pada kategori *strong structure*. Nilai *Silhouette Coefficient* berada dalam rentang nilai $0,7 < SC \leq 1$ termasuk kategori *strong structure*. Sedangkan nilai *Silhouette Coefficient* berada dalam rentang nilai $0,5 < SC \leq 0,7$ termasuk kategori *Medium Structure*.

SIMPULAN

Berdasarkan hasil penelitian, metode K-Means terbukti efektif dalam mengelompokkan titik panas kebakaran hutan dan lahan. Pengolahan data dengan menggunakan metode K-Means cluster terbaik dan parameter $k=2$ menghasilkan nilai *Silhouette coefficient* tertinggi pada tahun 2021, yang menunjukkan adanya struktur yang kuat dalam pengelompokan tersebut. Kesimpulan ini menunjukkan bahwa metode K-Means dapat digunakan sebagai alat yang dapat membantu dalam pemantauan dan analisis kebakaran hutan dan lahan. Implikasinya, dengan menggunakan metode ini, kita dapat mengidentifikasi dan memahami pola serta tren kebakaran hutan dan lahan, yang dapat membantu dalam pengambilan keputusan dan upaya pencegahan serta penanggulangan kebakaran di masa depan.

REFERENSI

- D. Oleh, B. (2018). IRBI indeks risiko bencana indonesia. *Nasional, dan Bencana*.
- I. Sumariyanto & J. Hutauruk, (2021). Konsep pergelaran teknologi persenjataan batalyon arhanud dalam menangkal ancaman kalimantan barat. *Jurnal DEFENDONESIA*, 5(2), 7–14.
- M. Firman, A. Halik, & L. Septiana. (2022). Analisa Data Untuk Prediksi Daerah Rawan Bencana Alam Di Jawa Barat Menggunakan Algoritma K-Means *Clustering*. *Journal of Information System, Applied, Management, Accounting and Research*, 6(4), 856–870.
- Kementerian Lingkungan Hidup dan Kehutanan Republik Indonesia. (2019). SiPongi - Karhutla Monitoring Sistem. In [Www.Sipongi.Menlhk.Go.Id](http://www.Sipongi.Menlhk.Go.Id).
- Athifaturrofifah, Goejantoro, R., & Yuniarti, D. (2019). Perbandingan Pengelompokan K-Means dan K-Medoids Pada Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Studi Kasus : Data Titik Panas Di Indonesia Pada 28 April 2018). *Jurnal EKSPONENSIAL*, 10(2), 143–152.
- Shoolihah, A.-M., Furqon, M. T., & Widodo, A. W. (2017). Implementasi Metode Improved K-Means untuk Mengelompokkan Titik Panas Bumi. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya*, 1(11), 1270–1276.
- Hermawati, R., & Sitanggang, I. S. (2016). Web-Based Clustering Application Using Shiny Framework and DBSCAN Algorithm for Hotspots Data in Peatland in Sumatra. *Procedia Environmental Sciences*, 33, 317–323. <https://doi.org/10.1016/j.proenv.2016.03.082>
- Gunawan, W. (2021). Implementasi Algoritma DBScan dalam Pemngambilan Data Menggunakan Scatterplot. *Techno Xplore : Jurnal Ilmu Komputer Dan Teknologi Informasi*, 6(2), 91–98. <https://doi.org/10.36805/technoxplore.v6i2.1179>
- Handayani, F., Teguh, R., & Lestari, A. (2021). *PENDETEKSIAN POTENSIAL HOTSPOT DATA SATELIT CHECK ALGORITMA CLUSTERING K-MEANS*. 15(2), 164–173.
- Id, I. D., & Angraini, T. R. (2018). *Penentuan Daerah Rawan Titik Api di Provinsi Riau Menggunakan Clustering Algoritma K-Means (Determination of Fire Point Prone Areas in Riau Province Using Clustering K-Means Algorithm)*. VI(November), 137–147.

- Mardalius, M. (2018). Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 4(2), 123-132.
- Sari, R. W., Wanto, A., & Windarto, A. P. (2018). Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi). *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 2(1).
- Rolansa, F., Yunita, Y., & Suheri, S. (2020). Sistem prediksi dan evaluasi prestasi akademik mahasiswa di Program Studi Teknik Informatika menggunakan data mining. *Jurnal Pendidikan Informatika dan Sains*, 9(1), 75-85.
- Maulana, A., Akbar, K. N., & Nurahman, N. (2021). Penerapan Clustering Menggunakan Algoritma K-Means Sebagai Analisis Produksi Komoditas Perikanan Provinsi di Indonesia. *EJECTS: E-Journal Computer, Technology and Informations System*, 1(1).
- Serliani, K., Mustofa, Y. A., & Kumala, I. S. (2020). Clustering Petani Penerima Pupuk Berdasarkan Luas Lahan Menggunakan Algoritma K-Means. *Jurnal Cosphi*, 4(2).