

Penerapan Algoritma DBSCAN Untuk Clustering Penjualan Supermarket

Talitha Naifa Audrey¹, Muhammad Amiruddin W.²,
Mitha Islamiyah³, Rinci Kembang Hapsari^{4*}

^{1,2,3,4}*Prodi Teknik Informatika, Fakultas Teknik Elektro dan Teknologi Informasi, Institut Teknologi Adhi Tama Surabaya, Indonesia*

* Penulis Korespondensi. Email: rincikembang@itats.ac.id

ABSTRAK

Supermarket adalah tempat perbelanjaan dimana menyediakan bermacam-macam bahan kebutuhan sehari-hari. Di supermarket banyak customer yang membeli barang kebutuhannya. Pertumbuhan supermarket semakin meningkat. Supermarket mempunyai berbagai macam produk yang berbeda merek, bermacam cabang dan berbagai tipe pelanggan. Untuk membuat strategi penjualan perlu diketahui produk-produk yang diminati customer. Dalam penelitian ini dilakukan klusterisasi produk supermarket berdasarkan dari data penjualan. Algoritma klusterisasi dalam penelitian ini adalah algoritma *Density-Based Spatial Clustering with Noise* (DBSCAN). Algoritma ini adalah sebuah algoritma pengelompokan obyek data berdasarkan densitas atau kepadatan, yang dipengaruhi oleh parameter input, yaitu nilai parameter Eps yang merupakan ukuran jarak yang akan digunakan untuk menemukan titik-titik di sekitar titik mana pun dan nilai parameter MinPts yang merupakan jumlah titik minimum yang dikelompokkan bersama agar suatu kawasan dianggap padat. Dalam penelitian ini Dataset yang dipakai adalah data sekunder, yang berisi 100 data penjualan supermarket, dengan mengambil 2 atribut. Hasil klusterisasi menunjukkan dengan menggunakan nilai parameter Eps = 6 dan nilai MinPts = 9 data produk barang terbagi menjadi 3 cluster, Dimana cluster 1 beranggotakan produk yang tidak diminati atau jarang dibeli customer, Cluster 2 beranggotakan data barang yang diminati customer, dan cluster 3 beranggotakan produk yang sangat diminati.

Kata Kunci: Supermarket, Penjualan, DBSCAN, Indeks, clustering

ABSTRACT

Supermarkets are shopping places that provide a variety of daily necessities. In supermarkets, many customers buy the goods they need. The growth of supermarkets is increasing. Supermarkets have various products with different brands, branches, and types of customers. To create a sales strategy, you need to know the products customers are interested in. In this research, supermarket product clustering was carried out based on sales data. The clustering algorithm in this research is the DBSCAN algorithm. This algorithm groups data objects based on density, which is influenced by input parameters, namely the Eps and MinPts values. In this research, the dataset used is secondary data, which contains 100 supermarket sales data, taking two attributes. The clustering results show that by using the Eps parameter value = 6 and the MinPts value = 9, the product data is divided into 3 clusters, where cluster 1 consists of products that are not in demand or are rarely purchased by customers, Cluster 2 consists of data on goods that customers are interested in, and cluster 3 consists of products that are not in demand or are rarely purchased by customers. Products that are in great demand.

Keyword: Supermarkets, Sales, DBSCAN, Index, clustering

Article info: Submitted: 25 May 2024

Accepted: 29 November 2024

How to cite this article:

Audrey, T. N., Widodo, M.A., Islamiyah, M., & Hapsari, R. K. (2024). Penerapan Algoritma DBSCAN Untuk Clustering Penjualan Supermarket. Zeta - Math Journal, 9(2), 55-59. <https://doi.org/10.31102/zeta.2024.9.2.83-89>



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](#).

1. PENDAHULUAN

Perkembangan teknologi informasi yang semakin pesat majunya dimana setiap waktu ada kemajuan teknologi. Perkembangan ini berperan penting di dalam aspek kehidupan manusia. Dimana aspek perekonomian adalah salah satu aspek yang tidak bisa menghindar dari teknologi informasi, terutama pada sistem penjualan.

Semakin ketatnya persaingan di dunia bisnis khususnya pada industri supermarket, menurut para ahli untuk memprediksi sebuah pola yang bisa meningkatkan transaksi penjualan ataupun pemasaran barang di suatu supermarket, dengan menggunakan manfaat data transaksi (Adha et al., 2017). Penggunaan data yang terdapat pada sistem informasi, melibatkan sumber daya pada sebuah organisasi, data, prosedur, serta sarana pendukung dalam menunjang kegiatan penentuan keputusan, tidak hanya sekedar menggantungkan pada data transaksi operasional, tetapi diperlukan juga suatu analisis data untuk menggali informasi implisit yang ada (Nur, 2019).

Informasi yang telah ada bisa dibuat sebagai dasar sistem penentuan keputusan dalam penyelesaian permasalahan bisnis serta support infrastruktur pada bidang teknologi, merupakan pencetus berkembangnya teknologi dalam data mining (Siregar, 2018). Data mining diperlukan untuk memperoleh informasi, trend, pola atau interaksi dalam big data yang dapat dimanfaatkan sebagai dasar untuk menentukan sebuah keputusan.

Pada data mining proses pengelompokkan data terdapat beberapa teknik. Analysis clustering memilah kumpulan data menjadi beberapa kategori dengan kesamaan keunikan atau perhitungan jarak (Hardiani, 2022). Sebagian besar research terkini menggunakan teknik clustering di karenakan sifat dan karakteristik data yang belum diketahui karakteristik data tersebut. Metode clustering memiliki perbedaan, pada clustering tidak mempunyai variabel target dalam melakukan pengelompokkan (Sindi et al., 2020).

Pada penelitian lainnya yang membahas mengenai metode clustering, terdapat berbagai macam pembahasan mengenai penggunaan metode clustering DBSCAN. Metode DBSCAN digunakan untuk seleksi penetapan mahasiswa yang berhak mendapatkan beasiswa dari Yayasan dengan menggunakan 10 data mahasiswa Yayasan universitas Budi Darma pada Angkatan tahun 2019. Dengan DBSCAN didapatkan hasil pengelompokkan 2 cluster dan tidak terdapat data noise karena semua titik data masuk kedalam kluster (Indini et al., 2022). Penelitian serupa yang membahas pengelompokkan produksi padi ladang dan padi sawah di wilayah provinsi Jawa Tengah, didapatkan hasil pengelompokkan 3 cluster dan tidak terdapat noise (Safitri et al., 2017). Penerapan metode DBSCAN dalam pertanian yang lain, adalah proses clustering pada lahan pertanian padi di Kabupaten Karawang (Sari & Primajaya, 2019). Metode DBSCAN juga digunakan dalam proses penentuan keputusan dalam membantu perusahaan menentukan pelanggan yang potensial (Devi et al., 2015).

Dalam penelitian ini algoritma DBSCAN clustering diimplementasikan untuk analisa data transaksi penjualan. Dengan mengimplementasikan algoritma ini dataset yang telah didapatkan dikelompokkan menjadi beberapa klaster berdasarkan kesamaan dan ketidaksamaan dari dataset tersebut. Hasil pengelompokkan difokuskan untuk mengetahui popularitas penjualan barang di supermarket berdasarkan cluster yang karakteristiknya sama (Ashari et al., 2019).

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining merupakan proses mengkolek serta pengolahan dataset dengan tujuan untuk menggali informasi penting dalam data, yang mana merupakan bidang interdisiplin yang merupakan bagian dari proses menemukan pengetahuan yang lebih besar (Meilani et al., 2021). Data mining memanfaatkan teknik-teknik yang terdapt dalam ilmu matematika, statistika, kecerdasan buatan serta *machine learning* untuk menghasilkan dan mengidentifikasi informasi serta pola-pola yang berasal dari dataset yang sangat besar (Sibuea & Safta, 2017).

Data mining merupakan proses untuk mendapatkan korelasi, memprediksi tren dimasa datang, dan pola baru yang memiliki makna dengan mengolah dataset dengan jumlah yang besar yang tersimpan pada repositori,

dengan memanfaatkan teknologi pengenalan pola, teknik statistik serta matematika (Zulfa et al., 2021). Dengan melakukan pengenalan pola informasi pada data yang berfungsi untuk memanipulasi data membentuk informasi baru yang lebih bermanfaat (Indini et al., 2022), (Utomo & Purba, 2019).

2.2. Clustering

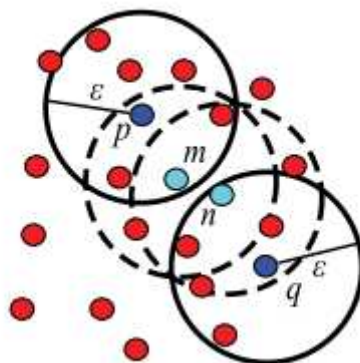
Teknik clustering adalah sebuah cara untuk mengelompokkan data berdasarkan kesamaan dan juga perbedaan yang terdapat di antara dataset tersebut. Cluster merupakan kumpulan dari data yang mempunyai kemiripan satu dengan yang lain, dan berbeda dengan data di cluster lain. Tujuan dari clustering adalah membagi dataset menjadi kluaster-kluaster di mana data dalam satu klaster memiliki karakteristik yang mirip dan juga berbeda terhadap data di klaster lain.

Dalam clustering, data dikelompokkan berdasarkan kemiripan dan perbedaan antar data, sehingga data dalam satu klaster memiliki tingkat kesamaan yang tinggi dan data antar klaster mempunyai kesamaan yang rendah (Devi et al., 2015). Clustering adalah salah satu teknik fungsionalitas data mining, dimana algoritma clustering adalah algoritma pengelompokkan dataset menjadi kluster-kluster data tertentu (Ashari et al., 2019).

2.3. Clustering berbasis Densitas

Clustering berdasarkan densitas merupakan suatu Teknik pengelompokkan data dimana mempartisi data menjadi kelompok-kelompok yang didasarkan pada kepadatan titik data. Pengelompokan berbasis kepadatan bergantung pada gagasan menemukan kepadatan suatu wilayah (Bhattacharjee & Mitra, 2021). Dalam Teknik ini, data yang mempunyai kepadatan yang sama dianggap sebagai bagian dari cluster yang sama. Sedangkan data yang memiliki kepadatan yang berbeda dianggap sebagai bagian dari kelompok yang berbeda. Dimana kepadatan didapatkan dengan menghitung sebagai rata-rata jarak antara titik data dalam suatu kelompok.

Tujuan klasterisasi berbasis densitas adalah untuk menemukan cluster pada tingkat granularitas yang berbeda dengan penyaringan noise yang sesuai. Dalam klasterisasi ini, cluster diidentifikasi sebagai area dengan kepadatan lebih tinggi dibandingkan ruang data lainnya. Klasterisasi ini juga memfasilitasi pendeteksian cluster dengan bentuk yang berubah-ubah (Bhattacharjee & Mitra, 2021).



Gambar 1. Konektivitas densitas (Kriegel et al., 2011)

Metode ini bertujuan untuk mengekstraksi kelompok dengan kepadatan yang relatif seragam yang terletak di seluruh ruang data. Algoritma yang berdasarkan densitas antara lain yaitu algoritma DBCLASD, DENCLUE, DBSCAN (Density Based Spatial Clustering of Applications with Noise), OPTICS, VDBSCAN, DVBSCAN, dan ST-DBSCAN (Bhuyan & Borah, 2013).

3. METODE PENELITIAN

3.1 Dataset

Dataset yang digunakan pada penelitian ini, yaitu data publik yang diambil dari situs Kaggle, diunggah oleh AUNG PYAE yang telah diperbarui pada 23 November 2019 dan data dengan format .csv. Dataset yang digunakan terdiri dari 1000 data pengunjung. Dataset terdiri dari 11 atribut, yaitu Branch, City, Customer type, Cogs, Date, Gender, Quantity, Gross Income, Tax 5%, Gross margin percentage, Invoice ID, Total, Payment, Product line, Rating, Time, dan Unit price.

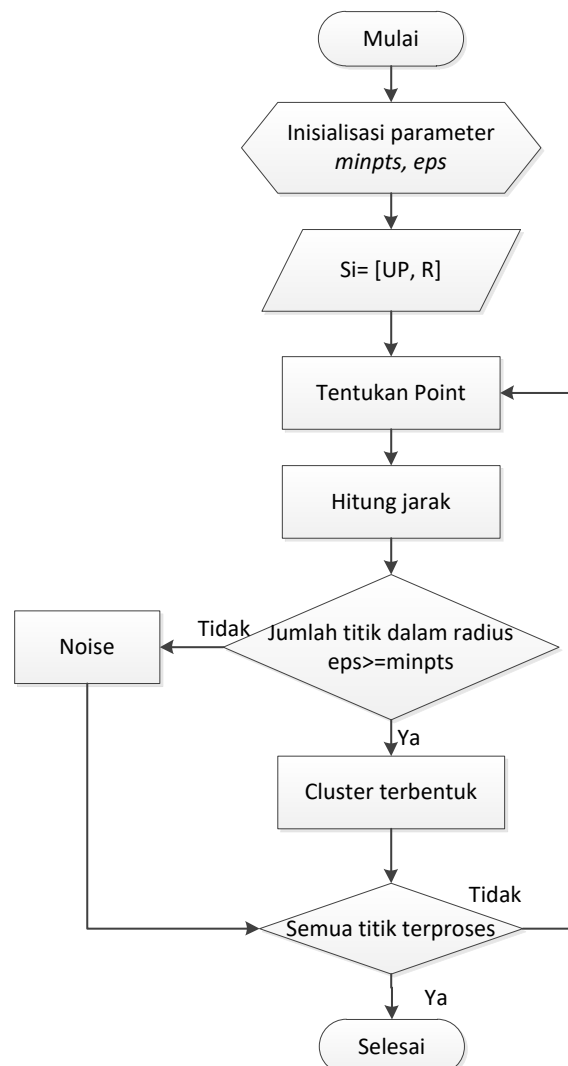
Pada dataset dilakukan reduksi data yang dilakukan secara random dengan memilih 100 data dan dilakukan seleksi fitur dengan memilih dua atribut dari 11 atribut, yaitu unit price dan rating, ditunjukkan pada Tabel 1.

Tabel 1. Dataset awal

Titik	Unit Price	Rating
0	74.69	9.1
1	15.28	9.6
2	46.33	7.4
3	58.22	8.4
...	...	...
97	12.45	4.1
98	52.75	9.3
99	82.7	7.4

3.2 Algoritma DBSCAN

Algoritma DBSCAN (*Density-based Spatial Clustering of Application with Noise*) adalah algoritma pengelompokan yang *density-based* (berbasis kepadatan), dimana mengelompokkan data berdasarkan tingkat kerapatan data. Algoritma ini dapat mendeteksi cluster yang memiliki bentuk dan ukuran berbeda dan juga bisa mendeteksi noise pada dataset. Algoritma ini juga tidak akan memasukkan noise pada cluster manapun (Ashari et al., 2019).



Gambar 2. Flowchart DBSCAN

Berdasarkan Gambar 2 menunjukkan langkah-langkah dari algoritma DBSCAN. Dengan input dataset S_i dengan parameter UP (*Unit Price*) dan R (*Rating*). Diman detail langkahnya ditunjukkan sebagai berikut:

1. Menentukan parameter Eps dan parameter $MinPts$ (Jumlah minimal data dalam radius). Dalam penelitian ini nilai $Eps=6$ dan $MinPts = 9$.
2. Menentukan titik awal (*core*) dengan bebas/random.
3. Melakukan pengukuran jarak titik awal dengan setiap titik dalam dataset, menggunakan persamaan jarak, rumus *Euclidean Distance*.

$$\text{Jarak} = \sqrt{(x - xp)^2 + (y - yp)^2} \quad (1)$$

Dengan,

- x : nilai atribut X titik tujuan
 - y : nilai atribut Y titik tujuan
 - x_p : nilai pusat atribut x
 - y_p : nilai pusat atribut y
4. Mengambil semua titik dimana kepadatan dapat dijangkau (*density reachable*) dari titik awal yang memenuhi nilai Eps dan nilai $MinPts$. Jika titik awal merupakan *core point*, maka cluster terpenuhi.
 5. Mengulangi langkah tiga (3) dan Langkah empat (4) hingga semua titik telah diproses. Apabila titik awal merupakan *border point* dan tidak memiliki hubungan kepadatan dapat dijangkau (*density reachable*) terhadap titik awal maka berikutnya akan memilih titik data lainnya pada database.

Dalam menetapkan nilai Eps dan nilai $MinPts$ secara optimal ditentukan dengan menggunakan *dist graph*. Dengan memperhatikan perubahan nilai Eps dari nilai k yang bermacam-macam. Pada titik yang mengalami perubahan tajam pada nilai k -dist digunakan menjadi nilai Eps dan nilai k merupakan nilai dari $MinPts$ (Rohalidyawati et al., 2020).

4. HASIL DAN PEMBAHASAN

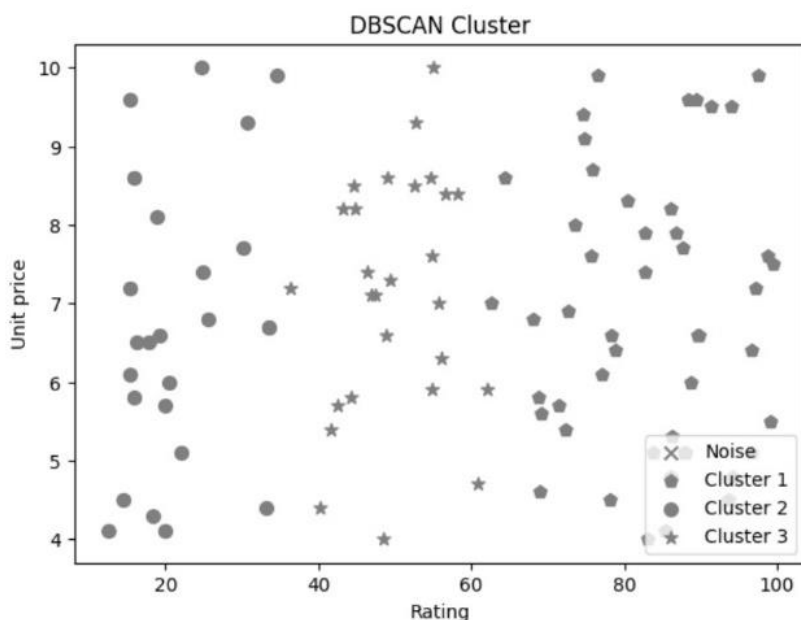
Berdasarkan penelitian yang telah dilakukan, dengan melakukan seleksi fitur, sehingga pada dataset transaksi penjualan, diambil dua atribut yaitu atribut harga per barang (*unit price*) dan atribut rating penjualan. Jumlah data yang digunakan sebanyak 100 data penjualan supermarket pada tahun 2019. Dimana hasil pengelompokan data penjualan di supermarket, disajikan dalam Tabel 2.

Tabel 2. Dataset yang telah dikelompokkan

Id	Unit Price	Rating	Cluster
0	74.69	9.1	1
34	99.42	7.5	1
35	68.12	6.8	1
...	...	...	...
22	33.2	4.4	2
69	18.33	4.3	2
44	15.37	7.2	2
...	...	...	...
25	52.59	8.5	3
8	36.26	7.2	3
72	48.52	4	3

Berdasarkan hasil pengelompokan yang tercantum pada Tabel 2, dapat terlihat bahwa dataset terbagi menjadi tiga (3) cluster dan tidak terdapat noise data. Pengelompokan data berdasarkan popularitas yang didapatkan dari penilaian pembeli berdasarkan unit price dan rating.

Hasil penilaian dengan numerik 1 yang artinya cluster yang paling kurang diminati, 2 yang artinya cluster yang diminati dan untuk 3 dengan arti cluster yang sangat diminati oleh pembeli.



Gambar 3. Hasil Clustering DBSCAN

Dengan mengimplementasikan algoritma DBSCAN didapatkan sebaran anggota setiap cluster yang ditunjukkan pada Gambar 3. Hasil pengelompokan dari dataset yang sudah dikelola dengan menggunakan metode DBSCAN tersaji pada Tabel 3.

Tabel 3. Pengelompokan Algoritma DBSCAN

Cluster	Jumlah Data
1	48
2	25
3	27

5. KESIMPULAN

Penelitian ini telah berhasil mengimplementasikan metode analisis mengenai Algoritma DBSCAN menunjukkan performa yang baik dalam proses clustering dengan data yang besar pada data penjualan di supermarket. Ditemukan faktor penentu popularitas penjualan barang di supermarket adalah harga per barang dan rating penjualan dari pembeli. Dari hasil pengujian, diperoleh kesimpulan pengujian penggunaan algoritma DBSCAN pada data penjualan di supermarket dalam empat tahun terakhir berhasil dilakukan dalam mendapatkan cluster. Dengan terbentuknya 3 cluster, untuk cluster 1 sebanyak 48 data dengan popularitas kurang diminati, cluster 2 sebanyak 25 data dengan popularitas diminati, dan cluster 3 sebanyak 27 data dengan popularitas sangat diminati.

UCAPAN TERIMA KASIH

Kami mengucapkan terimakasih kepada Bapak Pembina YPTS, Bapak Ketua YPTS, Bapak Pimpinan Institut Teknologi Adhi Tama Surabaya karena telah mensupport kami. Sehingga kolaborasi penelitian dosen dan mahasiswa dapat diselesaikan dan dipublikasikan.

DAFTAR PUSTAKA

- Aadha, N., Sianturi, L. T., & Siagian, E. R. (2017). IMPLEMENTASI DATA MINING PENJUALAN SABUN DENGAN MENGGUNAKAN METODE APRIORI (Studi Kasus : PT. Unilever). *Majalah Ilmiah INTI*, 12(2), 219–223.
- Ashari, B. S., Otniel, S. C., & Rianto. (2019). Perbandingan Kinerja K-Means Dengan DSCAN Untuk Metode Clustering Data Penjualan Online Retail. *Jurnal Siliwangi*, 5(2), 72–77. <http://jurnal.unsil.ac.id/index.php/jssainstek/article/view/1283>
- Bhattacharjee, P., & Mitra, P. (2021). A survey of density based clustering algorithms. *Frontiers of Computer Science*, 15(1). <https://doi.org/10.1007/s11704-019-9059-3>
- Bhuyan, R., & Borah, S. (2013). A survey of some density based clustering techniques. *Conference on Advancements in In-Formation, Computer and Communication*, July. <https://doi.org/10.13140/2.1.4554.6887>
- Devi, A. S., Putra, I. K. G. D., & Sukarsa, I. M. (2015). Implementasi Metode Clustering DBSCAN pada Proses Pengambilan Keputusan. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 6(3), 185. <https://doi.org/10.24843/lkjiti.2015.v06.i03.p05>
- Hardiani, T. (2022). Analisis Clustering Kasus COVID-19 Di Indonesia Menggunakan Algoritma K-Means. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 11(2), 156–165. <https://doi.org/10.23887/janapati.v11i2.45376>
- Indini, D. P., Siburian, S. R., & Utomo, D. P. (2022). Implementasi Algoritma DBSCAN untuk Clustering Seleksi Penentuan Mahasiswa yang Berhak Menerima Beasiswa Yayasan. 325–331.
- Kriegel, H. P., Kröger, P., Sander, J., & Zimek, A. (2011). Density-based clustering. *Advanced Review*, 1(3), 231–240. <https://doi.org/10.1002/widm.30>
- Meilani, B. D., Hapsari, R. K., & Novian, I. F. (2021). Classification of community opinion on the use of the Transjakarta bus based on twitter social network using naïve bayes method. *IOP Conference Series: Materials Science and Engineering*, 1010(1). <https://doi.org/10.1088/1757-899X/1010/1/012030>
- Nur, H. (2019). Penggunaan Metode Waterfall Dalam Rancang Bangun Sistem Informasi Penjualan. *Generation Journal*, 3(1), 1. <https://doi.org/10.29407/gj.v3i1.12642>
- Rohalidyawati, W., Rahmawati, R., & Mustafid, M. (2020). Segmentasi Pelanggan E-Money Dengan Menggunakan Algoritma Dbscan (Density Based Spatial Clustering Applications With Noise) Di Provinsi Dki Jakarta. *Jurnal Gaussian*, 9(2), 162–169. <https://doi.org/10.14710/j.gauss.v9i2.27818>
- Safitri, D., Wuryandari, T., & Rahmawati, R. (2017). Metode Dbscan Untuk Pengelompokan Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Produksi Padi Sawah Dan Padi Ladang. *Statistika*, 5(1), 8–13.
- Sari, B. N., & Primajaya, A. (2019). Penerapan Clustering Dbscan Untuk Pertanian Padi Di Kabupaten Karawang. *Jurnal Informatika Dan Komputer*, 4(1), 28–34.
- Sibuea, M. L., & Safta, A. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering. *Jurteks*, 4(1), 85–92. <https://doi.org/10.33330/jurteks.v4i1.28>
- Sindi, S., Ningse, W. R. O., Sihombing, I. A., R.H.Zer, F. I., & Hartama, D. (2020). Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia. *Jurnal Teknologi Informasi*, 4(1), 166–173. <https://doi.org/10.36294/jurti.v4i1.1296>
- Siregar, M. H. (2018). Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan). *Jurnal Teknologi Dan Open Source*, 1(2), 83–91. <https://doi.org/10.36378/jtos.v1i2.24>
- Utomo, D. P., & Purba, B. (2019). Penerapan Datamining pada Data Gempa Bumi Terhadap Potensi Tsunami di Indonesia. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1(September), 846. <https://doi.org/10.30645/senaris.v1i0.91>
- Zulfa, N., Auliya, R. I., & Zaenal, A. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, 2(2), 100. <http://jim.teknokrat.ac.id/index.php/JTSI>