

Segmentasi Pelanggan dengan Algoritma DBSCAN dan K-Means: Mengungkap Pola dalam Analisis Kepribadian Pelanggan

Nabila Dea Azahra^{a,*}, Zulfa Laela Maulida^a, Muhammad Faza Farizqi^a, Iqbal Kharisudin^a

^a Statistika dan Sains Data, Universitas Negeri Semarang, Semarang 50229, Indonesia

* Alamat Surel: nabiladeaaz@students.unnes.ac.id

Abstrak

Dalam pemasaran, analisis kepribadian pelanggan adalah pendekatan yang digunakan untuk memahami karakteristik dan perilaku pelanggan. Penelitian ini mengidentifikasi segmentasi pelanggan dengan menerapkan algoritma *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) dan K-Means pada data *marketing campaign*. Sebagai metode klusterisasi berbasis kepadatan, DBSCAN mampu mendeteksi kelompok pelanggan yang memiliki karakteristik serupa tanpa memerlukan penetapan banyak kelompok sebelumnya. DBSCAN sering diterapkan pada data yang banyak mengandung *noise*, hal ini dikarenakan DBSCAN tidak akan memasukkan data yang dianggap *noise* ke dalam kluster manapun. Melalui analisis data demografis dan perilaku pembelian, penelitian ini berhasil mengungkap pola-pola tersembunyi di antara pelanggan dan mengidentifikasi pelanggan yang tidak terstruktur. Penelitian ini menunjukkan algoritma DBSCAN efektif dalam segmentasi pelanggan berdasarkan pola pembelian daging, buah, dan emas, dengan parameter optimal ($\epsilon = 0,4$, $\text{minPts} = 3$). DBSCAN membentuk empat kluster utama, mengidentifikasi data *noise*, dan mengungkap pola konsumsi unik, memberikan wawasan strategis untuk pemasaran.

Kata kunci:

Segmentasi Pelanggan, DBSCAN, Klusterisasi Non-Hierarki, R Programming

© 2025 Dipublikasikan oleh Jurusan Matematika, Universitas Negeri Semarang

1. Pendahuluan

Segmentasi dengan analisis kepribadian pelanggan merupakan salah satu strategi utama dalam pemasaran yang digunakan untuk memahami dan melayani pelanggan dengan lebih efektif (Saliano et al., 2021). Dengan pendekatan ini, perusahaan dapat membagi pasar yang luas menjadi kelompok-kelompok kecil berdasarkan kesamaan perilaku, preferensi, dan kebutuhan (Faza et al., 2024). Strategi yang efisien dan tepat sasaran, serta peningkatan peluang pelanggan tetap setia dan menjadikan lebih banyak orang membeli bisa dilakukan oleh perusahaan (Djogo, 2022). Dalam konteks persaingan yang semakin ketat, segmentasi pelanggan yang akurat menjadi semakin krusial bagi keberhasilan perusahaan (Anom & Ali, 2024).

Salah satu metode segmentasi pelanggan yang mulai banyak digunakan saat ini adalah DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) (Miftahurrahmi et al., 2024). Algoritma ini dirancang untuk menemukan pola atau kluster pada data yang padat dan heterogen, bahkan ketika terdapat *noise* (Safitri et al., 2017). DBSCAN memiliki keunggulan yaitu tidak memerlukan penentuan jumlah kluster di awal, sehingga cocok digunakan pada data pelanggan yang kompleks (Kristianto, 2022). Dengan metode ini, perusahaan dapat mengidentifikasi kelompok pelanggan berdasarkan pola kepribadian mereka yang tersembunyi dan mengidentifikasi pelanggan yang tidak terstruktur (Fadila & Voutama, 2023). Hasil penelitian dapat memberikan wawasan yang mendalam ketika menyusun strategi pemasaran yang lebih efektif dan memberikan penawaran produk yang lebih relevan.

Seiring berkembangnya teknologi digital dan melimpahnya data pelanggan memberikan peluang bagi perusahaan untuk memahami perilaku konsumen dengan lebih baik. Namun, tantangan utamanya

To cite this article:

Azahra, N. D., Maulida, Z. L., Farizqi, M. F., & Kharisudin, I. (2025). Segmentasi Pelanggan dengan Algoritma DBSCAN dan K-Means: Mengungkap Pola dalam Analisis Kepribadian Pelanggan. *PRISMA, Prosiding Seminar Nasional Matematika* 8, 370-376

adalah memanfaatkan data yang tidak terstruktur, penuh *noise*, dan memiliki distribusi tidak seragam (Hajari et al., 2024). Algoritma segmentasi tradisional seperti K-Means sering tidak optimal dalam menangani data seperti ini. Oleh karena itu, penelitian ini menggunakan algoritma DBSCAN yang lebih fleksibel, mampu mendeteksi cluster dengan kepadatan tinggi dan mengabaikan data *noise* (Mutiah et al., 2024). Hal ini membantu perusahaan mengidentifikasi pelanggan potensial dan merancang strategi pemasaran yang lebih efektif.

Berdasarkan latar belakang tersebut, penelitian ini berfokus pada beberapa pertanyaan utama: (1) bagaimana algoritma DBSCAN dapat digunakan untuk segmentasi pelanggan yang lebih akurat; (2) mengapa DBSCAN lebih unggul dibandingkan metode segmentasi lain, khususnya dalam menangani data dengan *noise*; dan (3) pola kepribadian pelanggan apa saja yang dapat diungkap melalui segmentasi berbasis DBSCAN, serta bagaimana pola-pola ini dapat digunakan untuk menyusun strategi pemasaran yang lebih efektif bagi perusahaan.

Penelitian ini bertujuan untuk menganalisis dan menerapkan algoritma DBSCAN dalam segmentasi pelanggan berbasis data yang tidak terstruktur, serta mengungkap pola kepribadian pelanggan yang tersembunyi melalui pendekatan tersebut (Pratama, 2023). Dengan menggunakan DBSCAN, diharapkan perusahaan dapat memperoleh wawasan yang lebih dalam tentang perilaku pelanggan, yang nantinya dapat digunakan untuk menyusun rekomendasi strategi pemasaran yang lebih efektif. Secara keseluruhan, proyek ini bertujuan untuk mengevaluasi apakah penerapan DBSCAN pada data pelanggan dapat menghasilkan klaster yang bermakna dan relevan untuk pengambilan keputusan bisnis (Rohalidyawati et al., 2020)-(Aazmaan Ahmed Khan, 2023).

Penelitian ini juga berkontribusi pada pengembangan metodologi segmentasi pelanggan dalam pemasaran, khususnya dengan menawarkan solusi yang lebih canggih untuk mengatasi data yang kompleks dan tidak terstruktur. Dengan demikian, penelitian ini tidak hanya bermanfaat bagi praktisi pemasaran, tetapi juga bagi peneliti yang ingin mengembangkan lebih lanjut metode segmentasi berbasis data.

2. Metode

Clustering adalah teknik pembelajaran tanpa pengawasan yang mengelompokkan sekumpulan objek sedemikian rupa sehingga objek dalam kelompok atau klaster yang sama lebih mirip satu sama lain dibandingkan dengan objek dalam kelompok lain (Faza et al., 2024). Ada berbagai pendekatan untuk *clustering*, termasuk metode hierarki, partisional, berbasis grid, berbasis kepadatan, dan berbasis model, masing-masing dengan serangkaian teknik dan aplikasinya sendiri (Shetty & Singh, 2021).

Metodologi *clustering* hierarki mengkategorikan titik data dengan memisahkan entitas secara berulang menggunakan strategi aglomeratif (*bottom-up*) atau divisif (*top-down*), yang juga merupakan dua varian utama dari metodologi hierarki (Lutfi Hamzah, 2016). Sedangkan, *clustering* partisi berbeda dari metodologi pengelompokan hierarki, menetapkan data ke dalam sejumlah klaster yang telah ditetapkan dengan mengoptimalkan beberapa fungsi kriteria, tanpa struktur hierarki apa pun (Ismail et al., 2022). Metrik jarak Euclidean adalah kriteria yang paling umum digunakan, yang mengidentifikasi jarak minimum antara titik data dan setiap klaster yang tersedia, kemudian menetapkan titik tersebut ke klaster masing-masing. Beberapa algoritma *clustering* partisi yang terkenal termasuk K-means, K-Medoids, DBSCAN, dan pengelompokan spectral (Pribadi et al., 2022).

Dataset yang digunakan mencakup 2.240 data pelanggan dari dataset *marketing campaign*. Data ini berisi informasi demografis (seperti usia, tingkat pendidikan, dan status pernikahan), kebiasaan berbelanja, serta jumlah pengeluaran untuk berbagai kategori produk. Analisis diarahkan pada pola pembelian pelanggan untuk tiga kategori utama: produk daging, buah, dan emas. Penelitian ini difokuskan untuk mengeksplorasi hubungan antar variabel tersebut dalam membentuk pola pembelian pelanggan.

2.1 Pra-pemrosesan Data

Tahap ini mencakup pengumpulan data transaksi pelanggan yang mencerminkan pola pembelian mereka. Setelah data terkumpul, langkah berikutnya adalah memastikan kualitas data melalui proses pembersihan data (*data cleaning*) seperti mengatasi data yang hilang (*missing values*) atau pencilan. Selain itu, dilakukan transformasi data jika diperlukan, seperti normalisasi atau standarisasi, untuk memastikan data sesuai dengan algoritma DBSCAN.

2.2 Analisis Model DBSCAN

Dalam penelitian ini, digunakan instrumen dengan beberapa *library* utama pada *software* R untuk mendukung analisis data secara komprehensif. *Library* ‘tidyverse’ digunakan untuk manipulasi dan eksplorasi data, ‘ggplot2’ untuk menghasilkan visualisasi data, ‘dbscan’ untuk penerapan algoritma DBSCAN, ‘cluster’ dan ‘plotly’ untuk analisis cluster dan visualisasi 3D hasil clustering, serta ‘lubridate’ untuk pengolahan data waktu.

Proses analisis dimulai dengan menentukan dua parameter utama dalam algoritma DBSCAN, yaitu: (1) Epsilon (ϵ), yang menentukan jarak maksimum antar titik data untuk mengelompokkan data ke dalam cluster yang sama. Parameter ini menentukan seberapa besar wilayah atau radius sebuah cluster dapat terbentuk; (2) Minimum Points (minPts), yang merupakan jumlah minimum titik data yang diperlukan untuk membentuk sebuah cluster. Dalam penelitian ini, ϵ diatur pada 0.4, dan minPts diatur pada 3. Penentuan nilai optimal ϵ dan minPts menggunakan analisis k-distance plot yang membantu menentukan nilai-nilai ini berdasarkan distribusi jarak antar titik dalam data.

Setelah parameter ditentukan, algoritma DBSCAN diterapkan pada data clustering yang mencakup tiga variabel utama yaitu jumlah pembelian produk: (1) daging, (2) buah, dan (3) emas. Data ini telah melalui tahap pra-pemrosesan yang mencakup transformasi skala untuk memastikan bahwa setiap variabel memiliki bobot yang seimbang. Hasil dari penerapan DBSCAN ini menghasilkan pengelompokkan data ke dalam beberapa cluster berdasarkan kepadatan data dan mengidentifikasi outlier sebagai entitas terpisah.

Untuk memahami lebih lanjut hasil clustering, dilakukan visualisasi 3D yang menampilkan distribusi cluster berdasarkan variabel pembelian produk dan memberikan gambaran yang lebih jelas tentang karakteristik setiap cluster. Sebagai perspektif awal, dilakukan juga visualisasi scatter plot 2D untuk memvisualisasikan pola hubungan antar variabel sebelum beralih ke analisis 3D.

3. Hasil dan Pembahasan

Dataset yang digunakan dalam penelitian ini terdiri dari 2.208 observasi dan 9 variabel utama setelah melalui proses pra-pemrosesan. Variabel utama yang dianalisis mencakup total pembelian produk daging (MntMeatProducts), buah (MntFruits), dan emas (MntGoldProds). Selain itu, data pelanggan juga dilengkapi dengan informasi tambahan seperti tingkat pendidikan (Education), status pernikahan (Marital_Status), dan pendapatan (Income).

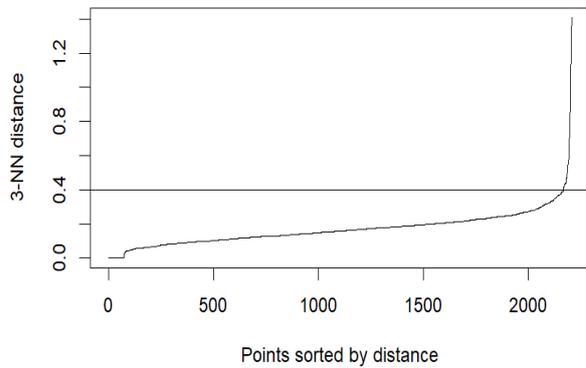
Pra-pemrosesan data dilakukan dengan langkah-langkah berikut: (1) Penyaringan data berdasarkan usia pelanggan, observasi dengan usia lebih dari 100 tahun dihapus untuk memastikan validitas data; (2) Transformasi skala menggunakan metode normalisasi untuk menyamakan bobot setiap variabel, karena terdapat perbedaan rentang antar-variabel cukup besar; (3) Menghapus atau mengganti nilai yang hilang sesuai kondisi data. Dataset hasil pra-pemrosesan ditampilkan pada Tabel 1 untuk memberikan gambaran tentang struktur data yang digunakan.

Tabel 1. Data Pelanggan untuk Analisis Clustering DBSCAN

MntMeat Products	MntFruits	MntGold Prods	Education	Marital_Status	Income
1725	1	1	Graduation	Married	2447
1725	0	1	Graduation	Together	157146
1622	16	4	PhD	Married	160803
1582	2	1	PhD	Married	157243
984	129	18	Graduation	Single	80695

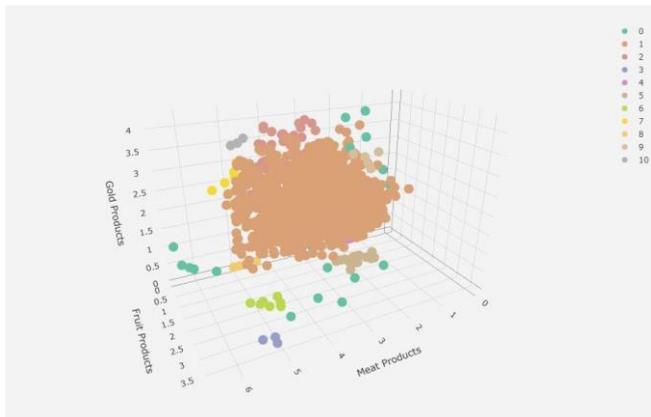
Data pelanggan yang telah melalui tahap pra-pemrosesan kemudian digunakan untuk proses clustering dengan algoritma DBSCAN. Langkah awal dalam proses ini adalah menentukan parameter utama, yaitu nilai epsilon (ϵ) dan minPts, yang akan memengaruhi pembentukan klaster. Dalam penelitian ini, nilai epsilon (ϵ) dan minPts dipilih melalui analisis k-distance plot. Grafik ini digunakan untuk mengidentifikasi *elbow point* yang menunjukkan jarak ideal antar data dalam klaster. Nilai ϵ sebesar 0.4

dipilih karena menunjukkan transisi signifikan pada grafik, sedangkan minPts sebesar 3 ditentukan berdasarkan jumlah minimum titik untuk membentuk kluster yang bermakna.



Gambar 1. K-Distance Plot untuk Penentuan Nilai Epsilon (ϵ)

Berdasarkan hasil analisis, nilai epsilon sebesar 0.4 dan minPts sebesar 3 digunakan sebagai parameter optimal untuk algoritma DBSCAN dalam penelitian ini. Selanjutnya, algoritma DBSCAN diterapkan pada variabel pembelian produk: MntMeatProducts, MntFruits, dan MntGoldProds. Untuk memberikan gambaran yang lebih jelas tentang distribusi kluster, hasil *clustering* divisualisasikan dalam bentuk *scatter plot* 3D. Visualisasi ini dipilih dengan tujuan mempermudah identifikasi hubungan antar cluster yang mungkin tidak terlihat jelas pada *scatter plot* 2D.



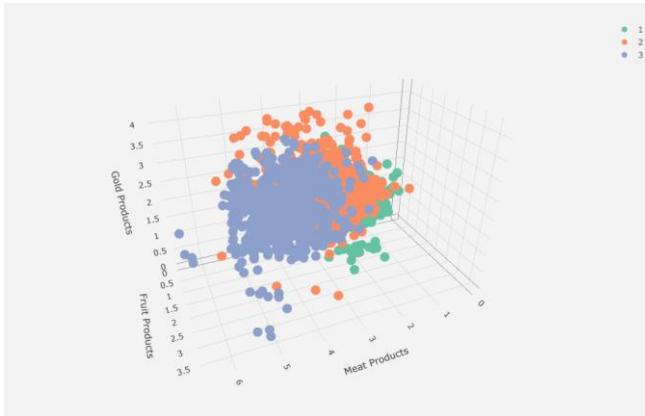
Gambar 2. Scatter Plot 3D Hasil Clustering dengan DBSCAN

Gambar 2 menampilkan hasil clustering menggunakan algoritma DBSCAN dalam bentuk scatter plot 3D yang menampilkan hubungan antara tiga variabel utama: *Meat Products* pada sumbu X, *Fruit Products* pada sumbu Y, dan *Gold Products* pada sumbu Z. Warna titik pada grafik sesuai legenda yang merepresentasikan 11 kluster yang terbentuk (0–10), dengan beberapa titik yang terpisah dari kluster utama yang menunjukkan *outlier* atau data *noise*.

Scatter plot menunjukkan bahwa sebagian besar titik data terkumpul dalam satu kluster besar yang terkonsentrasi di bagian tengah grafik dengan nilai sekitar 1 hingga 3 pada semua sumbu, mencerminkan pola belanja yang seragam pada pelanggan. Sementara itu, keberadaan *outlier* menggambarkan adanya pelanggan dengan perilaku pembelian yang berbeda secara signifikan. Kluster dominan (berwarna jingga, label 2) mencakup sebagian besar pengamatan, menunjukkan kesamaan karakteristik pada kelompok ini. Sementara itu, kluster yang lebih kecil seperti hijau, biru, dan abu-abu, tersebar di area luar grafik, merepresentasikan kelompok dengan karakteristik unik. Beberapa titik data menunjukkan nilai ekstrim, seperti konsumsi tinggi untuk *Meat Products* (mendekati 4) tetapi rendah untuk *Fruit Products* dan *Gold Products* (mendekati 0). Sebaran data ini mencerminkan adanya pola signifikan yang dapat dianalisis lebih lanjut untuk memahami hubungan antara variabel atau segmentasi berdasarkan preferensi konsumen terhadap produk tertentu. Visualisasi tersebut menegaskan bahwa algoritma DBSCAN mampu

menangkap struktur kepadatan data secara efektif, termasuk mengidentifikasi cluster utama dan *outlier* dalam dataset pelanggan.

Selain DBSCAN, algoritma K-Means juga diterapkan untuk mengelompokkan pelanggan. Tidak seperti DBSCAN yang bergantung pada kepadatan data, K-Means membagi dataset ke dalam sejumlah cluster berdasarkan jarak centroid. Hasil clustering menggunakan K-Means ditunjukkan pada Gambar 3.



Gambar 3. Scatter Plot 3D Hasil Clustering dengan *K-Means*

Gambar 3 memperlihatkan distribusi hasil clustering menggunakan algoritma *K-Means* dalam ruang tiga dimensi, berdasarkan variabel *MntMeatProducts*, *MntFruits*, dan *MntGoldProds*. Titik-titik pada grafik tersebut mewakili individu pelanggan, sedangkan warna yang berbeda menunjukkan pembagian ke dalam tiga kluster utama. Kluster 1 didominasi oleh pelanggan dengan konsumsi tinggi pada kategori produk daging, sementara kluster 2 dan kluster 3 menunjukkan pola konsumsi yang lebih seimbang antara buah dan emas.

Dibandingkan dengan hasil clustering DBSCAN pada Gambar 2, *K-Means* membentuk tiga kluster dengan batasan yang lebih terpusat. Meski demikian, terlihat adanya tumpang tindih antara kluster 2 dan kluster 3, terutama pada pelanggan dengan konsumsi *MntGoldProds* yang tinggi. Hal ini mengindikasikan bahwa *K-Means* cenderung membagi data secara merata berdasarkan jarak Euclidean, tanpa mempertimbangkan kemungkinan adanya data yang tidak termasuk dalam cluster tertentu (*outlier*).

Hasil *clustering* menunjukkan bahwa *K-Means* membagi data menjadi tiga kluster, lebih sedikit dibandingkan DBSCAN yang menghasilkan sebelas kluster. Meskipun *K-Means* lebih mudah diimplementasikan, algoritma ini kurang optimal dalam menangani data yang memiliki distribusi tidak seragam dan *outlier*. Sebaliknya, DBSCAN memberikan fleksibilitas lebih besar dalam mengenali pola data yang kompleks, termasuk mengidentifikasi data yang tidak masuk ke dalam cluster tertentu.

Dari perspektif bisnis, kemampuan DBSCAN untuk mendeteksi kluster kecil dan *outlier* membuka peluang untuk strategi pemasaran yang lebih personal dan tertarget. Kluster kecil dapat digunakan untuk mengidentifikasi pelanggan dengan kebutuhan khusus, seperti penawaran eksklusif untuk pembelian produk tertentu. Sementara itu, pelanggan yang terdeteksi sebagai *outlier* mungkin memerlukan perhatian lebih, misalnya dengan memberikan insentif khusus untuk mendorong keterlibatan mereka lebih lanjut.

Sebaliknya, hasil clustering *K-Means* memberikan gambaran yang lebih menyeluruh tentang segmentasi pelanggan dengan hanya tiga kluster utama. Ini sangat berguna untuk strategi pemasaran berskala besar yang membutuhkan pembagian kelompok pelanggan yang lebih sederhana. Kombinasi kedua algoritma dapat dimanfaatkan, di mana *K-Means* digunakan untuk membangun kerangka segmentasi awal, sementara DBSCAN memperkaya analisis dengan mendeteksi pola spesifik dan *outlier*.

4. Simpulan

Penelitian ini menunjukkan bahwa algoritma DBSCAN efektif dalam mengidentifikasi segmentasi pelanggan berdasarkan pola pembelian produk daging, buah, dan emas. Dengan parameter optimal ϵ sebesar 0,4 dan *minPts* sebesar 3, algoritma DBSCAN berhasil membentuk empat kluster utama serta mengidentifikasi data yang tergolong noise. Pendekatan ini terbukti unggul dibandingkan algoritma lain,

seperti K-Means, dalam menangani data dengan distribusi tidak teratur dan keberadaan outlier. Hasil clustering mengungkap pola konsumsi yang unik di antara pelanggan, termasuk kelompok dengan konsumsi moderat dan kelompok dengan preferensi ekstrem pada kategori produk tertentu. Sebagai contoh, mayoritas pelanggan berada dalam kluster dengan konsumsi seimbang pada semua kategori produk, sementara kluster-kluster kecil merepresentasikan pola pembelian yang lebih spesifik. Penelitian ini memberikan wawasan penting bagi perusahaan untuk menyusun strategi pemasaran yang lebih efektif, seperti penawaran produk yang relevan dan terarah berdasarkan karakteristik kluster pelanggan. Kedepannya, penelitian serupa dapat dilakukan dengan menambahkan variabel lain, seperti perilaku pembelian online atau loyalitas pelanggan, guna memperluas wawasan tentang preferensi pelanggan secara lebih komprehensif.

Daftar Pustaka

- Aazmaan Ahmed Khan, R. A. (2023). Customer segmentation using machine learning techniques. *Tuijin Jishu/Journal of Propulsion Technology*, 44(3), 2051–2061. <https://doi.org/10.52783/TJJPT.V44.I3.653>
- Anom, S., & Ali, H. (2024). Pengaruh Kualitas Produk, Harga, dan Lokasi Terhadap Sistem Informasi Pemasaran. *Jurnal Manajemen Pendidikan Dan Ilmu Sosial*, 5(3), 293–304. <https://doi.org/10.38035/JMPIS.V5I3.1944>
- Djogo, O. (2022). Strategi manajemen pemasaran dalam era digital pada masa sekarang. *Kebijakan : Jurnal Ilmu Administrasi*, 13(1), 43–47. <https://doi.org/10.23969/KEBIJAKAN.V13I1.5028>
- Fadila, M. A., & Voutama, A. (2023). Strategi promosi dengan pengelompokan pelanggan menggunakan clustering Kmeans. *Journal of Information System, Applied, Management, Accounting and Research*, 7(3), 672–680. <https://doi.org/10.52362/JISAMAR.V7I3.1151>
- Faza, A. H., Mutmainah, M., Kusumadini, L. R., Hidayat, R., & Ikaningtyas, M. (2024). Analisis segmentasi pasar dalam perencanaan bisnis industri ritel. *Merdeka : Jurnal Ilmiah Multidisiplin*, 1(4), 40–49. <https://doi.org/10.62017/MERDEKA.V1I4.1006>
- Hajari, V., Pratiwi, I., Putra, D. A., & Suhairi. (2024). Pemahaman riset pasar untuk pemasaran global primer dan sekunder. *Jurnal Masharif Al-Syariah: Jurnal Ekonomi Dan Perbankan Syariah*, 9(1). <https://doi.org/10.30651/JMS.V9I1.21454>
- Ismail, R., Retnawati, H., & Imawan, O. R. (2022). Cluster analysis of the national examination: School grouping to maintain the sustainability of high school quality. *REID (Research and Evaluation in Education)*, 8(1), 13–23. <https://doi.org/10.21831/REID.V8I1.45872>
- Kristianto, A. (2022). Implementasi DBSCAN dalam clustering data minat mahasiswa setelah pandemi Covid19. *KONSTELASI: Konvergensi Teknologi Dan Sistem Informasi*, 2(2), 426–431. <https://doi.org/10.24002/KONSTELASI.V2I2.5638>
- Lutfi Hamzah, R. (2016). *Penerapan metode hierarchical agglomerative clustering untuk segmentasi pelanggan potensial di jeger jersey Indonesia*.
- Miftahurrahmi, S., Zilrahmi, Amalita, N., & Mukhti, T. O. (2024). DBSCAN method in clustering Provinces in Indonesia based on crime cases in 2022. *UNP Journal of Statistics and Data Science*, 2(3), 330–337. <https://doi.org/10.24036/UJSDS/VOL2-ISS3/203>
- Mutiah, S., Hasnataeni, Y., Fitrianto, A., Risman Dwi Jumansyah, L., & dan Sains, S. (2024). *Perbandingan Metode Klastering K-Means dan DBSCAN dalam Identifikasi Kelompok Rumah Tangga Berdasarkan Fasilitas Sosial Ekonomi di Jawa Barat*. 09(02), 247–260. <https://doi.org/10.25157/teorema.v9i2.16290>
- Pratama, E. (2023). *Analisis metode clustering dengan Algoritma Spectral Clustering dalam pengelompokan tingkat tindak kriminalitas*.

- Pribadi, W. W., Yunus, A., & Wiguna, A. S. (2022). Perbandingan metode K-Means euclidean distance dan manhattan distance pada penentuan zonasi covid-19 di Kabupaten Malang. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 493–500. <https://doi.org/10.36040/JATI.V6I2.4808>
- 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.9.2.162-169>
- Safitri, D., Wuryandari, T., & Rahmawati, R. (2017). Metode DbSCAN Untuk Pengelompokan Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Produksi Padi Sawah Dan Padi Ladang. *Jurnal Statistika Universitas Muhammadiyah Semarang*, 5(1). <https://doi.org/10.26714/JSUNIMUS.5.1.2017>
- Saliano, Manan, L. O. A., & Utha, R. (2021). Strategi pemasaran dalam meningkatkan penjualan pada waroeng x-bro di Kota Kendari. *Sultra Journal of Economic and Business*, 2(1), 15–41. <https://doi.org/10.54297/SJEB.V2I1.151>
- Shetty, P., & Singh, S. (2021). Hierarchical clustering: a survey. *International Journal of Applied Research*, 7(4), 178–181. <https://doi.org/10.22271/ALLRESEARCH.2021.V7.I4C.8484>