

Penerapan Density-Based Spatial Clustering of Applications with Noise untuk Analisis Data Peserta Didik pada SMA Negeri 2 Samboja

Muhammad Ridho Sani^{*1}, Darmansyah²

^{1,2} STMIK Borneo Internasional

e-mail: ^{*1} mridho_sani.21@stmik-borneo.ac.id, ² darmansyah@lecturer.itk.ac.id

Abstrak

Pengelolaan data peserta didik yang jumlahnya besar seringkali menghadapi kendala dalam proses analisis, terutama untuk menemukan pola yang bermanfaat bagi pengambilan keputusan sekolah. Penelitian ini bertujuan menerapkan algoritma Density-Based Spatial Clustering of Applications with Noise (DBSCAN) untuk menganalisis data peserta didik pada SMA Negeri 2. DBSCAN dipilih karena mampu mengelompokkan data dengan bentuk cluster tidak beraturan dan mengidentifikasi outlier (noise). Penelitian menggunakan metode kuantitatif dengan data sekunder berupa profil peserta didik. Tahapan penelitian meliputi preprocessing data, penentuan parameter epsilon dan minPts, serta penerapan algoritma DBSCAN. Hasil penelitian menunjukkan bahwa DBSCAN dapat mengelompokkan peserta didik ke dalam beberapa cluster yang menggambarkan karakteristik tertentu, serta mendeteksi data yang menyimpang. Temuan ini dapat menjadi dasar bagi pihak sekolah dalam menyusun strategi pembinaan siswa sesuai karakteristik kelompok.

Kata kunci—DBSCAN; Clustering; Peserta Didik; SMA Negeri 2

Abstract

Managing large volumes of student data often poses challenges in the analysis process, particularly in identifying useful patterns for school decision-making. This study aims to apply the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to analyze student data at SMA Negeri 2. DBSCAN was chosen because it is capable of clustering data with irregular shapes and detecting outliers (noise). This research employed a quantitative method using secondary data from student profiles. The research stages included data preprocessing, determination of epsilon and minPts parameters, and the application of the DBSCAN algorithm. The results show that DBSCAN successfully grouped students into several clusters that represent specific characteristics, while also identifying outlier data. These findings can serve as a basis for schools in formulating student development strategies tailored to group characteristics.

Keywords— Data Mining; DBSCAN; Clustering; Student Data; SMA Negeri 2.

1. PENDAHULUAN

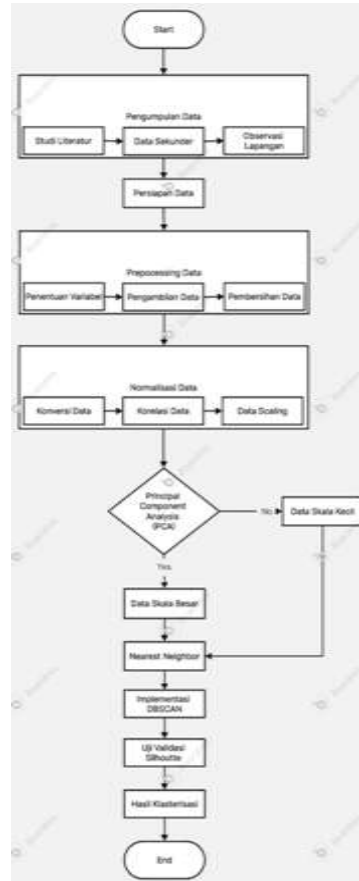
Perkembangan teknologi informasi membawa dampak signifikan pada dunia pendidikan, khususnya dalam pengelolaan data peserta didik yang semakin kompleks. Data siswa yang meliputi informasi akademik, kehadiran, serta aspek non-akademik dapat memberikan wawasan penting jika dianalisis dengan metode yang tepat. Namun, volume data yang besar seringkali menimbulkan kesulitan dalam menemukan pola yang relevan untuk mendukung pengambilan keputusan.

Data mining menjadi salah satu pendekatan yang dapat digunakan untuk menggali informasi tersembunyi dari data pendidikan. Teknik clustering memungkinkan pengelompokan data siswa berdasarkan kesamaan karakteristik sehingga dapat dimanfaatkan untuk pembinaan dan pengembangan potensi. Dari berbagai algoritma clustering, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) memiliki keunggulan dalam mengidentifikasi cluster dengan bentuk tidak beraturan sekaligus mendeteksi data outlier (noise).

SMA Negeri 2 dipilih sebagai objek penelitian karena memiliki jumlah peserta didik yang cukup besar dengan latar belakang yang beragam. Analisis menggunakan DBSCAN diharapkan dapat memberikan gambaran pola kelompok siswa yang berguna dalam strategi pembelajaran maupun kebijakan sekolah. Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma DBSCAN dalam analisis data peserta didik pada SMA Negeri 2 serta mengidentifikasi cluster yang terbentuk.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan teknik data mining berupa algoritma Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Objek penelitian adalah data peserta didik di SMA Negeri 2, yang dipilih karena memiliki jumlah siswa cukup besar dengan keragaman karakteristik. Tujuan dari metode ini adalah untuk menemukan pola pengelompokan siswa berdasarkan atribut akademik dan non-akademik.



Gambar 1. Langkah Penelitian

2.1. Jenis dan Sumber Data

Data yang digunakan adalah data sekunder berupa profil peserta didik yang diperoleh dari pihak sekolah. Atribut data mencakup nilai akademik, kehadiran, dan informasi lain yang relevan untuk analisis clustering.

2.2. Teknik Pengumpulan Data

Data dikumpulkan melalui dokumentasi dari arsip sekolah. Selanjutnya dilakukan pemeriksaan awal untuk memastikan kelengkapan dan konsistensi data sebelum dianalisis lebih lanjut.

2.3. Tahap Preprocessing Data

Sebelum algoritma DBSCAN diterapkan, dilakukan beberapa tahap preprocessing, yaitu:

- Data Cleaning: Menghapus data ganda, mengisi atau menghapus data yang kosong, dan mengoreksi kesalahan pencatatan.
- Data Transformation & Normalization: Mengubah format data agar seragam dan menormalkan atribut numerik sehingga berada pada skala yang sama.
- Attribute Selection: Menentukan atribut penting yang relevan dengan analisis, misalnya nilai rata-rata, jumlah ketidakhadiran, atau indikator prestasi siswa.

2.4. Penerapan Algoritma DBSCAN

Algoritma DBSCAN membutuhkan dua parameter utama:

- Epsilon (ϵ): Menentukan radius pencarian tetangga data.
- MinPts (minimum points): Menentukan jumlah minimum titik yang harus ada dalam radius ϵ untuk membentuk sebuah cluster. Penentuan nilai parameter dilakukan dengan metode trial and error serta analisis

distribusi data. DBSCAN kemudian dijalankan menggunakan perangkat lunak analisis data untuk membentuk cluster serta mengidentifikasi data noise (outlier).

2.5. Evaluasi Hasil Clustering

Cluster yang terbentuk dianalisis secara deskriptif untuk menafsirkan karakteristik tiap kelompok siswa, misalnya kelompok dengan nilai tinggi, kelompok dengan tingkat kehadiran rendah, atau kelompok dengan karakteristik campuran. Data yang masuk kategori noise juga ditinjau lebih lanjut karena berpotensi menunjukkan siswa dengan kondisi khusus.

Metode DBSCAN dipilih karena memiliki keunggulan dibandingkan algoritma clustering lain, yaitu mampu mendeteksi cluster dengan bentuk tidak beraturan, tidak memerlukan jumlah cluster ditentukan di awal, serta efektif dalam mengidentifikasi outlier pada dataset pendidikan.

3. HASIL DAN PEMBAHASAN

Penerapan algoritma DBSCAN pada data peserta didik SMA Negeri 2 Samboja menghasilkan empat cluster utama dan satu kategori noise. Evaluasi menggunakan silhouette score menunjukkan bahwa kombinasi parameter terbaik diperoleh pada $\epsilon = 0,09$ dan $\text{minPts} = 5$, dengan nilai $0,805574$, yang menandakan kualitas klasterisasi sangat baik.

3.1. Hasil Klasterisasi

Distribusi hasil pengelompokan adalah sebagai berikut:

- Cluster 0 (123 siswa / 29,93%) → mayoritas perempuan usia 16–18 tahun, berasal dari keluarga ekonomi menengah ke bawah, sebagian besar layak menerima PIP, dengan pekerjaan orang tua dominan petani atau wiraswasta.
- Cluster 1 (60 siswa / 14,60%) → mayoritas laki-laki, keluarga berpenghasilan rendah (Rp 500 ribu – Rp 1 juta), sebagian besar juga penerima PIP, dengan pekerjaan orang tua umumnya petani.
- Cluster 2 (114 siswa / 27,74%) → mayoritas laki-laki, berasal dari keluarga menengah ke atas dengan penghasilan stabil (Rp 2–5 juta), pekerjaan orang tua karyawan swasta atau wiraswasta, dan sebagian besar tidak menerima PIP.
- Cluster 3 (107 siswa / 26,03%) → mayoritas perempuan dari keluarga menengah ke atas, dengan pekerjaan orang tua sebagai PNS/TNI/Polri atau karyawan swasta, serta sebagian besar tidak menerima PIP.
- Noise (7 siswa / 1,70%) → data anomali, misalnya umur orang tua tidak realistis, lingkaran kepala bernilai 0 atau 5, serta kategori tempat tinggal atau pekerjaan orang tua yang jarang muncul.

3.2. Pembahasan

Hasil klasterisasi ini memberikan gambaran yang jelas mengenai profil sosial-ekonomi siswa:

- Cluster 0 dan 1 menggambarkan kelompok siswa dengan keterbatasan ekonomi, sehingga dapat menjadi prioritas dalam pemberian bantuan pendidikan.
- Cluster 2 dan 3 merepresentasikan siswa dari keluarga menengah ke atas, yang lebih tepat difokuskan pada strategi pengembangan akademik dan minat bakat.
- Noise mengindikasikan pentingnya validasi data administratif sekolah agar analisis lebih akurat.

Secara akademis, penelitian ini menegaskan efektivitas DBSCAN dalam menganalisis data pendidikan dengan distribusi yang kompleks dan tidak merata. Secara praktis, hasil penelitian bermanfaat bagi sekolah dalam menyusun kebijakan berbasis data, baik untuk penyaluran bantuan maupun program peningkatan kualitas belajar.

4. KESIMPULAN

Penelitian ini membuktikan bahwa algoritma DBSCAN (Density-Based Spatial Clustering of Applications with Noise) dapat diterapkan secara efektif untuk menganalisis data peserta didik di SMA Negeri 2 Samboja. Hasil klasterisasi menunjukkan kombinasi parameter terbaik pada $\epsilon = 0,09$ dan $\text{minPts} = 5$, dengan nilai silhouette score sebesar $0,805574$. Dari analisis tersebut terbentuk empat cluster utama dan satu kategori noise. Klaster terbesar adalah Cluster 0 (29,93%), diikuti oleh Cluster 2 (27,74%), Cluster 3 (26,03%), dan Cluster 1 (14,60%), sedangkan noise hanya mencakup 1,70% data.

Setiap cluster memiliki karakteristik berbeda yang mencerminkan kondisi sosial-ekonomi siswa, meliputi variabel jenis kelamin, umur, penghasilan orang tua, status kelayakan PIP, pekerjaan orang tua, serta alat transportasi. DBSCAN terbukti efektif dalam menangani noise dan mengelompokkan data tanpa perlu menentukan jumlah cluster sebelumnya.

DAFTAR PUSTAKA

- [1] Anggara, R., Rahman, A., & Si, S. (2022). Implementasi algoritma DBSCAN dalam mengelompokkan data pasien terdiagnosa penyakit ginjal kronis (PGK). *Jurnal AlgoritmeCCS*, 1(1), 114–123. <https://doi.org/10.35957/algoritme.xxxx>
- [2] Asrulla, A., Risnita, R., Jailani, M. S., & Jeka, F. (2023). Populasi dan sampling (kuantitatif), serta pemilihan informan kunci (kualitatif) dalam pendekatan praktis. Retrieved from <https://www.researchgate.net/publication/386875018>
- [3] 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>
- [4] Hasan, Y. (2024). Pengukuran silhouette score dan Davies-Bouldin index pada hasil cluster K-Means dan DBSCAN. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(3S1). <https://doi.org/10.23960/jitet.v12i3s1.5001>
- [5] Lastari, W. (2023). Penerapan data mining untuk memprediksi prestasi siswa SMA pada Dinas Pendidikan Provinsi Jambi.
- [6] Mai, P., et al. (2022). Implementasi data mining menggunakan algoritma Apriori dalam menentukan persediaan barang (Studi kasus: Toko Sinar Harahap). Retrieved from <https://jurnal.umj.ac.id/index.php/just-it/index>
- [7] Nugraha, A., Nurdiawan, O., & Dwilestari, G. (2022). Penerapan data mining metode K-Means clustering untuk analisa penjualan pada Toko Yana Sport.
- [8] Nurina Sari, B., Primajaya, A., & Ronggowaluyo, J. H. (2019). Penerapan clustering DBSCAN untuk pertanian padi di Kabupaten Karawang. Retrieved from <http://www.mapcoordinates.net/en>
- [9] Oktaviani, F., & Damayanti, N. R. (2024). Analisis data peserta didik sekolah menengah atas (SMA) menggunakan visualisasi Google Looker Studio. *JSAI: Journal Scientific and Applied Informatics*, 7(3). <https://doi.org/10.36085>
- [10] Prana, A., Sembiring, U., & Ginting, M. (n.d.). Analisis faktor-faktor yang mempengaruhi pengunduran diri mahasiswa dengan aplikasi data mining add-ins (Studi kasus pada STMIK Mikroskil).
- [11] Qadrini, L., et al. (2020). Metode K-Means dan DBSCAN pada pengelompokan data dasar kompetensi laboratorium ITS tahun 2017. Retrieved from <http://www.unipasby.ac.id>
- [12] Rahmah, S., Jamil, M., Aslindah, A., Fawait, A. B., & Saputra, Y. F. (2025). Pengelompokan hasil pembelajaran mahasiswa dengan algoritma K-Means clustering. *Jurnal Pengabdian Masyarakat dan Riset Pendidikan*, 3(4), 4215–4221. <https://doi.org/10.31004/jerkin.v3i4.1217>
- [13] Resnawati, R., Fadryani, F., Najar, A. M., Puspita, J. W., Bin Mardi, A., & Abu, M. (2024). Pelatihan dan pendampingan pemrograman Python dalam meningkatkan kompetensi siswa SMKN 5 Palu. *Jurnal Pengabdian Farmasi dan Sains*, 2(2), 6–12. <https://doi.org/10.22487/jpsf.2024.v2.i2.16879>
- [14] Wardhana, A. (2024). Teknik pengumpulan data penelitian. Retrieved from <https://www.researchgate.net/publication/382060598>
- [15] Yadav, S. K., & Pal, S. (2012). Data mining: A prediction for performance improvement of engineering students using classification. *World of Computer Science and Information Technology Journal*, 2(2), 51–56.
- [16] Yanuar, A., & Andarsyah, R. (2024). Sentimen analisis aplikasi PosAja pada Google Playstore untuk peningkatan Pospay Superapp menggunakan Support Vector Machine.