



SKRIPSI

PENERAPAN METODE HDBSCAN (*HIERARCHICAL DENSITY-BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE*) UNTUK PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR KETENAGAKERJAAN DALAM ANALISIS DAYA SAING ANGKATAN KERJA

CITRA AMELIA INTAN PERMADANI
NPM 21083010004

DOSEN PEMBIMBING

Aviolla Terza Damaliana, S.Si., M.Stat.
Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025**



SKRIPSI

PENERAPAN METODE HDBSCAN (*HIERARCHICAL DENSITY-BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE*) UNTUK PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR KETENAGAKERJAAN DALAM ANALISIS DAYA SAING ANGKATAN KERJA

CITRA AMELIA INTAN PERMADANI
NPM 21083010004

DOSEN PEMBIMBING

Aviolla Terza Damaliana, S.Si., M.Stat.
Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025**



SKRIPSI

PENERAPAN METODE HDBSCAN (*HIERARCHICAL DENSITY-BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE*) UNTUK PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR KETENAGAKERJAAN DALAM ANALISIS DAYA SAING ANGKATAN KERJA

CITRA AMELIA INTAN PERMADANI
NPM 21083010004

DOSEN PEMBIMBING

Aviolla Terza Damaliana, S.Si., M.Stat.
Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025**

LEMBAR PENGESAHAN

PENERAPAN METODE HDBSCAN (*HIERARCHICAL DENSITY-BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE*) UNTUK PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR KETENAGAKERJAAN DALAM ANALISIS DAYA SAING ANGKATAN KERJA

Oleh:
CITRA AMELIA INTAN PERMADANI
NPM. 21083010004

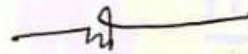
Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 10 September 2025:

Menyetujui,

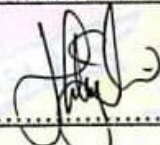
Aviolla Terza Damaliana, S.Si., M.Stat.
NIP. 19940802 202203 2 015


..... (Pembimbing I)

Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
NIP. 19830310 202121 1 006


..... (Pembimbing II)

Wahyu Syaifullah J. S, S.Kom., M.Kom.
NIP. 19860825 202121 1 003


..... (Ketua Penguji)

Andri Fauzan Adziima, M. Si.
NIP. 19950512 202406 1 001


..... (Penguji I)

Mengetahui,
Dekan Fakultas Ilmu Komputer


Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

PENERAPAN METODE HDBSCAN (*HIERARCHICAL DENSITY-BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE*) UNTUK PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR KETENAGAKERJAAN DALAM ANALISIS DAYA SAING ANGKATAN KERJA

Oleh:

CITRA AMELIA INTAN PERMADANI
NPM. 21083010004

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer

Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., Asean, Eng.
NIP. 198012052005011002

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertandatangan di bawah ini:

Nama Mahasiswa	:	Citra Amelia Intan Permadani
NPM	:	21083010004
Program	:	Sarjana (S1)
Program Studi	:	Sains Data
Fakultas	:	Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperilah gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokuman ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 15 September 2025
Yang Membuat Pernyataan,



CITRA AMELIA INTAN PERMADANI
NPM. 21083010004

ABSTRAK

Nama Mahasiswa / NPM : Citra Amelia Intan Permadani / 21083010004
Judul Skripsi : Penerapan Metode HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Application with Noise*) Untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Ketenagakerjaan dalam Analisis Daya Saing Angkatan Kerja
Dosen Pembimbing : 1. Aviolla Terza Damaliana, S.Si., M.Stat
2. Dr. Ir. Mohammad Idhom, SP, S. Kom., MT.

Angkatan kerja merupakan penduduk berusia 15 tahun ke atas yang aktif dalam kegiatan ekonomi dan mencerminkan potensi sumber daya manusia yang siap bekerja. Namun, Indonesia menghadapi tantangan berupa peningkatan jumlah angkatan kerja yang tidak sebanding dengan ketersediaan lapangan kerja serta rendahnya kualifikasi pendidikan yang menghambat daya saing sebagaimana ditegaskan dalam Peraturan Presiden Nomor 12 Tahun 2025. Variabel usia produktif (15–64 tahun) memiliki tingkat partisipasi lebih tinggi, pendidikan yang lebih baik berkorelasi dengan sektor formal dan upah lebih layak, sedangkan ketersediaan lapangan kerja berpengaruh pada mobilitas tenaga kerja. Penelitian ini bertujuan menerapkan metode HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*) untuk mengelompokkan provinsi di Indonesia berdasarkan indikator ketenagakerjaan meliputi usia produktif, lulusan pendidikan, dan sektor pekerjaan untuk menganalisis kesiapan daya saing angkatan kerja kelompok provinsi. Metode HDBSCAN dipilih karena mampu mengeksplorasi berbagai skala kepadatan untuk membentuk klaster yang stabil serta mengidentifikasi dan memisahkan pencila (*outlier*) yang umum ditemukan pada data ketenagakerjaan akibat perbedaan kondisi sosial dan ekonomi antarprovinsi. Dengan parameter terbaik *min_cluster_size* = 2 dan metrik *Euclidean*, diperoleh nilai *Silhouette Coefficient* sebesar 0,546 dengan hasil tiga klaster utama dan satu klaster *noise*, yakni klaster 0 berjumlah 3 provinsi, klaster 1 berjumlah 25 provinsi, klaster 2 berjumlah 3 provinsi, dan klaster -1 (*noise*) berjumlah 3 provinsi. Hasil ini menegaskan bahwa distribusi ketenagakerjaan di Indonesia tidak merata, terutama terkait tingkat pendidikan dan dominasi sektor pekerjaan. Penelitian ini juga mengembangkan GUI berbasis web interaktif menggunakan *framework Streamlit* untuk memvisualisasikan hasil klasterisasi dalam bentuk peta interaktif.

Kata kunci : Angkatan Kerja, *Outlier*, Klasterisasi, *Hierarchical Density-Based Spatial Clustering of Application with Noise* (HDBSCAN), Streamlit

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Citra Amelia Intan Permadani / 21083010004
Thesis Title : *Application of the HDBSCAN Method (Hierarchical Density-Based Spatial Clustering of Applications with Noise) for Clustering Provinces in Indonesia Based on Employment Indicators in Labor Force Competitiveness Analysis*
Advisor : 1. Aviolla Terza Damaliana, S.Si., M.Stat
2. Dr. Ir. Mohammad Idhom, SP, S. Kom., MT.

The labor force consists of people aged 15 years and above who are active in economic activities and reflects the potential of human resources who are ready to work. However, Indonesia faces challenges in the form of an increase in the labor force that is not proportional to the availability of jobs and low educational qualifications that hamper competitiveness, as stated in Presidential Regulation No. 12 of 2025. The productive age variable (15–64 years) has a higher participation rate, better education correlates with the formal sector and more decent wages, while job availability affects labor mobility. This study aims to apply the HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) method to group provinces in Indonesia based on employment indicators, including productive age, education graduates, and employment sectors, to analyze the competitiveness of the provincial workforce. The HDBSCAN method was chosen because it is capable of exploring various density scales to form stable clusters and identify and separate outliers commonly found in labor data due to differences in social and economic conditions between provinces. With the best parameters of `min_cluster_size = 2` and the Euclidean metric, a Silhouette Coefficient value of 0.546 was obtained with three main clusters and one noise cluster, namely cluster 0 consisting of 3 provinces, cluster 1 consisting of 25 provinces, cluster 2 consisting of 3 provinces, and cluster -1 (noise) consisting of 3 provinces. These results confirm that the distribution of employment in Indonesia is uneven, especially in terms of education levels and the dominance of employment sectors. This study also developed an interactive web-based GUI using the Streamlit framework to visualize the clustering results in the form of interactive maps.

Keywords: *Labor, Outlier, Clustering, Hierarchical Density-Based Spatial Clustering with Noise (HDBSCAN), Streamlit*

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Penerapan Metode HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Application with Noise*) Untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Ketenagakerjaan dalam Analisis Daya Saing Angkatan Kerja”** dapat terselesaikan dengan baik. Dalam kesempatan ini, penulis ingin menyampaikan rasa terima kasih kepada pihak-pihak yang telah memberikan bantuan dan bimbingan baik itu berupa moril, spiritual, maupun materiil. Untuk itu penulis mengucapkan terima kasih kepada:

1. Universitas Pembangunan Nasional “Veteran” Jawa Timur sebagai tempat menimba ilmu dan pengalaman bagi penulis.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT., selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Bapak Dr. Eng. Ir. Dwi Arman Prasetya., ST., MT., IPU., Asean. Eng., selaku Ketua Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Aviolla Terza Damaliana, S.Si., M.Stat., selaku Dosen Pembimbing Pertama skripsi dalam proses memberikan bimbingan, arahan, kesabaran serta semangat kepada penulis sejak tahap awal hingga akhir pengerjaan dan penyusunan skripsi ini.
5. Bapak Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T., selaku Dosen Pembimbing Kedua skripsi yang senantiasa memberikan bimbingan, masukan, saran, koreksi, dan semangat kepada penulis dalam proses penyempurnaan skripsi ini.
6. Kepada Ayah dan Ibu sebagai pahlawan dihidup penulis yang selalu memberikan seluruh kasih sayang, doa, tenaga, waktu, rezeki serta dukungan sehingga menjadi salah satu sumber inspirasi dan kekuatan bagi penulis dalam menyelesaikan setiap proses saat ini hingga selamanya. Ayah dan ibu yang selalu mengusahakan anak pertamanya ini menempuh pendidikan setinggi-tingginya meskipun ayah dan ibu tidak sempat merasakan pendidikan samapi

bangku perkuliahan. Terima kasih atas segala pengertian, perhatian, dan kesabaran dalam memahami dan menemani penulis dalam menyelesaikan setiap proses studinya hingga akhir.

7. Kakek dan nenek yang dengan tulus selalu memberikan doa, kasih sayang, dan perhatian sehingga menjadi salah satu sumber kekuatan bagi penulis dalam menyelesaikan setiap proses hingga akhir dan selamanya.
8. Adik Salsabila dan Mas Syariq Muhammad H.H yang selalu memberikan dukungan melalui hiburan, perhatian, semangat membuat penulis tersenyum dan terhibur sehingga lebih semangat dalam menyelesaikan proses hingga akhir.
9. Kepada teman-teman Sains Data angkatan 2021 yang menemani, memberikan informasi-informasi dan semangat kepada penulis selama proses penyusunan skripsi hingga mampu menyelesaikan skripsi ini hingga akhir.
10. Kepada rekan-rekan penulis: Dwinggrit Oktaviani Putri, Delinda Brilliant Cahayani, Adelia Azizatul Haq, Ellexia Leoni, dan rekan-rekan lainnya yang sudah memberikan semangat, motivasi, dan doa-doa tulus kepada penulis selama masa perkuliahan hingga lulus.
11. Tak lupa, penulis juga menyampaikan terima kasih kepada diri saya sendiri sebagai penulis atas ketekunan, kesabaran, kekuatan, ketahanan, dan semangat yang berusaha selalu terjaga selama proses perkuliahan dan penyusunan skripsi ini. Meski banyak sekali tantangan baru, penulis selalu bersyukur mampu dimudahkan melewati berbagai tantangan tersebut sehingga penulis akhirnya dapat menyelesaikan skripsi ini dengan baik dan lulus.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, September 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iv
ABSTRAK	v
<i>ABSTRACT</i>	vii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xv
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN	xix
DAFTAR NOTASI.....	xxi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	7
1.3. Batasan Masalah.....	7
1.4. Tujuan Penelitian	8
1.5. Manfaat Penelitian	8
BAB II TINJAUAN PUSTAKA.....	11
2.1. Penelitian Terdahulu	11
2.2. Dasar Teori.....	16
2.1.1. Ketenagakerjaan.....	16
2.1.2. <i>Clustering</i>	20
2.1.3. Standarisasi Data	21
2.1.4. Deteksi <i>Outlier</i>	22
2.1.5. HDBSCAN.....	24
2.1.6. <i>Silhouette Coefficient</i>	27
2.1.7. <i>Streamlit</i>	28
BAB III DESAIN DAN IMPLEMENTASI SISTEM	31
3.1. Variabel Penelitian dan Sumber Data	31
3.2. Langkah Analisis.....	32

3.2.1.	Identifikasi Masalah	33
3.2.2.	Pengumpulan Data dan Studi Literatur	34
3.2.3.	Eksplorasi Data	34
3.2.4.	Data <i>Preprocessing</i>	35
3.2.5.	Deteksi <i>Outlier</i>	35
3.2.6.	Klasterisasi dengan Algoritma <i>Hierarchical Density-Based Spatial Clustering of Application with Noise</i> (HDBSCAN)	36
3.2.7.	Evaluasi dengan <i>Silhouette Coefficient</i>	37
3.2.8.	Analisis Hasil Klasterisasi	37
3.2.9.	Pembuatan <i>Graphical User Interface</i> (GUI)	38
3.3.	Desain Sistem	38
BAB IV HASIL DAN PEMBAHASAN		41
4.1.	Dataset Awal	41
4.2.	Eksplorasi Data	42
4.3.	Data <i>Preprocessing</i>	44
4.3.1.	Pengecekan <i>Missing Value</i> dan Data Duplikat	44
4.3.2.	Standarisasi Data	45
4.4.	Deteksi <i>Outlier</i>	47
4.5.	Klasterisasi Menggunakan Algoritma HDBSCAN	49
4.5.1.	Menentukan Parameter	49
4.5.2.	Implementasi Perhitungan Manual	56
4.5.3.	Melakukan Klasterisasi	59
4.6.	Evaluasi Hasil Klasterisasi	62
4.7.	Analisis Hasil Klasterisasi	63
4.7.1	Hasil Klasterisasi HDBSCAN	64
4.7.2	Perbandingan HBDSCAN dengan <i>Agglomerative Hierarchical Clustering</i>	66
4.7.3	Analisis <i>Noise</i> dalam Klaster dan Performa HDBSCAN	68
4.8.	Interpretasi Hasil Klasterisasi	71
4.9.	Implementasi GUI	73
4.9.1.	Menu Home	74
4.9.2.	Menu Materi	74

4.9.3. Menu Klasterisasi.....	75
BAB V PENUTUP	81
5.1. Kesimpulan	81
5.2. Saran Pengembangan	83
DAFTAR PUSTAKA	85
LAMPIRAN.....	91

Halaman ini sengaja dikosongkan

DAFTAR GAMBAR

Gambar 3. 1 Diagram Alir Penelitian	33
Gambar 3. 2 Desain Sistem GUI.....	39
Gambar 4. 1 Visualisasi Outlier dengan LOF	49
Gambar 4. 2 Pengujian <i>min_cluster_size</i> = 2 dan metrik <i>Euclidean</i>	51
Gambar 4. 3 Pengujian <i>min_cluster_size</i> = 3 dan metrik <i>Euclidean</i>	52
Gambar 4. 4 Pengujian <i>min_cluster_size</i> = 4 dan metrik <i>Euclidean</i>	52
Gambar 4. 5 Pengujian <i>min_cluster_size</i> = 2 dan metrik <i>Manhattan</i>	53
Gambar 4. 6 Pengujian <i>min_cluster_size</i> = 3 dan metrik <i>Manhattan</i>	53
Gambar 4. 7 Pengujian <i>min_cluster_size</i> = 4 dan metrik <i>Manhattan</i>	54
Gambar 4. 8 Pengujian <i>min_cluster_size</i> = 5 dan metrik <i>Manhattan</i>	54
Gambar 4. 9 Pengujian <i>min_cluster_size</i> = 6 dan metrik <i>Manhattan</i>	55
Gambar 4. 10 Condensed Tree Plot	60
Gambar 4. 11 Visualisasi Klaster.....	71
Gambar 4. 12 Menu Home.....	74
Gambar 4. 13 Menu Materi.....	75
Gambar 4. 14 Upload Data.....	76
Gambar 4. 15 Stastika Deskriptif.....	77
Gambar 4. 16 Standarisasi Data	77
Gambar 4. 17 Deteksi Outlier	78
Gambar 4. 18 Parameter HDBSCAN.....	78
Gambar 4. 19 Evaluasi dan Visualisasi Klaster	79
Gambar 4. 20 Peta Interaktif Hasil Klaster	79
Gambar 4. 21 Interpretasi Hasil Klaster.....	80

Halaman ini sengaja dikosongkan

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	11
Tabel 2. 2 Nilai Silhouette Coefficient	28
Tabel 3. 1 Variabel Penelitian.....	31
Tabel 3. 2 Struktur Data Penelitian	32
Tabel 3. 3 Pengujian Parameter HDBSCAN	36
Tabel 4. 1 Dataset Awal Variabel Usia, Pendidikan, dan Sektor Pekerjaan per Provinsi.	41
Tabel 4. 2 Eksplorasi Data Variabel Usia Produktif hingga Perguruan Tinggi	42
Tabel 4. 3 Eksplorasi Data Variabel Sektor Pekerjaan Primer hingga Tersier	43
Tabel 4. 4 Hasil Cek <i>Missing Value</i>	44
Tabel 4. 5 Hasil Cek Data Duplikat	45
Tabel 4. 6 Hasil Standarisasi Data	46
Tabel 4. 7 Hasil Standar Deviasi Skor LOF untuk Threshold	48
Tabel 4. 8 Pengujian Parameter HDBSCAN	50
Tabel 4. 9 5 Data Awal untuk Perhitungan Manual.....	56
Tabel 4. 10 Hasil Perhitungan Jarak Euclidean 5 Data Awal	57
Tabel 4. 11 <i>Core Distance</i> 5 Data Awal	57
Tabel 4. 12 <i>Mutual Reachability Distance</i> 5 Data Awal	58
Tabel 4. 13 Nilai <i>Minimum Spanning Tree</i> 5 Data Awal.....	59
Tabel 4. 14 Hasil Klaster Perhitungan Manual 5 Data Awal	59
Tabel 4. 15 Hasil Klasterisasi dengan HDBSCAN.....	61
Tabel 4. 16 Evaluasi Algoritma DBSCAN vs HDBSCAN.....	62
Tabel 4. 17 Analisis Hasil Klasterisasi	64
Tabel 4. 18 Hasil Metode <i>Single Linkage</i> sebagai Perbandingan dengan HDBSCAN	66
Tabel 4. 19 Probabilitas Hasil Klasterisasi dengan HDBSCAN.....	68
Tabel 4. 20 Persentasi Variabel tiap Klaster	70
Tabel 4. 21 Rata-Rata Hasil Klaster.....	72

Halaman ini sengaja dikosongkan

DAFTAR LAMPIRAN

Lampiran 1. Data Penelitian <i>X1 – X5</i>	91
Lampiran 2. Data Penelitian <i>X6 – X8</i>	92
Lampiran 3. Source Code Program	94
Lampiran 4. GUI Streamlit	95
Lampiran 5. LoA Jurnal	96

Halaman ini sengaja dikosongkan

DAFTAR NOTASI

x_i	:	Data ke- i
$\bar{x}$	:	Rata-rata data
s	:	Standar deviasi dari data
p	:	Titik data yang sedang dihitung <i>Local Reachability Density</i> -nya
o	:	Titik data tetangga
$ N_{MinPts}(p) $	:	Jumlah tetangga p dalam suatu <i>minPts</i>
$\in$	:	Anggota himpunan
$k = M_{pts}$	:	Minimum klaster
a	:	Titik yang sedang dihitung
$neighbor_k$	:	Tetangga ke- k terdekat dari titik a
d	:	Jarak <i>euclidean</i>
$(x_1 - y_1)$	:	Koordinat titik pertama
$(x_n - y_n)$	:	Koordinat titik ke- n
s_i	:	<i>Silhouette coefficient</i>
a_i	:	Rata-rata jarak dari suatu objek dengan objek lain yang berada dalam satu <i>cluster</i>
b_i	:	Nilai jarak yang terkecil
$ A $	:	Banyaknya objek yang berada di klaster A
li, j	:	Indeks dari objek
$d_{i, j}$	:	Selisih antar objek ke- i dengan objek ke- j
$d_{i, A'}$	:	Rata-rata jarak objek i dengan keseluruhan objek yang berada di klaster lain
A'	:	Jumlah objek yang berada di klaster selain klaster A
b_i	:	Nilai jarak yang terkecil