

**PERBANDINGAN ALGORITMA *CLUSTERING DBSCAN* DAN
K-MEANS DALAM PENGELOMPOKAN SISWA TERBAIK**

SKRIPSI

MELDA NOVITA SARI

0701193181



**UNIVERSITAS ISLAM NEGERI
SUMATERA UTARA MEDAN**

**UNIVERSITAS ISLAM NEGERI
SUMATERA UTARA MEDAN**

**PROGRAM STUDI ILMU KOMPUTER
FAKULTAS SAINS DAN TEKNOLOGI**

UNIVERSITAS ISLAM NEGERI SUMATERA UTARA

MEDAN

2024

PERSETUJUAN SKRIPSI

Hal : Surat Persetujuan Skripsi

Lampiran : -

Kepada Yth.,

Dekan Fakultas Sains dan Teknologi

Universitas Islam Negeri Sumatera Utara Medan

Assalamu'alaikum Wr. Wb.

Setelah membaca, meneliti, memberikan petunjuk, dan mengoreksi serta mengadakan perbaikan, maka kami selaku pembimbing berpendapat bahwa skripsi saudara,

Nama : Melda Novita Sari

NIM : 0701193181

Program Studi : Ilmu Komputer

Judul : Perbandingan algoritma clustering dbscan dan
k-means dalam menentukan siswa terbaik

dapat disetujui untuk segera *dimunqasyahkan*. Atas perhatiannya kami ucapkan terimakasih.

Medan, 10 Juli 2024

Komisi Pembimbing,

Pembimbing Skripsi I



Muhammad Siddik Hasibuan, M.Kom
NIP. 198611152019031008

Pembimbing Skripsi II



Aidil Halim Lubis, M.Kom
NIP. 198805272019031010



KEMENTERIAN AGAMA REPUBLIK INDONESIA
UNIVERSITAS ISLAM NEGERI SUMATERA UTARA MEDAN
FAKULTAS SAINS DAN TEKNOLOGI

Jl. Lap. Golf, Durin Jangak, Kec. Pancur Batu, Kab. Deli Serdang K. Pos 20353
Url: www.saintek.uinsu.ac.id, E-mail: saintek@uinsu.ac.id

PENGESAHAN SKRIPSI

Nomor : B.237/ST/ST.V.2/PP.01.1/03/2024

Judul : Perbandingan Algoritma Clustering DBSCAN
dan K-Means dalam Mengelompokkan Siswa Terbaik
Nama : Melda Novita Sari
NIM : 0701193181
Program Studi : Ilmu Komputer
Fakultas : Sains dan Teknologi

Telah dipertahankan di hadapan Dewan Penguji Skripsi Program Studi Ilmu Komputer
Fakultas Sains dan Teknologi UIN Sumatera Utara Medan dan dinyatakan LULUS

Pada hari/tanggal : Selasa, 6 Agustus 2024
Tempat : Ruang Sidang Fakultas Sains dan Teknologi

Tim Ujian Munaqasyah

Ketua,

Ilka Zufria S. Kom. M. Kom.
NIP. 198506042015031006

Dewan Penguji,

Penguji 1

Dr. Mhd. Furqan S. Si S.H M. Comp. Sc.
NIP. 198008062006041003

Pembimbing 1

Muhammad Siddik Hasibuan M. Kom.
NIP. 198611252019031008

Penguji 2

Sriani M. Kom.
NIP. 198407032023212029

Pembimbing 2

Aidil Halim Lubis M. Kom.
NIP. 198805272019031010

Mengesahkan,

Dekan Fakultas Sains dan Teknologi
Universitas Islam Negeri Sumatera Utara Medan



Muhammad S.H.I., M.Hum.
NIP. 197703212009011008

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Saya yang bertanda tangan di bawah ini,

Nama : Melda Novita Sari

NIM : 0701193181

Program Studi : Ilmu Komputer

Judul : Perbandingan Algoritma *Clustering Dbscan* dan *K-Means*
dalam Pengelompokan Siswa Terbaik

menyatakan bahwa Skripsi ini adalah hasil karya saya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya. Apabila di kemudian hari ditemukan plagiat dalam Skripsi ini maka saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksi lainnya sesuai dengan peraturan yang berlaku.

Medan, 10 Juli 2024



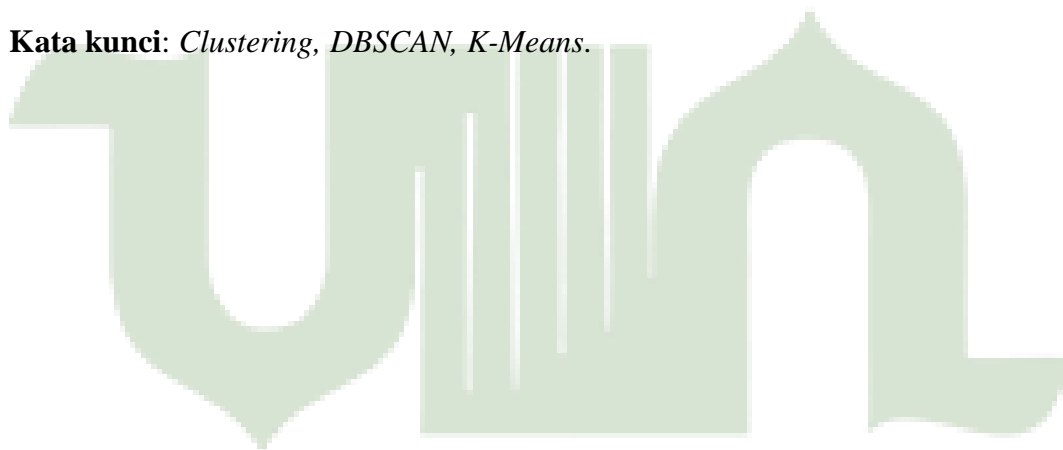
Melda Novita Sari

NIM. 0701193181

ABSTRAK

Pendidikan merupakan aspek penting dalam mengembangkan sumber daya manusia yang berkualitas. Evaluasi prestasi siswa adalah bagian krusial untuk pengelompokan siswa terbaik. Metode *clustering* adalah teknik yang dapat digunakan untuk mengelompokkan data prestasi siswa berdasarkan berbagai kriteria, sehingga memudahkan proses penilaian. Skripsi ini membahas kinerja dua algoritma *clustering*, yaitu *DBSCAN* dan *K-Means*, dalam pengelompokan siswa SMP terbaik. Penelitian ini bertujuan untuk mengetahui algoritma mana yang lebih efisien dan efektif dalam mengelompokkan data siswa SMP berdasarkan prestasi akademik mereka. Metode pengumpulan data pada penelitian ini dengan melakukan observasi dan wawancara. Hasil penelitian menunjukkan bahwa algoritma *DBSCAN* dengan epsilon 11 dan sampel data 24 menghasilkan 2 kluster dengan Silhouette Score sebesar 0.258030877243884. Sedangkan algoritma *K-Means* dengan metode Elbow menghasilkan 4 kluster dengan Silhouette Score sebesar 0.5697019340266847. Dari perbandingan tersebut, algoritma *K-Means* terbukti lebih efisien. *K-Means* juga menunjukkan kluster yang lebih bersih, terstruktur, dan memiliki kemiripan antar data yang lebih baik, sehingga lebih disarankan untuk pengelompokan siswa SMP terbaik.

Kata kunci: *Clustering, DBSCAN, K-Means.*



UNIVERSITAS ISLAM NEGERI

SUMATERA UTARA MEDAN

ABSTRACT

Education is a crucial aspect of developing quality human resources. Evaluating student performance is critical in determining the best students. *Clustering* methods are techniques that can be used to group student performance data based on various criteria, thus facilitating the assessment process. This thesis examines the performance of two *clustering* algorithms, *DBSCAN* and *K-Means*, in identifying the best junior high school (SMP) students. The research aims to determine which algorithm is more efficient and effective in *clustering* SMP student data based on their academic achievements. Data collection methods for this study involved observation and interviews. The research findings indicate that the *DBSCAN* algorithm with epsilon 11 and a sample of 24 data points resulted in 2 clusters with a Silhouette Score of 0.258030877243884. Meanwhile, the *K-Means* algorithm using the Elbow method produced 4 clusters with a Silhouette Score of 0.5697019340266847. From this comparison, the *K-Means* algorithm proved to be more efficient. *K-Means* also showed cleaner, more structured clusters with better similarity among data points, making it preferable for determining the best SMP students.

Keywords: *Clustering, DBSCAN, K-Means*



UNIVERSITAS ISLAM NEGERI

SUMATERA UTARA MEDAN

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Puji syukur kepada Allah SWT berkat Rahmat, Hidayah, dan Karunia-Nya kepada kita semua sehingga peneliti dapat menyelesaikan skripsi dengan judul “Perbandingan Algoritma *Clustering DBSCAN* Dan *K-Means* Dalam Pengelompokan Siswa Terbaik”. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan program sarjana di program studi Ilmu Komputer, Fakultas Sains dan Teknologi, UIN Sumatera Utara.

Peneliti menyadari dalam penyusunan skripsi ini tidak akan selesai tanpa bantuan dari berbagai pihak. Karena itu pada kesempatan ini peneliti mengucapkan terima kasih kepada:

1. Ibu Prof. Dr. Nurhayati, M.Ag, selaku Rektor UIN Sumatera Utara Medan beserta staf-stafnya yang telah memberikan berbagai macam fasilitas selama menjalankan perkuliahan.
2. Bapak Dr. Zulham, S. H, M. Hum., selaku Dekan Fakultas Sains dan Teknologi UIN Sumatera Utara Medan, beserta staf-stafnya.
3. Bapak Ilka Zufria, M. Kom selaku Ketua Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi, UIN Sumatera Utara Medan.
4. Bapak Dr. M. Fakhriza, S.T, M. Kom selaku Sekretaris Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi, UIN Sumatera Utara Medan.
5. Bapak Muhammad Siddik Hasibuan, M. Kom selaku dosen pembimbing I yang telah memberikan banyak kritikan, saran, motivasi, dan bimbingan kepada penulis selama proses penyelesaian skripsi ini.
6. Bapak Aidil Halim Lubis, M. Kom selaku dosen pembimbing II yang telah memberikan banyak kritikan, saran, motivasi, dan bimbingan kepada penulis selama proses penyelesaian skripsi ini.
7. Bapak Abdul Halim Hasugian, M. Kom selaku dosen pembimbing akademik yang telah memberikan banyak masukan dan motivasi dari awal perkuliahan hingga akhir.

8. Segenap Dosen Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi, UIN Sumatera Utara Medan yang tidak bisa peneliti sebutkan satu persatu yang telah memberikan ilmunya kepada peneliti.
9. Keluarga tercinta terutama orang tua peneliti, Ayah tercinta yang sangat dirindukan Alm.Dahler Lubis, S.Pd dan Ibu tercinta yang paling saya sayangi Faridah Hannum, S.Pd yang telah mendidik, memberikan dukungan, dan tanpa henti berdoa untuk anaknya tercinta. Dan empat saudara peneliti yang juga turut memberikan dukungan, Rachmat Suhendra Lubis, S.Pi, Sri Yanti Aisyah Lubis, M.E, Ridwan Harun Lubis, S.Kep dan Khaliq Ahmad Lubis.
10. Pemilik NIM 21060013, terima kasih telah menjadi penyemangat, pendengar keluh kesah, memberikan motivasi, dan penasehat yang baik serta support system dalam penulisan skripsi.

Peneliti menyadari skripsi ini tidak luput dari berbagai kekurangan, pada teknis penulisan ataupun dalam materi. Peneliti mengharapkan saran dan kritik demi kesempurnaan dan perbaikannya sehingga skripsi ini dapat memberikan manfaat bagi masyarakat, menjadi sumbangsi pemikiran bagi pihak yang membutuhkan serta bisa dikembangkan lagi lebih lanjut.

Medan, 30 September 2023

UNIVERSITAS ISLAM NEGERI
SUMATERA UTARA MEDAN

Melda Novita Sari
NIM. 0701193181

DAFTAR ISI

ABSTRAK	i
ABSTRACT	ii
KATA PENGANTAR	iii
DAFTAR ISI	v
DAFTAR GAMBAR	viii
DAFTAR TABEL	ix
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	5
1.4 Tujuan Masalah.....	5
1.5 Manfaat Penelitian.....	5
BAB II TINJAUAN PUSTAKA	6
2.1 Data Mining	6
2.2 Analisis Kluster (Cluster Analysis)	9
2.2.1 Metode Berbasis Partisi (Partitioning Method).....	10
2.2.2 Metode Berbasis Hirarki (Hierarchical Method).....	10
2.2.3 Metode Berbasis Kepadatan (Density-Based Method).....	11
2.2.4 Metode Berbasis Kisi (Grid-Based Method).....	11
2.3 <i>K-Means Clustering</i>	12
2.4 <i>DBSCAN Clustering</i>	15
2.5 Normalisasi	17

2.6	Min Max Scaller.....	18
2.7	Python.....	19
2.8	Indeks Silhouette.....	20
2.9	<i>Flowchart</i>	Error! Bookmark not defined.
2.10	Penelitian Terdahulu.....	23
BAB III METODOLOGI PENELITIAN		30
3.1	Tempat dan Waktu Penelitian.....	30
3.1.1	Tempat Penelitian.....	30
3.1.2	Waktu Penelitian.....	30
3.2	Bahan dan Alat Penelitian.....	30
3.2.1	Perangkat Keras.....	30
3.2.2	Perangkat Lunak.....	30
3.3	Kerangka Kerja Penelitian.....	31
3.3.1	Perencanaan Penelitian.....	31
3.3.2	Studi Literatur.....	31
3.3.3	Pengumpulan Data.....	32
3.3.4	Perancangan Sistem.....	34
3.3.5	Pengujian.....	38
BAB IV HASIL DAN PEMBAHASAN.....		39
4.1	Representasi Data.....	39
4.1.1	Data Preprocessing.....	39
4.1.2	Min Max.....	40
4.1.3	Normalisasi Data (Feature Scaling).....	41
4.2	Penerapan Algoritma <i>DBSCAN</i>	43
4.2.1	Implementasi <i>DBSCAN</i>	50

4.3	Uji Validasi <i>DBSCAN</i>	51
4.4	<i>K-Means Clustering</i>	52
4.5	Implementaasi Algoritma <i>K-Means</i>	55
4.6	Uji Validasi <i>DBSCAN K-Means</i>	57
4.7	Perbandingan Algoritma <i>DBSCAN</i> dan <i>K-Means</i>	58
BAB V KESIMPULAN DAN SARAN		57
5.1	KESIMPULAN	57
5.2	SARAN.....	57
DAFTAR PUSTAKA		58



UNIVERSITAS ISLAM NEGERI
SUMATERA UTARA MEDAN

DAFTAR GAMBAR

Gambar 2. 1 Skema Knowledge Discovery in Database (KDD)	6
Gambar 2. 2 Diagram Hubungan Data Mining (K. C. Gouda)	7
Gambar 2. 3 Visualisasi <i>K-Means</i> (Ryan Reza, 2020)	12
Gambar 2. 4 Algoritma pengelompokan	16
Gambar 2. 5 Kasus <i>DBSCAN</i> (Inayah dkk. 2022).....	17
Gambar 3. 1 Kerangka Kerja Penelitian	31
Gambar 3. 2 <i>Flowchart</i> Model Penelitian	35
Gambar 3. 3 <i>Flowchart</i> Metode <i>DBSCAN</i>	36
Gambar 3. 4 <i>Flowchart</i> Metode <i>K-Means</i>	37
Gambar 4. 1 Import Library Dan Memanggil Data.....	42
Gambar 4. 2 Proses normalisasi data dengan <i>StandardScaler</i>	42
Gambar 4. 3 Hasil normalisasi data.....	43
Gambar 4. 4 <i>NearestNeighbors</i>	50
Gambar 4. 5 Grafik hasil <i>NearestNeighbors</i>	50
Gambar 4. 6 Jumlah Klaster <i>DBSCAN Clustering</i>	51
Gambar 4. 7 Implementasi <i>Silhouette Score</i> pada <i>DBSCAN</i>	51
Gambar 4. 8 Implementasi Algoritma <i>K-Means</i>	56
Gambar 4. 9 Grafik Hasil Implementasi.....	56
Gambar 4. 10 Menampilkan hasil <i>K- Means Clustering</i>	57
Gambar 4. 11 Grafik hasil <i>K-Means Clustering</i>	57
Gambar 4. 12 Uji Validasi <i>Silhouette</i> pada <i>K-Means Clustering</i>	57

DAFTAR TABEL

Tabel 2. 1 Kasus <i>K-Means clustering</i>	14
Tabel 2. 2 Hasil Pengelompokan Data Dengan Centroid	15
Tabel 2. 3 Tabel Keterangan Nilai Silhoutte	21
Tabel 2. 4 Simbol <i>Flowchart</i>	21
Tabel 2. 5 Keuntungan dan Kekurangan <i>Flowchart</i>	23
Tabel 2. 6 Penelitian Terdahulu	23
Tabel 3. 1 Waktu Penelitian.....	30
Tabel 3. 2 Data hasil wawancara dan observasi.....	32
Tabel 3. 3 Data siswa hasil wawancara kelas VII-I	33
Tabel 3. 4 Keterangan Format ke Numerik	33
Tabel 4. 1 Data Hasil Praprocessing.	40
Tabel 4. 2 Data Hasil MinMax.....	41
Tabel 4. 3 Data Jarak Iterasi I	46
Tabel 4. 4 Data Jarak Iterasi II.....	47
Tabel 4. 5 Data Jarak Iterasi III.....	48
Tabel 4. 6 Data Jarak Iterasi IV	48
Tabel 4. 7 Data Jarak Iterasi V.....	49
Tabel 4. 8 Contoh Data.....	52
Tabel 4. 9 Centroid Klaster yang Terpilih	52
Tabel 4. 10 Iterasi 2	55