

TEKNIK KLAUSTERING

UNTUK DATA MINING

BIG DATA



PENULIS

SANDI ARDIANSYAH

SUTRISNO

AKBAR SANJAYA RAMBE

RAMADHAN ADY PRATAMA

MUHAMMAD FUAD HAKIKI

SAYUTI RAHMAN

HARTONO

ALGORITMA :

K-MEANS, K-MEDOIDS (PAM), HIERARCHICAL CLUSTERING (AGGLOMERATIVE/DIVISIVE), DBSCAN, OPTICS, GAUSSIAN MIXTURE MODEL (GMM), MEAN-SHIFT CLUSTERING, SELF-ORGANIZING MAP (SOM / KOHONEN), SPECTRAL CLUSTERING.

TEKNIK KLASTERING UNTUK DATA MINING

Penulis

SANDI ARDIANSYAH

SUTRISNO

AKBAR SANJAYA RAMBE

RAMADHAN ADY PRATAMA

MUHAMMAD FUAD HAKIKI

SAYUTI RAHMAN

HARTONO

DITERBITKAN OLEH:

UNIVERSITAS MEDAN AREA PRESS

TEKNIK KLASTERING UNTUK DATA MINING

Penulis : Sandi Ardiansyah
Sutrisno, ST, MT
Akbar Sanjaya Rambe
Ramadhan Ady Pratama
Muhammad Fuad Hakiki
Dr. Sayuti Rahman, ST, M.Kom
Dr. Hartono, S.Kom, M.Kom

Desain Cover: Ramadhan Ady Pratama
Edit Layout : Dr. Sayuti Rahman, ST, M.Kom
Editor : Yuan Nisa, S.Si, M.Si

ISBN:

Copyright 2025
Uk: 21 x 29,7 cm hlm.121
Copyright 2025

Universitas Medan Area Press

**Address: Jalan Kolam Nomor 1, Kenangan Baru, Kec. Percut Sei Tuan,
Deliserdang, Sumatera Utara
Telephone:061-7366878,
e-mail: pghc@uma.ac.id**

KATA PENGANTAR

Puji syukur ke hadirat Tuhan Yang Maha Esa, karena atas rahmat dan karunia-Nya buku berjudul “Teknik Klastering untuk Data Mining” ini dapat disusun dan diselesaikan dengan baik. Buku ini disusun sebagai upaya untuk memberikan pemahaman yang komprehensif, sistematis, dan aplikatif mengenai teknik klastering sebagai salah satu metode utama dalam *data mining* dan *unsupervised learning*.

Perkembangan pesat teknologi informasi serta meningkatnya ketersediaan data dalam berbagai bidang telah mendorong kebutuhan akan metode analisis data yang mampu mengekstraksi pola dan struktur tersembunyi secara efektif. Klastering merupakan pendekatan fundamental yang banyak digunakan dalam berbagai sektor, mulai dari pendidikan, kesehatan, ekonomi, bisnis, hingga teknologi dan kecerdasan buatan. Oleh karena itu, pemahaman yang kuat terhadap konsep, algoritma, serta implementasi teknik klastering menjadi kompetensi penting bagi mahasiswa, peneliti, maupun praktisi.

Buku ini membahas berbagai algoritma klastering secara terstruktur, dimulai dari metode berbasis partisi seperti K-Means dan K-Medoids, dilanjutkan dengan Hierarchical Clustering, DBSCAN, OPTICS, Gaussian Mixture Model, Mean-Shift, Self-Organizing Map (SOM), hingga Spectral Clustering. Setiap bab disusun dengan pendekatan bertahap yang mencakup landasan teori, kelebihan dan keterbatasan metode, perhitungan, serta contoh implementasi.

Penyusunan buku ini ditujukan sebagai bahan ajar, referensi akademik, serta panduan praktis bagi mahasiswa, dosen, dan pemerhati data mining dan machine learning. Penulis berharap buku ini dapat menjadi kontribusi dalam mendukung pengembangan literasi data dan riset berbasis analisis klaster di lingkungan akademik maupun profesional.

Akhir kata, penulis menyadari bahwa buku ini masih memiliki keterbatasan dan terbuka terhadap penyempurnaan. Oleh karena itu, saran dan masukan yang konstruktif sangat diharapkan demi perbaikan pada edisi selanjutnya. Semoga buku ini memberikan manfaat dan menjadi referensi yang berguna.

Medan, Februari 2026

Penulis

Sinopsis

Buku Teknik Klustering untuk Data Mining menyajikan pembahasan komprehensif dan sistematis mengenai metode klustering sebagai salah satu teknik utama dalam data mining dan unsupervised learning. Buku ini dirancang untuk membantu pembaca memahami bagaimana pola dan struktur tersembunyi dalam data dapat diidentifikasi melalui proses pengelompokan tanpa label, yang semakin penting di era big data dan kecerdasan buatan.

Pembahasan diawali dengan pengenalan konsep dasar klustering dan algoritma K-Means, mencakup sejarah, prinsip kerja, kelebihan, keterbatasan, perhitungan manual, hingga contoh implementasi. Selanjutnya, buku ini mengulas K-Medoids (PAM) sebagai pengembangan K-Means yang lebih robust terhadap outlier, diikuti dengan Hierarchical Clustering yang menekankan struktur bertingkat dan interpretasi dendrogram. Metode berbasis kepadatan seperti DBSCAN dan OPTICS juga dibahas untuk menangani data dengan bentuk kluster tidak beraturan dan keberadaan noise.

Pada bagian lanjutan, buku ini mengupas metode klustering berbasis probabilistik dan nonlinier, seperti Gaussian Mixture Model (GMM) dengan algoritma Expectation-Maximization, Mean-Shift Clustering, serta Self-Organizing Map (SOM/Kohonen) yang terinspirasi dari mekanisme biologis. Bab terakhir membahas Spectral Clustering dengan pendekatan teori graf dan matriks Laplacian, yang cocok untuk data berdimensi tinggi dan struktur kompleks.

Setiap bab disusun secara bertahap dengan landasan teori, formulasi matematis, kelebihan dan kekurangan metode, contoh perhitungan, studi kasus, serta panduan interpretasi hasil. Dengan pendekatan yang konseptual sekaligus aplikatif, buku ini ditujukan sebagai bahan ajar, referensi akademik, dan panduan praktis bagi mahasiswa, dosen, peneliti, serta praktisi yang ingin memperdalam pemahaman dan penerapan teknik klustering dalam berbagai bidang seperti pendidikan, kesehatan, bisnis, dan teknologi informasi

DAFTAR ISI

	Halaman
KATA PENGANTAR.....	iii
DAFTAR ISI	iv
BAB I. ALGORITMA K-MEANS	1
1.1 Pendahuluan.....	1
1.1.1. Sejarah Klustering.....	1
1.1.2. Munculnya Algoritma <i>K-Means</i>	2
1.2 Algoritma K-Means	2
1.2.1. Apa Itu Algoritma <i>K-Means</i> ?.....	2
1.2.2. Fungsi Algoritma <i>K-Means</i>	3
1.2.3. Proses Algoritma <i>K-Means</i>	4
1.2.4. Kelebihan Algoritma <i>K-Means</i>	5
1.2.5. Kekurangan Algoritma <i>K-Means</i>	6
1.3 Perhitungan Algoritma <i>K-Means</i>	7
1.4 Implementasi Algoritma <i>K-Means</i>	8
1.5 Kesimpulan	10
BAB II. ALGORITMA K-MEDOIDS (PAM).....	12
2.1 Pendahuluan.....	12
2.1.1. Sejarah Algoritma <i>K-Medoids</i> (PAM).....	12
2.1.2. Munculnya Algoritma <i>K-Medoids</i>	12
2.1.3. Algoritma <i>K-Medoids</i>	13
2.1.4. Apa Itu Algoritma <i>K-Medoids</i>	14
2.1.5. Fungsi Algoritma <i>K-Medoids</i>	15
2.1.6. Proses Algoritma <i>K-Medoids</i>	15
2.1.7. Kelebihan Algoritma <i>K-Medoids</i>	16
2.1.8. Kekurangan Algoritma <i>K-Medoids</i>	17
2.2 Perhitungan Algoritma <i>K-Medoids</i>	18
2.2.1. Langkah-Langkah Algoritma <i>K-Medoids</i>	19

2.2.2. Diagram Alur (Flowchart) Algoritma K-Medoids.....	20
2.3 Implementasi Algoritma K- <i>Medoids</i>	21
2.4 Kesimpulan	22
BAB III. HIERARCHICAL CLUSTERING	24
3.1. Pendahuluan	24
3.2. Tujuan <i>Hierarchical Clustering</i>	25
3.3. Defenisi <i>Hierarchical Clustering</i>	26
3.4. Metode <i>Hierarchical Clustering</i>	28
3.5. Impelementasi Metode	29
3.6. Contoh Kasus	31
3.7. Penutup.....	33
BAB IV. DBSCAN (DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE)	35
4.1. Pendahuluan.....	35
4.2. Tujuan dan Capaian.....	36
4.3. Konsep Dasar Clustering dalam Analisis Data	38
4.4. Jenis-Jenis Metode Clustering.....	39
4.5. Konsep Kepadatan dalam Clustering	40
4.6. Konsep Jarak dan Neighborhood	40
4.7. Metode.....	41
4.8. Impelementasi Metode	43
4.9. Contoh Kasus	45
4.10. Penutup.....	47
BAB V. ORDERING POINTS TO IDENTIFY THE CLUSTERING	50
5.1. Pendahuluan.....	50
5.2. Metode (Algoritma)	52
5.2.1. Konsep Dasar	52
5.2.2. Parameter	52
5.2.3. Langkah Algoritma	53
5.2.4. Output dan Interpretasi	54
5.2.5. Kelebihan dan Kekurangan OPTICS	56

5.3. Contoh Hitungan Algoritma OPTICS	58
5.4. Penelitian Terkait	60
5.5. Penutup.....	62
BAB VI. GAUSSIAN MIXTURE MODEL (GMM)	64
6.1 Pendahuluan	64
6.2. Konsep Dasar Gaussian Distribution	65
6.2.1. Definisi distribusi Gaussian	65
6.2.2. Visualisasi kurva Gaussian.....	66
6.2.3. Definisi GMM sebagai kombinasi beberapa Gaussian	66
6.2.4. Fungsi Likelihood dan Parameter GMM.....	67
6.3 Algoritma Expectation-Maximization (EM).....	68
6.3.1. Formulasi Dasar EM	68
6.3.2. Tahapan Algoritma EM.....	68
6.3.3. Tahapan Algoritma EM.....	69
6.4. Kelebihan dan Kekurangan	71
6.4.1. Kelebihan.....	71
6.4.2. Kekurangan	72
6.5. Studi Kasus.....	72
BAB VII. MEAN-SHIFT CLUSTERING	75
7.1. Pendahuluan	75
7.2. Konsep Dasar	76
7.2.1. Definisi kernel density estimation (KDE).....	76
7.2.3. Langkah-langkah algoritma.....	78
7.2.4. Parameter.....	79
7.3. Kelebihan dan Kekurangan Mean-Shift.....	80
7.3.1. Kelebihan.....	80
7.3.2. Kekurangan	80
7.4. Ringkasan	81
BAB VIII. SELF-ORGANIZING MAP (SOM/KOHONEN)	82
8.1. Pendahuluan	82

8.2. Konsep Dasar dan Arsitektur SOM.....	83
8.2.1. Inspirasi Biologis dan Prinsip Pembelajaran.....	83
8.2.2. Inspirasi Biologis dan Prinsip Pembelajaran.....	83
8.2.3. Landasan Teori: Kuantisasi Vektor dan Voronoi Tessellation	83
8.2.4. Mekanisme Interaksi Lateral: Fungsi "Mexican Hat"	84
8.3. Algoritma Self-Organizing Map	85
8.3.1. Detail Matematis: Fungsi Peluruhan (Decay Functions)	86
8.3.2. Varian Algoritma: Batch SOM.....	87
8.3.3. Simulasi Perhitungan Manual: Studi Kasus Sederhana	87
8.4. Visualisasi dan Interpretasi Kluster	89
8.4.1. U-Matrix (Unified Distance Matrix)	89
8.4.2. Component Planes.....	90
8.5. Evaluasi Kualitas SOM: Metrik Kuantitatif.....	90
8.5.1. Kesalahan Kuantisasi (Quantization Error - Qe)	90
8.5.2. Kesalahan Topologi (Topographic Error - Te)	90
8.6. Kelebihan dan Kekurangan SOM.....	91
8.7. Isu Praktis dalam Implementasi	92
8.7.1. Isu Praktis dalam Implementasi	92
8.7.2. Efek Tepi (Border Effect).....	92
8.8. Studi Kasus: Segmentasi Pelanggan.....	92
8.8.1. Tahap Pra-pemrosesan Data: Normalisasi dan Pembobotan.....	93
8.8.2. Tahap Pra-pemrosesan Data: Normalisasi dan Pembobotan.....	94
8.9. Kesimpulan.....	95
BAB IX. SPECTRAL CLUSTERING.....	96
9.1. Pendahuluan	96
9.2 Fondasi Teori Graf	96
9.2.1. Adjacency Matrix dan Weight Matrix.....	97
9.2.2. Degree Matrix.....	97
9.3 Matriks Laplacian Graf	97
9.3.1 Unnormalized Graph Laplacian	98

9.3.2 Normalized Graph Laplacians.....	98
9.4 Algoritma Spectral Clustering.....	99
9.4.1 Algoritma Spectral Clustering.....	100
9.5 Perspektif Graph Cut.....	100
9.5.1 RatioCut.....	101
9.5.2 Normalized Cut (Ncut).....	101
9.6 Pemilihan Parameter dan Tantangan.....	101
9.6.1. Pemilihan Parameter dan Tantangan.....	102
9.6.1. Pemilihan Skala Lokal (Local Scaling).....	102
9.7. Teori Perturbasi dan Stabilitas Spektral.....	103
9.8. Interpretasi Probabilistik: Random Walks.....	103
9.8.1. Jarak Komut (Commute Distance).....	104
9.9. Hubungan dengan Kernel PCA.....	105
DAFTAR PUSTAKA.....	107