

PERBANDINGAN KINERJA *PARTITION-BASED* DAN *DENSITY-BASED CLUSTERING* PADA DATA MENGANDUNG *OUTLIER*

DELITA NUR HASANAH



PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA MATEMATIKA DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Perbandingan Kinerja *Partition-Based* dan *Density-Based Clustering* pada Data Mengandung *Outlier*” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dalam penyusunan karya tugas akhir ini, penulis memanfaatkan teknologi kecerdasan buatan, yaitu ChatGPT, sebagai alat bantu dalam penyelarasan kalimat, peninjauan struktur paragraf, serta perbaikan kode analisis. Seluruh hasil yang diperoleh dari penggunaan teknologi tersebut telah ditelaah, disunting, dan disesuaikan kembali oleh penulis secara mandiri agar memenuhi kaidah akademik. Penulis bertanggung jawab sepenuhnya atas keabsahan, orisinalitas, serta keseluruhan isi karya tugas akhir ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2026

Delita Nur Hasanah
G1401221018



ABSTRAK

DELITA NUR HASANAH. Perbandingan Kinerja *Partition-Based* dan *Density-Based Clustering* pada Data Mengandung *Outlier*. Dibimbing oleh HARI WIJAYANTO dan LA ODE ABDUL RAHMAN.

Keberadaan *outlier* merupakan salah satu tantangan dalam analisis *clustering* karena dapat memengaruhi struktur data dan menurunkan ketepatan hasil pengelompokan. Penelitian ini bertujuan membandingkan kinerja metode *partition-based clustering* (K-Means dan K-Medoids) dengan *density-based clustering* (DBSCAN) pada data yang mengandung *outlier*, serta menerapkannya pada data Indeks Kualitas Lingkungan Hidup (IKLH) kabupaten/kota di Pulau Jawa. Data simulasi dibangkitkan menggunakan distribusi normal multivariat melalui 12 skenario yang mengombinasikan kondisi *overlap* dan *non-overlap*, jumlah observasi sebanyak 100 dan 500, serta proporsi *outlier* sebesar 0%, 5%, dan 10%, dengan 100 ulangan pada setiap skenario. Kinerja metode dievaluasi menggunakan *euclidean distance* antara pusat *cluster* hasil *clustering* dan pusat *cluster* sebenarnya. Pada data empiris, evaluasi dilakukan menggunakan *Silhouette Index*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index*. Hasil penelitian menunjukkan bahwa peningkatan proporsi *outlier* meningkatkan nilai *euclidean distance* pada seluruh metode, dengan K-Means sebagai metode yang paling sensitif terhadap *outlier*. Sebaliknya, DBSCAN secara konsisten menghasilkan *euclidean distance* terendah pada data simulasi serta memperoleh nilai *Silhouette Index* tertinggi dan *Davies-Bouldin Index* terendah pada data empiris. Hasil tersebut menunjukkan bahwa DBSCAN merupakan metode yang lebih *robust* terhadap keberadaan *outlier* dan mampu menghasilkan kualitas *clustering* yang lebih baik dibandingkan K-Means dan K-Medoids.

Kata kunci: DBSCAN, K-Means, K-Medoids, pencilan, penggerombolan.



ABSTRACT

DELITA NUR HASANAH. Performance Comparison of Partition-Based and Density-Based Clustering Methods on Data Containing Outliers. Supervised by HARI WIJAYANTO and LA ODE ABDUL RAHMAN.

Outliers are one of the major challenges in cluster analysis because they can distort the data structure and reduce clustering accuracy. This study aims to compare the performance of partition-based clustering methods (K-Means and K-Medoids) and the density-based clustering method (DBSCAN) on data containing outliers and to apply these methods to the Environmental Quality Index (EQI) data of regencies and municipalities in Java Island. Simulation data were generated from a multivariate normal distribution under 12 scenarios combining overlap and non-overlap cluster conditions, sample sizes of 100 and 500 observations, and outlier proportions of 0%, 5%, and 10%, with 100 replications for each scenario. The clustering performance was evaluated using the Euclidean distance between the estimated cluster centers and the true cluster centers. For the empirical data, clustering quality was assessed using the Silhouette Index, Davies–Bouldin Index, and Calinski–Harabasz Index. The results indicate that increasing the proportion of outliers generally increases the Euclidean distance for all clustering methods, with K-Means being the most sensitive to outliers. In contrast, DBSCAN consistently produced the lowest Euclidean distance in the simulation study and achieved the highest Silhouette Index and the lowest Davies–Bouldin Index in the empirical analysis. These findings demonstrate that DBSCAN is more robust to the presence of outliers and provides better clustering performance than K-Means and K-Medoids.

Keywords: clustering, DBSCAN, outliers, K-Means, K-Medoids.



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026¹
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

PERBANDINGAN KINERJA PARTITION-BASED DAN DENSITY-BASED *CLUSTERING* PADA DATA MENGANDUNG *OUTLIER*

DELITA NUR HASANAH

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA MATEMATIKA DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Skripsi:
1 Dr. Anwar Fitrianto, S.Si., M.Sc

Perpustakaan IPB University



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Judul Skripsi : Perbandingan Kinerja *Partition-Based* dan *Density-Based Clustering* pada Data Mengandung *Outlier*

Nama : Delita Nur Hasanah
NIM : G1401221018

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:
Prof. Dr. Ir. Hari Wijayanto, M.Si.

Pembimbing 2:
La Ode Abdul Rahman, S.Si., M.Si.

Diketahui oleh

Ketua Program Studi:
Dr. Bagus Sartono, SSi, MSi
NIP 197804112005011002

Tanggal Ujian:
15 Juli 2026

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga penelitian yang berjudul “Perbandingan Kinerja *Partition-Based* dan *Density-Based Clustering* pada Data Mengandung *Outlier*”. Penulisan karya ini bisa diselesaikan karena dukungan, bantuan, motivasi, dan bimbingan dari berbagai pihak. Pada kesempatan ini, penulis menyampaikan banyak terima kasih kepada:

1. Prof. Dr. Ir. Hari Wijayanto, M.Si. dan Pak Laode Abdul Rahman, MSi, selaku dosen pembimbing yang senantiasa bersedia meluangkan waktunya untuk memberikan bimbingan, masukan, dukungan, dan arahan kepada penulis.
2. Dr. Anwar Fitrianto, S.Si., M.Sc selaku penguji di luar komisi pembimbing.
3. Seluruh dosen pengajar Program Studi Statistika dan Sains Data atas ilmu yang telah diberikan
4. Keluarga yang selalu memberikan dukungan dan doa selama berkuliah.
5. STK Ngantuk, Shalma, Afi, Haidar, Deden, Iyol, Mukni, Rai, Apdan. Teman dari tingkat satu, Arla. Fasil RSH 2026, Ikhda, Aenur, Atim. Teman dari kecil, Jia, Caca, Deden. Serta seluruh Marinestic dan teman-teman yang mendukung dan memberikan warna di perkuliahan.
6. Bang Nabil dan Bang Much yang sudah membantu menjelaskan metode yang digunakan pada skripsi ini.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan. Penulis menyadari masih ada kesalahan dan kekurangan dalam penyusunan penelitian ini, untuk itu penulis memohon maaf.

Bogor, Agustus 2026

Delita Nur Hasanah



DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	3
II TINJAUAN PUSTAKA	4
2.1 Analisis <i>Clustering</i>	4
2.1.1 K-Means	4
2.1.2 K-Medoids	5
2.1.3 DBSCAN	5
2.2 Metrik Evaluasi <i>Clustering</i>	6
2.2.1 <i>Silhouette Index</i>	6
2.2.2 <i>Davies-Bouldin Index</i>	7
2.2.3 Calinski-Harabaz Index	8
2.2.4 <i>Euclidean Distance</i>	8
2.3 <i>Outlier</i>	9
III METODE	10
3.1 Data	10
3.1.1 Data Simulasi	10
3.1.2 Data Empiris	10
3.2 Prosedur Analisis Data Simulasi	11
3.3 Prosedur Analisis Data Empiris	12
IV HASIL DAN PEMBAHASAN	13
4.1 Kajian Data Simulasi	13
4.1.1 Pembangkitan Data	13
4.1.2 Evaluasi <i>Clustering</i>	14
4.2 Kajian Data Empiris	18
4.2.1 Eksplorasi Data	18
4.2.2 <i>Clustering</i> Metode K-Means	20
4.2.3 Interpretasi Hasil <i>Clustering</i> K-Means	20
V SIMPULAN DAN SARAN	23
5.1 Simpulan	23
5.2 Saran	23
DAFTAR PUSTAKA	24



DAFTAR TABEL

1	Skenario data simulasi	10
2	Peubah yang digunakan	10
3	Nilai validitas internal metode K-Means	20
4	Rata-rata peubah pada setiap <i>cluster</i>	21
5	Anggota pada setiap <i>cluster</i>	22

DAFTAR GAMBAR

1	Visualisasi contoh salah satu data simulasi pada 500 observasi	13
2	Visualisasi contoh salah satu data simulasi pada 100 observasi	14
3	Pengaruh <i>outlier</i> pada kondisi <i>overlap</i> dan observasi 100	15
4	Pengaruh <i>outlier</i> pada kondisi <i>overlap</i> dan observasi 500	16
5	Pengaruh <i>outlier</i> pada kondisi <i>non overlap</i> dan observasi 100	17
6	Pengaruh <i>outlier</i> pada kondisi <i>non overlap</i> dan observasi 500	17
7	Matriks korelasi antar peubah	18
8	Boxplot pada setiap peubah	19
9	Plot PCA	19
10	Evaluasi <i>clustering</i> K-Means	20
11	Pemetaan hasil <i>clustering</i> K-Means	21