

**MENGUKUR KINERJA *CLUSTERING* MENGGUNAKAN
ALGORITMA *K-MEANS* DAN *K-MEDOIDS*
DENGAN OPTIMASI *SILHOUETTE COEFFICIENT***

Diajukan Sebagai Syarat untuk Menyelesaikan
Pendidikan Program Strata-1 pada
Jurusan Teknik Informatika



Oleh :

Karinda Amelia

NIM : 09021282025054

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2024**

LEMBAR PENGESAHAN SKRIPSI

**MENGUKUR KINERJA *CLUSTERING* MENGGUNAKAN
ALGORITMA *K-MEANS* DAN *K-MEDOIDS*
DENGAN OPTIMASI *SILHOUETTE COEFFICIENT***

Oleh :

Karinda Amelia

NIM : 09021282025054

Palembang, 19 Juni 2024


Mengetahui,

Ketua Jurusan Teknik Informatika



Dr. M. Fachrudrozi, S.Si., M.T.
NIP 198005222008121002

Pembimbing



Dian Palupi Rini, M.Kom., Ph.D.

NIP 197802232006042002

TANDA LULUS UJIAN KOMPREHENSIF


Pada hari Rabu tanggal 5 Juni 2024 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Karinda Amelia
NIM : 09021282025054
Judul : Mengukur Kinerja *Clustering* Menggunakan Algoritma *K-Means* dan *K-Medoids* dengan Optimasi *Silhouette Coefficient*

dan dinyatakan **LULUS**.

1. Ketua Penguji


Desty Rodiah, M.T.
NIP 198912212020122011



.....

2. Penguji

Yunita, M.Cs.
NIP 198306062015042002



.....

3. Pembimbing

Dian Palupi Rini, M.Kom., Ph.D.
NIP 197802232006042002



.....

Mengetahui,
Ketua Jurusan Teknik Informatika



Dr. M. Fachrurrozi, S.Si., M.T.
NIP 198005222008121002

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Karinda Amelia
NIM : 09021282025054
Program Studi : Teknik Informatika
Judul Skripsi : Mengukur Kinerja *Clustering* Menggunakan Algoritma
K-Means dan *K-Medoids* dengan Optimasi *Silhouette*
Coefficient

Hasil Pengecekan Software *iThenticate/Turnitin* : 12%

Menyatakan bahwa laporan skripsi saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan skripsi ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, 19 Juni 2024



Karinda Amelia

NIM 09021282025054

MOTTO DAN PERSEMBAHAN

“You don't have to be great to start, but you have to start to be great.”

~ **Zig Ziglar**

“Life isn't about finding yourself. Life is about creating yourself.”

~ **George Bernard Shaw**

Kupersembahkan Karya Tulis Ini Kepada :

- Papa dan Mama Tercinta
- Diri Sendiri
- Adik dan Kakak Tersayang
- Keluarga Besar
- Dosen Pembimbing dan Penguji
- Sahabat dan Teman Seperjuangan
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

**MEASURING CLUSTERING PERFORMANCE USING
K-MEANS AND K-MEDOIDS ALGORITHMS
WITH SILHOUETTE COEFFICIENT OPTIMIZATION**

By :
Karinda Amelia
09021282025054

ABSTRACT

Clustering is one of the most widely used techniques in data processing that can be used to find patterns and characteristics in data. K-Means and K-Medoids are the most widely used clustering algorithms because they are simple, easy to implement, and relatively fast. Although proven to have good performance with high accuracy and efficiency for large data, previous studies have shown differences in clustering results between the two algorithms. Therefore, this study measures and compares the performance of K-Means and K-Medoids by applying Silhouette Coefficient optimization. Silhouette Coefficient is used to determine the optimal number of clusters based on the average silhouette value to optimize clustering performance. Tests were conducted by applying the K-Means, K-Medoids, and Silhouette Coefficient algorithms on three different datasets obtained from Kaggle in the range of 2 to 10 clusters. Evaluation of the clustering performance was done using Silhouette Score. The results show that the optimal number of clusters for all datasets is 2, both in the K-Means and K-Medoids algorithms. However, the K-Means algorithm shows better performance than K-Medoids with higher Silhouette Score values and close to 1. The highest Silhouette Score values of K-Means in the three datasets are 0.7099, 0.7674, and 0.7716, respectively. The K-Medoids are 0.7064, 0.7445, and 0.7257.

Keywords : *Clustering, K-Means, K-Medoids, Silhouette Coefficient, Silhouette Score*

Palembang, 19 June 2024

Approved,
Head of Informatics Department

Supervisor



Dian Palupi Rini, M.Kom., Ph.D.
NIP 197802232006042002

**MENGUKUR KINERJA *CLUSTERING* MENGGUNAKAN
ALGORITMA *K-MEANS* DAN *K-MEDOIDS*
DENGAN OPTIMASI *SILHOUETTE COEFFICIENT***

Oleh :
Karinda Amelia
09021282025054

ABSTRAK

Clustering menjadi salah satu teknik yang paling banyak digunakan dalam pengolahan data yang dapat digunakan untuk menemukan pola dan karakteristik pada data. *K-Means* dan *K-Medoids* menjadi algoritma *clustering* yang paling banyak digunakan karena sederhana, mudah diimplementasikan, dan relatif cepat. Meskipun terbukti memiliki kinerja yang baik dengan akurasi tinggi dan efisiensi untuk data besar, penelitian terdahulu menunjukkan perbedaan hasil *clustering* antara kedua algoritma tersebut. Oleh karena itu, penelitian ini melakukan pengukuran dan perbandingan kinerja *K-Means* dan *K-Medoids* dengan menerapkan optimasi *Silhouette Coefficient*. *Silhouette Coefficient* digunakan untuk menentukan jumlah *cluster* optimal berdasarkan nilai *silhouette* rata-rata untuk mengoptimalkan kinerja *clustering*. Pengujian dilakukan dengan menerapkan algoritma *K-Means*, *K-Medoids*, dan *Silhouette Coefficient* pada tiga dataset berbeda yang diperoleh dari *Kaggle* pada rentang 2 hingga 10 *cluster*. Evaluasi terhadap kinerja *clustering* dilakukan menggunakan *Silhouette Score*. Hasil penelitian menunjukkan bahwa jumlah *cluster* optimal untuk semua dataset adalah 2, baik pada algoritma *K-Means* maupun *K-Medoids*. Namun, algoritma *K-Means* menunjukkan kinerja yang lebih baik dibandingkan *K-Medoids* dengan nilai *Silhouette Score* yang lebih tinggi dan mendekati 1. Nilai *Silhouette Score* tertinggi *K-Means* di ketiga dataset berturut-turut yaitu 0,7099, 0,7674, dan 0,7716. Adapun *K-Medoids* yaitu 0,7064, 0,7445, dan 0,7257.

Kata Kunci : *Clustering, K-Means, K-Medoids, Silhouette Coefficient, Silhouette Score*

Palembang, 19 Juni 2024

Pembimbing

Mengetahui
Ketua Jurusan Teknik Informatika



Dr. M. Fachrudzi, S.Si., M.T.
NIP. 19800522008121002



Dian Palupi Rini, M.Kom., Ph.D.
NIP. 197802232006042002

KATA PENGANTAR

Puji dan syukur atas kehadiran Allah Swt. yang telah memberikan rahmat, hidayah dan petunjuk-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “Mengukur Kinerja *Clustering* Menggunakan Algoritma *K-Means* dan *K-Medoids* dengan Optimasi *Silhouette Coefficient*” sebagai salah satu syarat untuk menyelesaikan program Sarjana (S1) Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Dalam menyelesaikan tugas akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun tidak langsung. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Allah Swt. yang telah memberikan nikmat, kelancaran, kesanggupan dan segala kesempatan bagi penulis untuk menyelesaikan tugas akhir.
2. Kedua orang tua serta keluarga penulis tercinta, yang telah memberikan doa, dukungan, motivasi, dan semangat yang sangat berarti selama ini dalam melaksanakan perkuliahan di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya sehingga penulis dapat menyelesaikan tugas akhir ini.
3. Diri sendiri yang telah berjuang sejauh ini melawan rasa malas dan bekerja keras untuk menyelesaikan tugas akhir ini.
4. Bapak Dr. M. Fachrurrozi, S.Si., M.T. selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

5. Ibu Dian Palupi Rini, M.Kom., Ph.D. selaku Dosen Pembimbing yang senantiasa memberikan bimbingan, masukan, dan motivasi untuk menyelesaikan tugas akhir ini dengan baik.
6. Ibu Yunita, M.Cs. selaku Dosen Penguji tugas akhir yang telah memberikan saran dan masukan yang membangun.
7. Bapak Kanda Januar Miraswan, M.T. selaku Dosen Pembimbing Akademik yang senantiasa memberikan bimbingan selama masa perkuliahan.
8. Seluruh Dosen Jurusan Teknik Informatika Fakultas Ilmu Komputer yang telah banyak memberikan ilmu selama masa perkuliahan.
9. Seluruh staf administrasi dan pegawai Fakultas Ilmu Komputer yang telah membantu dalam urusan administrasi.
10. Teman-teman kelas dan jurusan Teknik Informatika Reguler Angkatan 2020 yang telah berbagi keluh kesah, motivasi, semangat, dan warna selama proses perkuliahan.

Penulis menyadari dalam penyusunan tugas akhir ini masih terdapat kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga tugas akhir ini dapat berguna dan bermanfaat bagi kita semua.

Palembang, 19 Juni 2024

Penulis

Karinda Amelia

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN KOMPREHENSIF.....	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT.....	vi
ABSTRAK.....	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xvii
BAB I PENDAHULUAN.....	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penelitian	I-5
1.5 Manfaat Penelitian	I-5
1.6 Batasan Masalah.....	I-6
1.7 Sistematika Penulisan	I-6
1.8 Kesimpulan	I-8
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori.....	II-1
2.2.1 <i>Exploratory Data Analysis (EDA)</i>	II-1
2.2.2 <i>Preprocessing</i>	II-5
2.2.3 <i>Clustering</i>	II-6

2.2.4	<i>K-Means</i>	II-6
2.2.5	<i>K-Medoids</i>	II-4
2.2.6	<i>Euclidean Distance</i>	II-5
2.2.7	<i>Silhouette Coefficient</i>	II-6
2.2.8	<i>Rational Unified Process (RUP)</i>	II-9
2.3	Penelitian Lain yang Relevan.....	II-11
2.4	Kesimpulan	II-12
BAB III METODOLOGI PENELITIAN.....		III-1
3.1	Pendahuluan	III-1
3.2	Pengumpulan Data	III-1
3.3	Tahapan Penelitian	III-1
3.3.1	Kerangka Kerja.....	III-2
3.3.2	Kriteria Pengujian.....	III-6
3.3.3	Format Data Pengujian	III-6
3.3.4	Alat yang Digunakan dalam Pelaksanaan Penelitian	III-7
3.3.5	Pengujian Penelitian	III-8
3.3.6	Analisis Hasil Pengujian dan Membuat Kesimpulan	III-8
3.4	Metode Pengembangan Perangkat Lunak	III-8
3.4.1	Fase Insepsi.....	III-8
3.4.2	Fase Elaborasi.....	III-9
3.4.3	Fase Konstruksi	III-9
3.4.4	Fase Transisi.....	III-9
3.5	Manajemen Proyek Penelitian	III-10
BAB IV PENGEMBANGAN PERANGKAT LUNAK		IV-1
4.1	Pendahuluan	IV-1

4.2	<i>Rational Unified Process (RUP)</i>	IV-1
4.2.1	Fase Insepsi.....	IV-1
4.2.1.1	Pemodelan Bisnis	IV-1
4.2.1.2	Kebutuhan Sistem (<i>System Requirements</i>).....	IV-2
4.2.1.3	Analisis Kebutuhan dan Desain	IV-2
4.2.1.4	Rancangan Perangkat Lunak.....	IV-20
4.2.2	Fase Elaborasi.....	IV-26
4.2.2.1	Kebutuhan Sistem	IV-26
4.2.2.2	Perancangan Data.....	IV-27
4.2.2.3	Perancangan Antarmuka.....	IV-27
4.2.2.4	Rancangan Diagram	IV-31
4.2.2.5	<i>Activity Diagram</i>	IV-32
4.2.2.6	<i>Sequence Diagram</i>	IV-35
4.2.2.7	<i>Class Diagram</i>	IV-37
4.2.3	Fase Konstruksi	IV-38
4.2.3.1	Implementasi <i>Library</i>	IV-38
4.2.3.2	Implementasi Kelas.....	IV-39
4.2.3.3	Implementasi Antarmuka	IV-40
4.2.4	Fase Transisi	IV-45
4.2.4.1	Kebutuhan Sistem	IV-45
4.2.4.2	Rencana Pengujian	IV-45

4.2.4.3 Implementasi Pengujian.....	IV-48
4.3 Kesimpulan	IV-53
BAB V HASIL DAN ANALISIS PENELITIAN.....	V-1
5.1 Pendahuluan	V-1
5.2 Data Hasil Penelitian.....	V-1
5.2.1 Konfigurasi Percobaan	V-12
5.2.2 Data Hasil Konfigurasi I.....	V-12
5.2.3 Data Hasil Konfigurasi II	V-14
5.2.4 Data Hasil Konfigurasi III	V-15
5.3 Analisis Hasil Penelitian	V-16
5.4 Data Hasil Pengelompokkan	V-21
5.5 Kesimpulan	V-25
BAB VI KESIMPULAN DAN SARAN	VI-1
6.1 Pendahuluan	VI-1
6.2 Kesimpulan	VI-1
6.3 Saran.....	VI-2
DAFTAR PUSTAKA	xx
LAMPIRAN.....	xxvi

DAFTAR TABEL

Tabel II-1.	Kriteria Pengukuran Nilai <i>Silhouette Coefficient</i>	II-8
Tabel III-1.	Format Pengujian Nilai <i>k</i> dan <i>Silhouette Score</i> Masing-Masing Dataset	III-7
Tabel III-2.	Format Pengujian Perbandingan Nilai <i>k</i> dan <i>Silhouette Score</i> Optimal	III-7
Tabel III-3.	Penjadwalan Penelitian dalam bentuk <i>Work Breakdown Structure</i> (WBS)	III-11
Tabel IV-1.	Kebutuhan Fungsional Perangkat Lunak.....	IV-2
Tabel IV-2.	Kebutuhan Non-Fungsional Perangkat Lunak	IV-2
Tabel IV-3.	Deskripsi Dataset I.....	IV-5
Tabel IV-4.	Sampel Data Beserta Notasi	IV-5
Tabel IV-5.	Hasil Normalisasi Data.....	IV-6
Tabel IV-6.	Penentuan <i>Centroid</i> untuk <i>K-Means</i>	IV-7
Tabel IV-7.	Jarak Data Terhadap <i>Centroid</i>	IV-8
Tabel IV-8.	<i>Cluster</i> Iterasi-1 <i>K-Means</i>	IV-8
Tabel IV-9.	<i>Centroid</i> Iterasi-2 <i>K-Means</i>	IV-9
Tabel IV-10.	Jarak Data Terhadap <i>Centroid</i> Iterasi-2 <i>K-Means</i>	IV-9
Tabel IV-11.	<i>Cluster</i> Iterasi-2 <i>K-Means</i>	IV-10
Tabel IV-12.	<i>Cluster</i> Iterasi-2 <i>K-Means</i>	IV-10
Tabel IV-13.	Nilai $a(i)$ <i>K-Means</i>	IV-11
Tabel IV-14.	Nilai $d(i)$ <i>K-Means</i>	IV-12
Tabel IV-15.	Nilai $b(i)$ <i>K-Means</i>	IV-13
Tabel IV-16.	Nilai $s(i)$ <i>K-Means</i>	IV-13
Tabel IV-17.	Penentuan <i>Medoid</i> untuk <i>K-Medoids</i>	IV-14

Tabel IV-18. Jarak Data Terhadap <i>Medoid</i>	IV-15
Tabel IV-19. <i>Cluster</i> Iterasi-2 <i>K-Medoids</i>	IV-15
Tabel IV-20. Penentuan <i>Medoid</i> untuk <i>K-Medoids</i> Iterasi-2	IV-16
Tabel IV-21. Jarak Data Terhadap <i>Medoid</i>	IV-16
Tabel IV-22. <i>Cluster</i> Iterasi-2 <i>K-Medoids</i>	IV-16
Tabel IV-23. Nilai $a(i)$	IV-18
Tabel IV- 24. Nilai $d(i)$	IV-19
Tabel IV-25. Nilai $b(i)$	IV-19
Tabel IV-26. Nilai $s(i)$	IV-19
Tabel IV- 27. Definisi Aktor	IV-21
Tabel IV-28. Definisi <i>Use Case</i>	IV-21
Tabel IV-29. Skenario <i>Use Case</i> Mengunggah Dataset.....	IV-22
Tabel IV-30. Skenario Melakukan <i>Exploratory Data Analysis</i> (EDA) dan <i>Preprocessing</i>	IV-23
Tabel IV-31. Skenario <i>Use Case</i> Melakukan <i>Clustering</i> Algoritma <i>K-Means</i> dengan Optimasi <i>Silhouette Coefficient</i>	IV-24
Tabel IV-32. Skenario <i>Use Case</i> Melakukan <i>Clustering</i> Algoritma <i>K-</i> <i>Medoids</i> dengan Optimasi <i>Silhouette Coefficient</i>	IV-25
Tabel IV-33. Kebutuhan Sistem.....	IV-27
Tabel IV-34. <i>Library</i> Pendukung Perangkat Lunak.....	IV-39
Tabel IV-35. Implementasi Kelas	IV-39
Tabel IV-36. Kebutuhan Sistem.....	IV-45
Tabel IV-37. Rencana Pengujian <i>Use Case</i> Mengunggah Dataset	IV-46

Tabel IV-38. Rencana Pengujian <i>Use Case</i> Melakukan <i>Exploratory Data Analysis</i> (EDA) dan <i>Preprocessing</i>	IV-46
Tabel IV-39. Rencana Pengujian Melakukan <i>Clustering</i> Algoritma <i>K-Means</i> dengan Optimasi <i>Silhouette Coefficient</i>	IV-46
Tabel IV-40. Melakukan <i>Clustering</i> Algoritma <i>K-Medoids</i> dengan Optimasi	IV-47
Tabel IV-41. Implementasi Pengujian <i>Use Case</i> Mengunggah Dataset	IV-48
Tabel IV-42. Rencana Pengujian <i>Use Case</i> Melakukan <i>Exploratory Data Analysis</i> (EDA) dan <i>Preprocessing</i>	IV-48
Tabel IV-43. Rencana Pengujian <i>Use Case</i> Melakukan <i>Clustering</i> Algoritma <i>K-Means</i> dengan Optimasi <i>Silhouette Coefficient</i>	IV-50
Tabel IV-44. Rencana Pengujian <i>Use Case</i> Melakukan <i>Clustering</i> Algoritma <i>K-Medoids</i> dengan Optimasi <i>Silhouette Coefficient</i>	IV-51
Tabel V-1. Statistik Deskriptif Dataset I.....	V-2
Tabel V-2. Statistik Deskriptif Dataset II	V-2
Tabel V-3. Statistik Deskriptif Dataset III.....	V-3
Tabel V-4. Hasil Pengujian Nilai <i>k</i> dan <i>Silhouette Score</i> Dataset I	V-20
Tabel V-5. Hasil Pengujian Nilai <i>k</i> dan <i>Silhouette Score</i> Dataset II.....	V-20
Tabel V-6. Hasil Pengujian Nilai <i>k</i> dan <i>Silhouette Score</i> Dataset III.....	V-20
Tabel V-7. Perbandingan Nilai <i>k</i> dan <i>Silhouette Score</i> Optimal	V-21

DAFTAR GAMBAR

Gambar II-1.	<i>Arsitektur Rational Unified Process</i>	II-9
Gambar III-1.	Diagram Tahapan Penelitian	III-2
Gambar III-2.	Kerangka Kerja Penelitian	II-3
Gambar IV-1.	Diagram <i>Use Case</i>	IV-20
Gambar IV-2.	Rancangan Antarmuka Halaman <i>Homepage</i> Sebelum Mengunggah Dataset.....	IV-28
Gambar IV-3.	Rancangan Antarmuka Halaman <i>Homepage</i> Setelah Mengunggah Dataset.....	IV-28
Gambar IV-4.	Rancangan Antarmuka Halaman <i>Exploratory and Preprocessing</i>	IV-29
Gambar IV-5.	Rancangan Antarmuka Halaman <i>K-Means Clustering</i>	IV-30
Gambar IV-6.	Rancangan Antarmuka Halaman <i>K-Medoids Clustering</i>	IV-31
Gambar IV-7.	<i>Activity Diagram</i> Mengunggah Dataset	IV-32
Gambar IV-8.	<i>Activity Diagram</i> Melakukan <i>Exploratory Data Analysis</i> (EDA) dan <i>Preprocessing</i>	IV-33
Gambar IV-9.	<i>Activity Diagram</i> Melakukan <i>Clustering</i> Algoritma <i>K- Means</i> dengan Optimasi <i>Silhouette Coefficient</i>	IV-34
Gambar IV-10.	<i>Activity Diagram</i> Melakukan <i>Clustering</i> Algoritma <i>K- Medoids</i> dengan Optimasi <i>Silhouette Coefficient</i>	IV-35
Gambar IV-11.	<i>Sequence Diagram</i> Mengunggah Dataset	IV-36
Gambar IV-12.	<i>Sequence Diagram</i> Melakukan <i>Exploratory Data Analysis</i> (EDA) dan <i>Preprocessing</i>	IV-36

Gambar IV-13. <i>Sequence Diagram</i> Melakukan <i>Clustering</i> Algoritma <i>K-Means</i> dengan Optimasi <i>Silhouette Coefficient</i>	IV-37
Gambar IV-14. <i>Sequence Diagram</i> Melakukan <i>Clustering</i> Algoritma <i>K-Medoids</i> dengan Optimasi <i>Silhouette Coefficient</i>	IV-37
Gambar IV-15. <i>Class Diagram</i>	IV-38
Gambar IV-16. Implementasi Antarmuka Halaman <i>Homepage</i> Sebelum Mengunggah Dataset.....	IV-41
Gambar IV-17. Implementasi Antarmuka Halaman <i>Homepage</i> Setelah Mengunggah Dataset.....	IV-41
Gambar IV-18. Rancangan Antarmuka Halaman <i>Exploratory and Preprocessing</i>	IV-42
Gambar IV-19. Implementasi Antarmuka Halaman <i>K-Means Clustering</i> ...	IV-43
Gambar IV-20. Implementasi Antarmuka Halaman <i>K-Medoids Clustering</i>	IV-44
Gambar V-1. Histogram Dataset I.....	V-6
Gambar V-2. QQ Plot Dataset I	V-8
Gambar V-3. <i>Box Plot</i> Dataset I.....	V-10
Gambar V-4. Grafik <i>Silhouette Coefficient K-Means Clustering</i> Dataset I	V-13
Gambar V-5. Grafik <i>Silhouette Coefficient K-Medoids Clustering</i> Dataset I	V-13
Gambar V-6. Grafik <i>Silhouette Coefficient K-Means Clustering</i> Dataset II.....	V-14

Gambar V-7.	Grafik <i>Silhouette Coefficient K-Medoids Clustering</i> Dataset II	V-14
Gambar V-8.	Grafik <i>Silhouette Coefficient K-Means Clustering</i> Dataset III.....	V-15
Gambar V-9.	Grafik <i>Silhouette Coefficient K-Medoids Clustering</i> Dataset III.....	V-16
Gambar V-10.	<i>Silhouette Score K-Means</i> Dataset I.....	V-17
Gambar V-11.	<i>Silhouette Score K-Medoids</i> Dataset I.....	V-17
Gambar V-12.	<i>Silhouette Score K-Means</i> Dataset II	V-18
Gambar V-13.	<i>Silhouette Score K-Medoids</i> Dataset II	V-18
Gambar V-14.	<i>Silhouette Score K-Means</i> Dataset III	V-19
Gambar V-15.	<i>Silhouette Score K-Medoids</i> Dataset III	V-19
Gambar V-16.	3D Plot <i>K-Means</i> Dataset I	V-22
Gambar V-17.	3D Plot <i>K-Medoids</i> Dataset I	V-22
Gambar V-18.	3D Plot <i>K-Means</i> Dataset II	V-23
Gambar V-19.	3D Plot <i>K-Medoids</i> Dataset II.....	V-23
Gambar V-20.	3D Plot <i>K-Means</i> Dataset III.....	V-24
Gambar V-21.	3D Plot <i>K-Medoids</i> Dataset III.....	V-24

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab pendahuluan ini akan diuraikan secara umum tentang keseluruhan penelitian yang mencakup latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, serta sistematika penulisan dokumen penelitian.

1.2 Latar Belakang

Seiring dengan perkembangan teknologi yang semakin pesat, volume data yang dihasilkan semakin besar. Data tersebut dapat berasal dari berbagai sumber, seperti media sosial, observasi, *log file*, dan lain sebagainya. Pertumbuhan data yang besar dan kompleks membutuhkan pengolahan atau pemrosesan data yang efektif dan efisien. Salah satu cara untuk meningkatkan efektivitas dan efisiensi dalam pemrosesan data ialah dengan melakukan *clustering* (Puspita & Ariyanti, 2021).

Clustering menjadi salah satu teknik yang sangat populer dalam pengolahan data karena dapat membantu menemukan pola dan karakteristik dalam data yang kompleks. Pada *clustering*, data dikategorikan berdasarkan kesamaan karakteristik dari seluruh kelompok data (Nugroho & Adhinata, 2022). Ada banyak algoritma *clustering* yang tersedia, dan setiap algoritma memiliki karakteristik dan kinerja yang berbeda-beda. Oleh karena itu, perlu dilakukan pengukuran kinerja antara algoritma *clustering* yang berbeda dalam menentukan algoritma yang paling sesuai untuk digunakan pada jenis data tertentu. Hal ini dapat membantu meningkatkan

efektivitas dan efisiensi dalam pengolahan data *clustering* dan menghasilkan *cluster* yang lebih berkualitas dan sesuai dengan kebutuhan.

Di antara banyaknya algoritma *clustering*, terdapat dua algoritma yang masih saling berkaitan, yaitu *K-Means* dan *K-Medoids*. *K-Means* merupakan algoritma *clustering* yang paling umum digunakan. Hal ini dikarenakan *K-Means* merupakan algoritma yang sederhana, mudah diimplementasikan, relatif cepat, dan mudah dimodifikasi (Pratiwi & Wibowo, 2022). Algoritma ini dapat bekerja pada jumlah data yang besar dengan waktu komputasi yang relatif cepat dan efisien (Darnita et al., 2020). Adapun *K-Medoids* atau *Partitioning Around Method* (PAM) merupakan salah satu pengembangan dari algoritma *K-Means* yang ditujukan untuk mengurangi sensitivitas dari partisi yang dihasilkan sehubungan dengan nilai-nilai ekstrem yang terdapat dalam dataset. Algoritma *K-Medoids* muncul sebagai penanggulangan kelemahan algoritma *K-Means* yang sensitif terhadap *outlier* dikarenakan sebuah objek dengan sebuah nilai yang besar mungkin secara substansial menyimpang dari distribusi data (Triyanto, 2015).

Penelitian yang dilakukan bertujuan membandingkan algoritma *K-Means* dan *K-Medoids* dengan menguji data menggunakan aplikasi berbasis *Java*, *Hadoop*, dan *Hive*. Parameter yang digunakan dalam pengujian ini adalah akurasi, waktu eksekusi, dan kompleksitas algoritma. Dari pengujian tersebut diperoleh kesimpulan bahwa algoritma *K-Medoids* lebih baik jika dibandingkan dengan algoritma *K-Means* (Nurhayati et al., 2018). Dalam penelitian lain mengenai pengelompokan data menggunakan *K-Means* dan *K-Medoids* menyatakan bahwa dari segi waktu *K-Means* hanya membutuhkan waktu 1 detik dalam mengolah data

sedangkan *K-Medoids* membutuhkan waktu 1 menit 38 detik. Selain itu, nilai *Davies-Bouldin Index* (DBI) pada *K-Means* lebih mendekati 0 dibandingkan *K-Medoids* (Kamila et al., 2019). Penelitian lain membandingkan algoritma *K-Means* dan *K-Medoids* untuk pengelompokan pelanggan berdasarkan karakteristik tertentu. Hasilnya *K-Means* lebih akurat daripada *K-Medoids*, dengan tingkat akurasi *K-Means* adalah 90,47% dan *K-Medoids* 76,19% (Muningsih, 2018). Adapun penelitian lainnya membandingkan algoritma *K-Means* dan *K-Medoids* yang bertujuan untuk menganalisa sebaran gempa dan pola spasial kejadian gempa di Indonesia. Hasilnya didapatkan nilai $k = 6$ adalah nilai k terbaik dan nilai *silhouette* tertinggi sebesar 0.4674067. Dari hasil perbandingan kedua algoritma, diketahui berdasarkan hasil perhitungan nilai *silhouette* diperoleh hasil *K-Medoids* lebih baik bila dibandingkan dengan *K-Means* (Senduk et al., 2019).

Berdasarkan penelitian terdahulu, terdapat perbedaan dalam hasil *cluster*. Oleh karena itu, diperlukan penelitian lebih lanjut untuk mendapatkan pemahaman yang lebih komprehensif tentang karakteristik dan kinerja dari algoritma *K-Means* dan *K-Medoids*. Berbeda dengan penelitian terdahulu yang hanya menggunakan satu jenis dataset, penelitian ini akan menggunakan tiga jenis dataset dengan jumlah atribut yang berbeda untuk menganalisis hasil dan kinerja algoritma pada dataset dengan jumlah atribut yang bervariasi. Hal ini diharapkan dapat memberikan wawasan yang lebih luas mengenai kinerja kedua algoritma tersebut dalam berbagai skenario data. Dalam melakukan pengklasteran data, baik *K-Means* maupun *K-Medoids* telah terbukti memiliki kinerja yang sangat baik dengan tingkat akurasi yang tinggi dan efisiensi dalam memproses jumlah objek yang besar. Namun,

tantangan terbesar dalam penggunaan algoritma tersebut adalah menentukan jumlah *cluster* (k) yang optimal untuk data yang dianalisis (Ramadhania et al., 2023). *Silhouette Coefficient* adalah salah satu algoritma yang dapat digunakan untuk mengatasi tantangan tersebut. Algoritma ini dirancang untuk menentukan jumlah *cluster* yang optimal melalui pendekatan nilai *silhouette* rata-rata, yang memberikan pengukuran objektif terhadap kualitas *cluster* yang dibentuk (Dewa & Jatipaningrum, 2019). Kemampuannya dalam menganalisa jumlah k *cluster* dapat digunakan untuk mengoptimalkan kinerja *K-Means* dan *K-Medoids*.

Berdasarkan uraian diatas, pada penelitian ini akan dilakukan klusterisasi tiga jenis dataset dengan jumlah atribut yang berbeda menggunakan algoritma *K-Means* dan *K-Medoids* yang dioptimasi menggunakan *Silhouette Coefficient* untuk mendapatkan hasil *cluster* yang optimal.

1.3 Rumusan Masalah

Berdasarkan uraian dari latar belakang masalah diatas, maka diperoleh rumusan masalah sebagai berikut :

1. Bagaimana implementasi algoritma *K-Means* dan *K-Medoids* dengan optimasi *Silhouette Coefficient* dalam pengklasteran tiga jenis dataset dengan jumlah atribut yang berbeda?
2. Bagaimana perbandingan kinerja algoritma *K-Means* dan *K-Medoids* dengan optimasi *Silhouette Coefficient* dalam pengklasteran tiga jenis dataset dengan jumlah atribut yang berbeda?

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini antara lain sebagai berikut :

1. Mengimplementasikan algoritma *K-Means* dan *K-Medoids* menggunakan optimasi *Silhouette Coefficient* dalam pengklasteran tiga jenis dataset dengan jumlah atribut yang berbeda.
2. Mengetahui perbandingan kinerja algoritma *K-Means* dan *K-Medoids* dengan optimasi *Silhouette Coefficient* dalam pengklasteran tiga jenis dataset dengan jumlah atribut yang berbeda.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini antara lain sebagai berikut:

1. Bagi Peneliti
 - Sebagai salah satu syarat untuk mendapatkan gelar Sarjana Komputer di Fakultas Ilmu Komputer Universitas Sriwijaya.
 - Sebagai tolak ukur peneliti selama mengemban pendidikan Teknik Informatika di Fakultas Ilmu Komputer Universitas Sriwijaya.
 - Meningkatkan pemahaman dan pengetahuan dalam bidang *data mining* terutama *clustering*.
2. Bagi Universitas
 - Penelitian ini dapat dijadikan sebagai referensi untuk penelitian-penelitian selanjutnya.

3. Bagi Pembaca

- Memahami kinerja *clustering* menggunakan algoritma *K-Means* dan *K-Medoids* dalam melakukan pengklasteran data dengan optimasi *Silhouette Coefficient*.
- Membantu pembaca dalam memilih algoritma *clustering* yang tepat untuk pengklasteran data tertentu dan meningkatkan kinerja sistem *clustering* yang digunakan.
- Dapat menjadi referensi bagi peneliti yang melakukan penelitian sejenis atau berhubungan dengan pengklasteran data.

1.6 Batasan Masalah

Agar pembahasan dalam penelitian ini dapat lebih jelas dan terarah, maka diperlukan batasan terhadap permasalahan yang akan diteliti yang berfokus pada :

1. Perangkat lunak yang dibangun mampu memproses semua jenis dataset dengan tipe data numerik dan kategorikal.
2. Penelitian ini hanya melakukan pengujian terhadap tiga dataset kanker dengan jumlah atribut yang berbeda yang diperoleh dari situs *Kaggle*.
3. Pengukuran jarak menggunakan *Euclidean Distance*.

1.7 Sistematika Penulisan

Standar penulisan tugas akhir ini merujuk pada sistematika penulisan tugas akhir Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya, yaitu :

BAB I PENDAHULUAN

Bab ini membahas mengenai latar belakang, perumusan masalah, tujuan, manfaat penelitian, batasan masalah, sistematika penulisan, dan kesimpulan.

BAB II KAJIAN LITERATUR

Bab ini akan membahas dasar-dasar teori yang digunakan dalam penelitian, seperti *Exploratory Data Analysis*, *Preprocessing*, *clustering* menggunakan algoritma *Euclidean Distance*, *K-Means*, *K-Medoids*, dan *Silhouette Coefficient* serta penelitian lain yang relevan dengan penelitian ini.

BAB III METODOLOGI PENELITIAN

Bab ini membahas mengenai tahapan yang akan dilakukan pada penelitian. Rencana tahapan penelitian akan dijelaskan secara rinci dengan mengacu pada kerangka kerja serta perancangan proyek perangkat lunak.

BAB IV PENGEMBANGAN PERANGKAT LUNAK

Bab ini berisi perancangan dan implementasi perangkat lunak dengan metode pemrograman berorientasi objek berdasarkan panduan *Rational Unified Process* yang di dalamnya terdapat fase inepsi, elaborasi, konstruksi, dan transisi.

BAB V HASIL DAN ANALISIS PENELITIAN

Bab ini akan membahas hasil pengujian serta analisis dari penelitian yang telah dilakukan. Analisis dari penelitian akan digunakan sebagai landasan pembuatan kesimpulan dalam penelitian ini.

BAB VI KESIMPULAN DAN SARAN

Bab ini akan membahas mengenai kesimpulan dari penelitian dan pengujian yang dilakukan beserta saran yang dapat digunakan untuk penelitian yang lebih baik kedepannya.

1.8 Kesimpulan

Pada bab ini telah dijelaskan mengenai latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, dan sistematika penulisan. Berdasarkan penjelasan di atas, penelitian menggunakan algoritma *K-Means* dan *K-Medoids* dengan optimasi *Silhouette Coefficient* untuk mengukur kinerja *clustering* dalam melakukan pengklasteran data.

DAFTAR PUSTAKA

- Anggara, M., Sujiani, H., & Nasution, H. (2016). Pemilihan Distance Measure pada K-Means Clustering untuk Pengelompokkan Member di Alvaro Fitness. *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, 1(1), 1–6.
- Azzahra Nasution, D., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *Journal of Computer Engineering System and Science*, 4(1), 78–82.
- Darnita, Y., Toyib, R., & Kurniawan, Y. (2020). Penerapan Metode K-Means Clustering pada Aplikasi Android pada Tanaman Obat Herbal. *Jurnal Pseudocode*, 7(2), 105–114.
- Dewa, F. A., & Jatipaningrum, M. T. (2019). Segmentasi E-Commerce dengan Cluster K-Means dan Fuzzy C-Means (Studi Kasus: Media Sosial di Indonesia yang diunduh di Play Store). *Jurnal Statistika Industri dan Komputasi*, 4(1), 53–67.
- Fath, M. S. A., & Karawang, R. A. (2022). Tinjauan Perancangan Produksi dan Kualitas pada Produk Rak Dies di CV Sarana Sejahtera Teknik. *Jurnal Ilmiah Wahana Pendidikan*, 8(2), 159–168.
- Febriani, S. (2022). Analisis Deskriptif Standar Deviasi. *Jurnal Pendidikan Tambusai*, 6(1), 910–913.

- Handari, A. F., & Santosa, B. (2022). Pengaruh Layanan Informasi dengan Teknik Modeling Simbolik terhadap Self Efficacy Karir Remaja. *Jurnal Pendidikan Tambusai*, 6(1), 2690–2269.
- Kamila, I., Khairunnisa, U., & Mustakim. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, 5(1), 119–125.
- Karo Karo, I. M., Dewi, S., Mardiana, M., Ramadhani, F., & Harliana, P. (2023). K-Means and K-Medoids Algorithm Comparison for Clustering Forest Fire Location in Indonesia. *Jurnal Ecotipe (Electronic, Control, Telecommunication, Information, and Power Engineering)*, 10(1), 86–94.
- Muningsih, E. (2018). Komparasi Metode Clustering K-Means dan K-Medoids dengan Model Fuzzy RFM untuk Pengelompokan Pelanggan. *Jurnal Evolusi*, 6(2), 106–113.
- Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering Berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, 4(1), 20–24.
- Novianto, R., & Goeirmanto, L. (2019). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering untuk Menganalisa Bisnis Perusahaan Asuransi. *Jurnal Teknik Informatika dan Sistem Informasi*, 6(1), 85–95.

- Nugraha, A., Nurdiawan, O., & Dwilestari, G. (2022). Penerapan Data Mining Metode K-Means Clustering untuk Analisa Penjualan pada Toko Yana Sport. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 849–855.
- Nugroho, N., & Adhinata, F. D. (2022). Penggunaan Metode K-Means dan K-Means++ Sebagai Clustering Data Covid-19 di Pulau Jawa. *Teknika*, 11(3), 170–179.
- Nurhayati, Sinatrya, S. N., Wardhani, L. K., & Busman. (2018). Analysis of K-Means and K-Medoids's Performance Using Big Data Technology. *The 6th International Conference on Cyber and IT Service Management (CITSM 2018)*, 1–5.
- Padang, Y., Abdunnur, & R, M. S. (2023). Analisis Kuartil, Desil dan Persentil pada Ukuran Panjang Udang Bintik Coklat (*Metapenaeus Demani*) di Perairan Muara Ilu kabupaten Kutai Kartanegara. *Tropical Aquatic Sciences*, 2(1), 44.
- Paembonan, S., Abduh, H., & Kunci, K. (2021). Penerapan Metode Silhouette Coeficient untuk Evaluasi Clustering Obat Clustering; K-means; Silhouette coefficient. *Jurnal Ilmiah - Ilmu Teknik*, 6(2), 48–54.
- Patricia, S., Marpaung, C. P., Wijaya, L. R., Paramartha, M. A., Atmadja, W. D., & Ningsih, R. Y. (2023). Implementasi Exploratory Data Analysis (EDA) untuk Menganalisis Berbagai Faktor Risiko Penyakit Jantung di Amerika Serikat. *Journal of Student Development Information System (JoSDIS)*, 3(2), 108–124.

- Perwitasari, R., Afwani, R., & Endang Anjarwani, S. (2020). *Penerapan Metode Rational Unified Process (RUP) dalam Pengembangan Sistem Informasi Medical Check Up pada Citra Medical Centre.*
- Pratiwi, H., & Wibowo, A. P. W. (2022). Mengidentifikasi Kelompok Sektor Perkebunan di Indonesia Berdasarkan Produktivitas Hasil Bumi. *Diffusion Journal of System and Information Technology*, 2(2), 54–61.
- Puspita, D., & Ariyanti, S. (2021). *Perbandingan Fuzzy C-Means dan K-Means untuk Text Clustering Menggunakan LSI.* Universitas Katolik Parahyangan.
- Rahman, A. T., Wiranto, & Anggrainingsih, R. (2017). Coal Trade Data Clustering Using K-Means (Case Study PT. Global Bangkit Utama). *ITSMART: Jurnal Ilmiah Teknologi Dan Informasi*, 6(1), 24–31.
- Ramadhania, H. L., Widiarti, Zakaria, L., & Nusyirwan, dan. (2023). Aplikasi Metode Sillhouette Coefficient, Metode Elbow dan Metode Gap Statistics dalam Menentukan K Optimal pada Analisis K-Medoids. *Jurnal Siger Matematika*, 04(01).
- Sabarudin, S., & Shar, S. A. (2020). Penerapan Data Mining Metode K-Means Clustering untuk Analisa Penjualan pada Toko Yana Sport. *Jurnal Nuansa Informatika*, 14(1), 63.
- Senduk, R. F., Indwiarti, & Nhita, F. (2019). Clustering of Earthquake Prone Areas in Indonesia Using K-Medoids Algorithm. *Indonesia Journal of Computing*, 4(3), 65–76.

- Sholeh Wibowo, B., Aini Masrurah, N., Uswatun Kasanah, Y., Trapsilawati, F., & Adha Ilhami, M. (2019). Toward a Taxonomy of Micro and Small Manufacturing Enterprises. *Communications in Science and Technology*, 4(2), 74–80.
- Sholikhah, A. (2016). Statistik Deskriptif dalam Penelitian Kualitatif. *Komunika*, 10(2), 342–362.
- Triwahyuni, A., & Saputra, N. (2015). Architecture E-Mall Using Rup (Rational Unified Process) Methods. *Cogito Smart Journal*, 1(1), 1–12.
- Triyanto, W. A. (2015). Algoritma K-Medoids untuk Penentuan Strategi Pemasaran Produk. *Jurnal SIMETRIS*, 6(1), 183–188.
- Wanto, A., Siregar, M. N. H., Windarto, A. P., Hartama, D., Ginantra, N. L. W. S. R., Napitupul, D., Negara, E. S., Lubis, M. R., & Prianto Cahyo. (2020). *Data Mining Algoritma dan Implementasi* (1st ed.).
- Werdiningsih, D., Rahmawati, D., Cholissodin, I., & Santoso, N. (2019). Optimasi K-Means untuk Pengelompokan Data Kinerja Akademik Dosen Menggunakan Particle Swarm Optimization. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(4), 2548–2964.
- Wira, B., Budianto, A. E., & Wiguna, A. S. (2019). Implementasi Metode K-Medoids Clustering untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 di Universitas Kanjuruhan Malang. *RAINSTEK Jurnal Terapan Sains & Teknologi*, 1(3), 54–69.

Yaziz, Kusnandar, D., & Wira Rizki, S. (2019). Analisis Regresi Robust Estimasi-M dengan Menggunakan Pembobotan Bisquare Tukey dan Welsch dalam Mengatasi Data Outlier. *Buletin Ilmiah Mat, Stat, dan Terapannya (Bimaster)*, 08(4), 799–804.