



**KOMPARASI *ELBOW METHOD*, *DAVIES BOULDIN INDEX*, DAN
SILHOUETTE COEFFICIENT PADA ALGORITMA K-MEDOIDS**

SKRIPSI

**untuk memenuhi persyaratan dalam menyelesaikan
program sarjana Strata-1 Matematika**

Oleh :

LITA AGUSTINA

NIM 1711011220010

**PROGRAM STUDI MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

2024

SKRIPSI

Komparasi *Elbow Method*, *Davies Bouldin Index*, dan *Silhouette Coefficient* Pada Algoritma *K-Medoids*

Oleh:
Lita Agustina
NIM 1711011220010

telah dipertahankan di depan Dosen Penguji pada tanggal 3 Juli 2024.
Susunan Dosen Penguji:

Pembimbing I

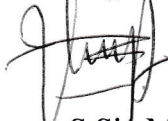


Akhmad Yusuf, S.Si., M.Kom.
NIP. 198004022005011001

Dosen Penguji:

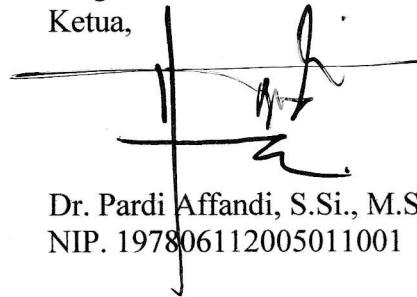
3. Oni Soesanto, S.Si., M.Si. (✓)
4. Saman Abdurrahman, S.Si., M.Sc. (✓)

Pembimbing II



Thresye, S.Si., M.Si.
NIP. 197205042000122002

Banjarbaru, 17 Juli 2024
Program Studi Matematika FMIPA ULM
Ketua,



Dr. Pardi Affandi, S.Si., M.Sc.
NIP. 197806112005011001



Mengetahui,
Wakil Dekan Bidang Akademik,

Dr. Gunawan, S.Si., M.Si.
NIP. 19791101 200501 1 002

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, Juli 2024

Lita Agustina
NIM. 1711011220010

ABSTRAK

KOMPARASI *ELBOW METHOD*, *DAVIES BOULDIN INDEX*, DAN *SILHOUETTE COEFFICIENT* PADA ALGORITMA *K-MEDOIDS* (Oleh : Lita Agustina; Pembimbing: Akhmad Yusuf, Thresye ; 2024 ; 39 Halaman)

Di era informasi saat ini, data mining menjadi aset penting dalam berbagai bidang. Dalam data mining terdapat tugas *clustering* yang bertujuan untuk mengelompokkan data berdasarkan ukuran kedekatan atau kemiripan. Metode *clustering* yang dapat digunakan yaitu *k-means*, *k-medoids*, dan lainnya. *K-medoids* hadir untuk mengatasi kelemahan *k-means* karena *k-means* dapat mengalami masalah ketika mengelompokkan data yang mengandung *outlier*. Penelitian ini menggunakan metode *clusterisasi K-Medoids* untuk mengelompokkan data set *Iris* dan *Wine*. Pengelompokkan jumlah *cluster* yang tepat menjadi permasalahan, sehingga diperlukan metode evaluasi *cluster*. Tujuan dari penelitian ini adalah untuk menentukan jumlah *cluster* optimal. Metode evaluasi *cluster* yang digunakan yaitu metode *elbow*, *Davies Bouldin Index* (DBI), dan *silhouette coefficient*. Hasil penelitian menunjukkan dengan menggunakan algoritma *K-Medoids*, metode *elbow* memberikan hasil optimal sebanyak 3 *cluster* untuk kedua data set. Metode *Davies Bouldin Index* (DBI) dan *silhouette coefficient* menunjukkan masing-masing 2 *cluster* sebagai hasil optimal untuk kedua data set. Metode *elbow* konsisten dalam mengidentifikasi jumlah *cluster* optimal sesuai dengan pengelompokan awal untuk kedua data set sehingga metode *elbow* ditetapkan sebagai metode evaluasi *cluster* yang optimal.

Kata Kunci: Algoritma *K-Medoids*, Metode *Elbow*, *Davies Bouldin Index*, *Silhouette Coefficient*.

ABSTRACT

COMPARATION OF ELBOW METHOD, DAVIES BOULDIN INDEX, AND SILHOUETTE COEFFICIENT ON THE K-MEDOIDS (By: Lita Agustina; Supervisors: Akhmad Yusuf, Thresye; 2024; 39 Pages)

In the current information era, data mining has become an important asset in various fields. In data mining there is a clustering task which aims to group data based on measures of closeness or similarity. Clustering methods that can be used are k-means, k-medoids, and others. K-medoids is here to overcome the weaknesses of k-means because k-means can experience problems when grouping data that contains outliers. This research uses the K-Medoids clusterization method to group the Iris and Wine data sets. Grouping the right number of clusters is a problem, so a cluster evaluation method is needed. The aim of this research is to determine the optimal number of clusters. The cluster evaluation methods used are the elbow method, Davies Bouldin Index (DBI), and silhouette coefficient. The research results show that using the K-Medoids algorithm, the elbow method provides optimal results with 3 clusters for both data sets. The Davies Bouldin Index (DBI) and silhouette coefficient methods show each of the 2 clusters as optimal results for both data sets. The elbow method is consistent in identifying the optimal number of clusters according to the initial grouping for both data sets so that the elbow method is determined as the optimal cluster evaluation method.

Keywords: K-Medoids Algorithm, Elbow Method, Davies Bouldin Index, Silhouette Coefficient

PRAKATA

Alhamdulillah, puji syukur kepada Allah SWT atas berkat dan rahmat-Nya sehingga penulis diberikan kesehatan dan kesempatan dalam menyelesaikan skripsi yang berjudul "Komparasi *Elbow Method*, *Davies Bouldin Index*, dan *Silhouette Coefficient* pada Algoritma *K-Medoids*" untuk memenuhi persyaratan dalam menyelesaikan program sarjana Strata-1 Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Dalam penyusunan skripsi ini tidak lepas dari dukungan dan bantuan dari berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih kepada :

1. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lambung Mangkurat
2. Koordinator Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lambung Mangkurat
3. Bapak Akhmad Yusuf, S.Si., M.Kom. dan Ibu Thresye, S.Si., M.Si. selaku dosen pembimbing yang telah banyak membantu, memberi arahan, dan meluangkan waktu dalam proses penyusunan dan penelitian skripsi ini.
4. Seluruh Dosen dan Staf Program Studi Matematika yang telah membantu dalam memberikan informasi yang bermanfaat dalam penyusunan skripsi.
5. Ayah, ibu, kakak-kakak, dan keluarga penulis yang selalu memberikan doa, kasih sayang, dukungan, dan kepercayaan selama perkuliahan hingga penulis menyelesaikan skripsi.
6. Sahabat penulis yang telah memberikan doa dan dukungan kepada penulis selama pengerjaan skripsi.
7. Seluruh rekan dan teman-teman mahasiswa Matematika angkatan 2017, serta seluruh pihak yang telah memberikan bantuan, baik berupa motivasi, saran, dan dukungan kepada penulis selama penyusunan skripsi.

Penulis menyadari bahwa dalam penyusunan skripsi ini terdapat kekurangan. Oleh karena itu, penulis berharap kritik dan saran untuk penelitian kedepannya.

Penulis berharap semoga skripsi ini dapat memberikan manfaat kepada pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, Juli 2024

Lita Agustina

NIM. 1711011220010

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
PERNYATAAN	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
PRAKATA	vi
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Tujuan Penelitian.....	3
1.3 Sistematika Penulisan.....	3
BAB II TINJAUAN PUSTAKA	4
2.1 Data Mining.....	4
2.2 <i>Outlier</i>	6
2.3 <i>Clustering</i>	8
2.4 <i>K-Medoids</i>	8
2.5 <i>Clustering</i> Data dengan Algoritma <i>K-Medoids</i>	9
2.6 Metode <i>Elbow</i> dan <i>Sum of Square Error (SSE)</i>	11
2.7 <i>Davies Bouldin Index</i>	11
2.8 <i>Silhouette Coefficient</i>	12

BAB III METODE PENELITIAN	14
3. 1. Deskripsi Data	14
3. 2. Prosedur Penelitian	14
BAB IV PEMBAHASAN	17
4.1 Hasil	17
4.2 Pengumpulan Data.....	17
4.3 <i>Data Preparation</i>	17
4.4 <i>Clusterisasi</i> Menggunakan Metode <i>K-Medoids</i>	19
4.5 Evaluasi Jumlah <i>Cluster</i> Optimal	24
4.5.1. Metode <i>Elbow</i> dengan <i>Sum of Square Error</i>	24
4.5.2. <i>Davies Bouldin Index</i>	26
4.5.3. <i>Silhouette Coefficient</i>	29
4.6 Pembahasan	31
BAB V PENUTUP	36
5.1 Kesimpulan	36
5.2 Saran	36
DAFTAR PUSTAKA	37
LAMPIRAN	

DAFTAR GAMBAR

Gambar	Halaman
1. Proses <i>Knowledge Discovery in Database</i>	5
2. Skema identifikasi <i>outlier</i> menggunakan <i>IQR</i> atau <i>boxplot</i>	7
3. Flowchart algoritma <i>K-Medoids</i>	10
4. Metode penelitian.....	16
5. <i>Boxplot</i> data set <i>Iris</i>	18
6. <i>Boxplot</i> data set <i>Iris</i> yang telah di transformasi.....	19
7. Plot perhitungan SSE.....	26
8. Plot perhitungan DBI.....	29
9. Plot perhitungan SC.....	31
10. Plot perhitungan SSE data set <i>Iris</i>	31
11. Plot perhitungan DBI data set <i>Iris</i>	32
12. Plot perhitungan SC data set <i>Iris</i>	32
13. Plot perhitungan SSE data set <i>Wine</i>	33
14. Plot perhitungan DBI data set <i>Wine</i>	33
15. Plot perhitungan SC data set <i>Wine</i>	34

DAFTAR TABEL

Tabel	Halaman
1. Penaksiran nilai <i>silhouette</i>	13
2. Kumpulan data set.....	14
3. Data set <i>Iris</i>	17
4. Data set <i>Iris</i> yang telah di transformasi	18
5. <i>Medoid</i> awal.....	20
6. Perhitungan jarak	21
7. Pengelompokan data	21
8. Pengelompokan <i>cluster</i> data	21
9. <i>Medoid</i> baru	22
10. Hasil pengelompokan dengan <i>medoid</i> baru	23
11. <i>Medoid</i> percobaan 2 <i>cluster</i>	24
12. Anggota <i>cluster</i> 1 percobaan 2 <i>cluster</i>	24
13. Anggota <i>cluster</i> 2 percobaan 2 <i>cluster</i>	25
14. Hasil perhitungan SSE	26
15. Hasil perhitungan antar <i>medoid</i>	28
16. Hasil perhitungan rasio	28
17. Hasil perhitungan DBI	29
18. Hasil perhitungan $S(i)$	30
19. Hasil perhitungan SC	31
20. Hasil perhitungan SSE data set <i>Iris</i>	31
21. Hasil perhitungan DBI data set <i>Iris</i>	32
22. Hasil perhitungan SC data set <i>Iris</i>	32
23. Hasil perhitungan SSE data set <i>Wine</i>	33
24. Hasil perhitungan DBI data set <i>Wine</i>	33
25. Hasil perhitungan SC data set <i>Wine</i>	34
26. Komparasi perhitungan keseluruhan data.....	34
27. Pembagian kelompok data.....	35

DAFTAR LAMPIRAN

Lampiran

1. Data set *Iris*
2. Data set *Wine*
3. *Script python*