



**EVALUASI METODE *ELLBOW*, *DAVIES BOULDIN INDEX* (DBI) DAN  
*SILHOUETTE COEFFICIENT* DALAM MEMPEROLEH *K* OPTIMAL  
PADA ALGORITMA *K-MEANS***

**Skripsi**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer**

**Oleh**

**M.FIQRIAN HAFIZ**

**1611016210013**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU**

**JUNI 2023**

## SKRIPSI

### EVALUASI METODE *ELLBOW*, *DAVIES BOULDIN INDEX (DBI)* DAN *SILHOUETTE COEFFICIENT* DALAM MEMPEROLEH K OPTIMAL PADA ALGORITMA *K-MEANS*

Oleh:

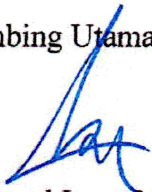
**M.FIQRIAN HAFIZ**

**NIM. 1611016210013**

telah dipertahankan di depan Dosen Penguji pada tanggal 23 Juni 2023

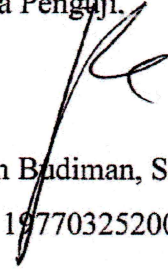
Susunan Dosen Penguji:

Pembimbing Utama,



Muhammad Itqan Mazdadi S.Kom, M.Kom  
NIP. 199006122019031013

Ketua Penguji,



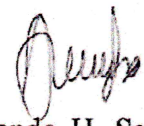
Irwan Budiman, S.T., M.Kom  
NIP. 197703252008121001

Pembimbing Pendamping,



Radityo Adi Nugroho, S.T., M.Kom.  
NIP. 198212042008011006

Anggota Penguji,



Triando H. Saragih.S.Kom,M.Kom  
NIP. 199308242019031012

**Banjarbaru, 23 Juni 2023**

Ketua Program Studi Ilmu Komputer



Irwan Budiman, S.T., M.Kom.

NIP. 197703252008121001

## 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, 2023

Yang menyatakan,



**M Fiqrian Hafiz**  
**NIM. 1611016210013**

# EVALUASI METODE *ELLBOW*, *DAVIES BOULDIN INDEX (DBI)* DAN *SILHOUETTE COEFFICIENT* DALAM MEMPEROLEH *K* OPTIMAL PADA ALGORITMA *K-MEANS*

(M.Fiqrian Hafiz; Pembimbing : Muhammad Itqan Mazdadi, S.Kom. M.Kom dan Radityo Adi Nugroho, S.T, M.Kom; 73 Halaman)

## ABSTRAK

*Clustering* merupakan bagian dari salah satu *Data Mining* yang memiliki tujuan pengelompokan data berdasarkan nilai yang memiliki kemiripan data satu dengan data lainnya. *Clustering* memiliki kelemahan yaitu tidak dapat diketahuinya jumlah *cluster* yang optimal untuk digunakan. Mengatasi kekurangan dari proses *clustering* yang akan dilakukan pada penelitian ini yaitu evaluasi tiga metode yaitu *Elbow*, *Davies-Bouldin Index* dan *Silhouette Coefficient* yang mana ketiga metode tersebut dapat diketahui *k* optimal. Hasil penelitian menunjukkan bahwa metode *Elbow* menyarankan jumlah cluster yang berbeda dari metode *DBI* dan *Silhouette* untuk sebagian besar *Dataset*. Untuk *Dataset Wine*, metode *Elbow* merekomendasikan 3 *Cluster*, sedangkan metode *DBI* dan *Silhouette* merekomendasikan 6 dan 2 cluster masing-masing. Sementara itu, dataset Kanker memiliki rekomendasi *cluster* sebanyak 4 dari metode *Elbow*, 6 dari metode *DBI*, dan 2 dari metode *Silhouette*. Metode clustering yang berbeda ini menghasilkan nilai evaluasi *Clustering* yang beragam. Nilai *Elbow*, *DBI*, dan *Silhouette* juga menunjukkan variasi yang signifikan untuk setiap dataset. Hasil penelitian ini menyoroti pentingnya mempertimbangkan berbagai metode *Clustering* dan evaluasi saat menganalisis data yang kompleks dan beragam. Dalam konteks ini, keputusan tentang jumlah *Cluster* yang optimal dapat bervariasi tergantung pada metode yang digunakan dan karakteristik dataset yang diamati. Oleh karena itu, pemilihan metode *Clustering* yang tepat sangat penting untuk mendapatkan hasil *Clustering* yang akurat dan relevan dalam analisis data.

**Kata Kunci** : *K*-optimal, *Elbow*, *Davies Bouldin Index*, *Silhouette Coefficient*, *K-Means*

# **AN EVALUATION OF THE ELBOW METHOD, DAVIES-BOULDIN INDEX (DBI), AND SILHOUETTE COEFFICIENT IN OBTAINING THE OPTIMAL K FOR THE K-MEANS ALGORITHM**

(M.Fiqrian Hafiz; Supervisors: Muhammad Itqan Mazdadi, S.Kom., M.Kom. and Radityo Adi Nugroho, S.T., M.Kom.; 73 Pages)

## **ABSTRACT**

Clustering is a part of Data Mining that aims to group data based on similarity between data points. However, clustering has a drawback, which is the inability to determine the optimal number of clusters to use. To address this limitation, this research will evaluate three methods: Elbow, Davies-Bouldin Index, and Silhouette Coefficient, all of which can help identify the optimal 'k' value. The research findings indicate that the Elbow method suggests a different number of clusters compared to the DBI and Silhouette methods for most datasets. For the Wine dataset, the Elbow method recommends 3 clusters, while the DBI and Silhouette methods recommend 6 and 2 clusters, respectively. On the other hand, the Cancer dataset has recommendations of 4 clusters from the Elbow method, 6 from the DBI method, and 2 from the Silhouette method. These different clustering methods result in diverse evaluation values. The Elbow, DBI, and Silhouette values also exhibit significant variation for each dataset. This research highlights the importance of considering various clustering methods and evaluations when analyzing complex and diverse data. In this context, the decision regarding the optimal number of clusters can vary depending on the method used and the characteristics of the observed dataset. Therefore, selecting the appropriate clustering method is crucial to obtain accurate and relevant clustering results in data analysis.

**Keywords:** Optimal K, Elbow, Davies Bouldin Index, Silhouette Coefficient, K-Means

## PRAKATA

Puji dan syukur kepada Allah SWT karena atas berkat rahmat-Nya penulis dapat menyelesaikan skripsi yang berjudul “Perbandingan Metode *Ellbow* dan *Davies Bouldin Index (DBI)* Dalam Memperoleh K Optimal Pada Algoritma *K-Means* ” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Prakata ini kami persembahkan sebagai ungkapan rasa syukur dan penghargaan yang tulus atas segala bantuan, dukungan, serta dorongan yang telah kami terima selama proses penulisan skripsi ini.

Skripsi ini adalah hasil dari perjalanan panjang yang tak terlupakan dalam menempuh pendidikan di perguruan tinggi. Selama proses ini, kami telah mendapatkan banyak ilmu, pelajaran, dan pengalaman berharga yang akan membekas sepanjang hayat.

Kami ingin menyampaikan terima kasih yang tak terhingga kepada seluruh dosen pembimbing kami, Muhammad Itqan Mazdadi S.Kom, M.Kom dan Radityo Adi Nugroho S.T, M.Kom atas bimbingan, arahan, dan dukungan penuh yang telah diberikan. Bantuan kalian telah memandu kami untuk mengembangkan skripsi ini dengan baik.

Kami juga berterima kasih kepada seluruh dosen dan staf pengajar di Universitas Lambung Mangkurat yang telah memberikan ilmu pengetahuan dan wawasan yang berharga dalam perjalanan studi kami.

Tak lupa pula, ucapan terima kasih kami sampaikan kepada keluarga kami yang selalu memberikan doa, dukungan, dan motivasi agar kami tetap semangat dalam menyelesaikan skripsi ini.

Kami mengakui bahwa penulisan skripsi ini tidak terlepas dari berbagai keterbatasan dan tantangan. Namun, semua itu telah membantu kami untuk tumbuh dan berkembang sebagai pribadi yang lebih tangguh dan bertanggung jawab.

Semoga Allah SWT senantiasa memberikan rahmat, hidayah, dan keberkahan dalam setiap langkah perjalanan hidup kami. Amin.

Banjarbaru, Maret 2023



M Fiqrihan Hafiz

## DAFTAR ISI

HALAMAN JUDUL.....	i
SKRIPSI .....	ii
PERNYATAAN .....	iii
ABSTRAK.....	iv
ABSTRACT.....	v
PRAKATA .....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xii
<b>BAB I PENDAHULUAN</b>	
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	1
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	3
1.6 Sistematika Penulisan.....	4
<b>BAB II TINJAUAN PUSTAKA</b>	
2.1 <i>Data Mining</i> .....	5
2.2 <i>Data Clustering</i> .....	9
2.3 <i>K-Means</i> .....	12
2.4 <i>Ellbow</i> .....	14
2.5 <i>Davies Bouldi Index</i> .....	16
2.6 <i>Silhoutte Cofficient</i> .....	17
2.7 Keaslian Penelitian.....	19
<b>BAB III METODE PENELITIAN</b>	
3.1 Alat Penelitian.....	23
3.2 Bahan Penelitian.....	23
3.3 Prosedur Penelitian.....	24



## **BAB IV HASIL DAN PEMBAHASAN**

4.1	Hasil Pengumpulan Data .....	26
4.2	<i>Pre-Processing</i> .....	26
4.3	Algoritma <i>K-Means</i> .....	27
4.4	K-Optimal .....	31
4.5	Evaluasi .....	39
4.6	Pembahasan.....	43

## **BAB V PENUTUP**

5.1	Kesimpulan .....	45
5.2	Saran .....	45
DAFTAR PUSTAKA .....		47
LAMPIRAN.....		49
RIWAYAT HIDUP.....		55

## DAFTAR TABEL

Tabel 1. Peranan <i>Data Mining</i> .....	7
Tabel 2 Representasi <i>Kauffman</i> dan <i>Rousseuw</i> .....	18
Tabel 3 Keaslian Penelitian.....	19
Tabel 4. Perancangan Penelitian .....	22
Tabel 5. Kumpulan Data Set.....	24
Tabel 6 Data Set.....	26
Tabel 7 Hasil <i>Presprocessing Data</i> .....	27
Tabel 8 Tabel <i>Data Set (Wine)</i> .....	27
Tabel 9 Tabel <i>Centroid Awal</i> .....	27
Tabel 10 Hasil Perhitungan Kuadrat Jarak <i>Cluster</i> .....	28
Tabel 11 Hasil Penentuan <i>Cluster</i> .....	28
Tabel 12 Tabel <i>Centroid Akhir</i> .....	31
Tabel 13 Tabel Kumpulan <i>Cluster 1</i> .....	32
Tabel 14 Tabel <i>Centroid Titik 1</i> .....	32
Tabel 15 Tabel Hasil Jarak.....	33
Tabel 16 Tabel Kumpulan <i>Cluster 2</i> .....	33
Tabel 17 Tabel <i>Centroid Cluster 2</i> .....	34
Tabel 18 Hasil Perhitungan Jarak .....	34
Tabel 19 Tabel SSE.....	35
Tabel 20 Hasil Perhitungan SSW 1.....	36
Tabel 21 Hasil Perhitungan SSW 2.....	37
Tabel 22 Hasil Perhitungan Memperoleh Nilai SSB .....	37
Tabel 23 Hasil Perhitungan Memperoleh Nilai Rasio .....	38
Tabel 24 Tabel Hasil Keseluruhan Perhitungan DBI.....	38
Tabel 25 Hasil Perhitungan Mencari Nilai Ai.....	40
Tabel 26 Keseluruhan Hasil Perhitungan Ellbow <i>Dataset Wine</i> .....	44
Tabel 27 Keseluruhan Hasil Perhitungan <i>DBI Dataset Wine</i> .....	44
Tabel 28 Keseluruhan Hasil Perhitungan <i>Silhouette Dataset Wine</i> .....	44
Tabel 29 Keseluruhan Hasil Perhitungan <i>Ellbow Dataset Kanker</i> .....	45
Tabel 30 Keseluruhan Hasil Perhitungan <i>DBI Dataset Kanker</i> .....	45

Tabel 31 Keseluruhan Hasil Perhitungan <i>Silhouette</i> Dataset Kanker .....	45
Tabel 32 Keseluruhan Hasil Perhitungan <i>Ellbow</i> Dataset Hewan .....	46
Tabel 33 Keseluruhan Hasil Perhitungan <i>DBI</i> Dataset Hewan.....	46
Tabel 34 Keseluruhan Hasil Perhitungan <i>Silhouette</i> Dataset Hewan .....	46
Tabel 35 Keseluruhan Hasil Perhitungan <i>Ellbow Dataset Glass</i> .....	47
Tabel 36 Keseluruhan Hasil Perhitungan <i>DBI Dataset Glass</i> .....	47
Tabel 37 Keseluruhan Hasil Perhitungan <i>Silhouette Dataset Glass</i> .....	47
Tabel 38 Keseluruhan Hasil Perhitungan <i>Ellbow Dataset Pendigits</i> .....	48
Tabel 39 Keseluruhan Hasil Perhitungan <i>DBI Dataset Pendigits</i> .....	48
Tabel 40. Keseluruhan Hasil Perhitungan <i>Silhouette Dataset Pendigits</i> .....	48
Tabel 41 Hasil Perhitungan Keseluruhan Data .....	49

## DAFTAR GAMBAR

Gambar 1 Tahapan pada proses <i>data mining</i> .....	6
Gambar 2 Penentuan K Optimal menggunakan <i>Ellbow</i> .....	15
Gambar 3 Penentuan K Optimal menggunakan <i>Silhouette Coefficient</i> .....	19
Gambar 4 Alur Penelitian.....	25
Gambar 5 <i>Cluster</i> optimal dengan metode <i>Ellbow</i> .....	35
Gambar 6 Grafik Ellbow Dataset Wine .....	44
Gambar 7 Grafik DBI Dataset Wine .....	44
Gambar 8 Grafik Silhouette Dataset Wine.....	44
Gambar 9 Grafik Ellbow Dataset Kanker .....	45
Gambar 10 Grafik DBI Dataset Kanker .....	45
Gambar 11 Grafik Silhouette Dataset Kanker .....	45
Gambar 12 Grafik Ellbow Dataset Hewan.....	46
Gambar 13 Grafik DBI Dataset Hewan .....	46
Gambar 14 Grafik Silhouette Dataset Hewan .....	46
Gambar 15 Grafik Ellbow Dataset Glass .....	47
Gambar 16 Grafik DBI Dataset Glass.....	47
Gambar 17 Grafik Silhouette Dataset Glass .....	47
Gambar 18 Grafik Ellbow Dataset Pendigits .....	48
Gambar 19 Grafik DBI Dataset Pendigits.....	48
Gambar 20 Grafik Silhouette Dataset Pendigits .....	48

## DAFTAR LAMPIRAN

Lampiran 1 Hasil Dataset Glass .....	55
Lampiran 2 Hasil Dataset Hewan .....	56
Lampiran 3 Hasil Dataset Kanker .....	57
Lampiran 4 Hasil Dataset Wine .....	58
Lampiran 5 Hasil Dataset Pendigits .....	59