

Perbandingan Metode *K-Means* dan Metode *DBSCAN* pada Pengelompokan Saham Indeks LQ45

Iqbal Ali Mansyah Wardana¹⁾, Finda Nur Lailiyah²⁾, Noviyanti Santoso³⁾

¹⁾²⁾³⁾ Departemen Statistika Bisnis, Fakultas Vokasi, Institut Teknologi Sepuluh Nopember
Kampus ITS Sukolilo, 60111, Kota Surabaya

¹⁾ bgiqbalwardana18@gmail.com

²⁾ findanur23@gmail.com

³⁾ noviyanti.santoso@its.ac.id

Abstrak

Peningkatan minat masyarakat terhadap investasi, terutama di kalangan muda, telah menyebabkan lonjakan jumlah investor di Indonesia sejak 2019. Kondisi ini diiringi dengan semakin banyaknya perusahaan yang melantai di Bursa Efek Indonesia (BEI), khususnya pada indeks saham LQ45 yang menjadi salah satu acuan utama investor. Namun, dinamika pasar saham sering menyulitkan investor dalam memilih saham dengan kinerja keuangan yang optimal. Analisis rasio keuangan dari laporan keuangan merupakan alat penting untuk menilai kinerja perusahaan. Akan tetapi, data rasio keuangan bersifat numerik dan belum terkelompokkan secara jelas, sehingga diperlukan metode pengelompokan untuk memberikan pemahaman yang lebih baik mengenai pola kinerja perusahaan. Penelitian ini membandingkan dua metode klustering, yaitu *K-Means* dan *DBSCAN*, dalam mengelompokkan saham-saham pada indeks LQ45 berdasarkan rasio keuangan seperti *Return on Equity* (ROE), *Net Profit Margin* (NPM), *Debt to Equity Ratio* (DER), *Debt to Asset Ratio* (DAR), Ukuran Perusahaan (UP), dan Volume Perdagangan Saham dari periode Februari hingga Juli 2023. *K-Means* dikenal efektif untuk data yang terdistribusi secara jelas, sementara *DBSCAN* unggul dalam menangani data dengan bentuk distribusi yang kompleks dan keberadaan *outlier*. Hasil penelitian ini menunjukkan bahwa metode *DBSCAN* lebih baik dalam mengelompokkan perusahaan karena memiliki nilai *Silhouette Score* dan *Index Dunn* yang lebih tinggi dibanding dengan metode *K-Means*.

Kata kunci: *DBSCAN, Klustering, K-Means, Kinerja Keuangan Perusahaan.*

Abstract

The increase in public interest in investment, especially among young people, has caused a surge in the number of investors in Indonesia since 2019. This condition is accompanied by the increasing number of companies listed on the Indonesia Stock Exchange (BEI), especially on the LQ45 stock index which is one of the main references for investors. However, stock market dynamics often make it difficult for investors to choose shares with optimal financial performance. Financial ratio analysis of financial reports is an important tool for assessing company performance. However, financial ratio data is numerical and has not been clearly grouped, so a grouping method is needed to provide a better understanding of company performance patterns. This research compares two clustering methods, namely *K-Means* and *DBSCAN*, in grouping shares in the LQ45 index based on financial ratios such as *Return on Equity* (ROE), *Net Profit Margin* (NPM), *Debt to Equity Ratio* (DER), *Debt to Asset Ratio* (DAR), *Company Size* (UP), and *Stock Trading Volume* from the period February to July 2023. *K-Means* is known to be effective for clearly distributed data, while *DBSCAN* excels in handling data with complex distribution forms and the existence of outliers. The results of this research indicate that the *DBSCAN* method is better in grouping companies because it has higher *Silhouette Score* and *Dunn Index* values compared to the *K-Means* method.

Keywords: *Company Financial Performance, Clustering, DBSCAN, K-Means.*

1. PENDAHULUAN

Minat masyarakat, terutama kalangan muda, terhadap investasi mengalami peningkatan signifikan sejak 2019 hingga 2022. Pada 2021, jumlah investor di Indonesia tercatat meningkat sebesar 92,99%, dan pada 2022 jumlah ini kembali naik sebesar 33,53% [1]. Lonjakan jumlah investor ini diikuti oleh semakin banyaknya perusahaan yang memilih untuk *go public* dan menawarkan saham mereka melalui Bursa Efek Indonesia (BEI). Meningkatnya jumlah perusahaan yang terdaftar memicu persaingan antarperusahaan untuk memperluas pangsa pasar. Hal ini menyebabkan kondisi saham menjadi dinamis, sehingga investor sering kali kesulitan dalam menentukan saham dengan kinerja keuangan yang baik untuk mengoptimalkan risiko dan return investasi. Dalam menentukan perusahaan yang tepat untuk berinvestasi, analisis rasio keuangan dari laporan keuangan menjadi alat penting. Analisis ini memberikan gambaran kinerja keuangan perusahaan baik pada periode saat ini maupun di masa mendatang [2]. Namun, data rasio keuangan bersifat numerik dan belum terkelompokkan secara jelas berdasarkan karakteristik kinerja keuangan perusahaan. Oleh karena itu, diperlukan metode pengelompokan untuk mengidentifikasi perusahaan dengan kinerja keuangan serupa. Cara ini sangat efektif dalam mengidentifikasi pola tersembunyi pada data yang besar dan kompleks, sehingga dapat mendukung investor memilih saham yang tepat untuk investasi [3].

Indeks LQ45, yang diluncurkan pertama kali pada Februari 1997 oleh BEI, sering menjadi acuan bagi investor dalam memilih saham. Indeks ini terdiri dari perusahaan dengan likuiditas tinggi, kapitalisasi pasar besar, serta didukung fundamental yang kuat. BEI melakukan evaluasi berkala terhadap indeks ini setiap enam bulan, sehingga perusahaan yang terdaftar dalam LQ45 merupakan perusahaan pilihan yang memenuhi kriteria tertentu. Saham dalam indeks LQ45 dianggap lebih aman dan berpotensi memberikan return stabil dalam jangka panjang.

Metode klustering sering digunakan untuk mengelompokkan perusahaan berdasarkan karakteristik kinerja keuangan. Klustering adalah teknik analisis multivariat yang bertujuan untuk mengelompokkan objek-objek ke dalam kelompok yang homogen berdasarkan karakteristik tertentu [4]. Dua metode klustering yang populer adalah *K-Means* dan *Density-Based Spatial Klustering of Applications with Noise* (DBSCAN). Metode *K-Means* bekerja dengan membagi data ke dalam kelompok berdasarkan kedekatan antar data menggunakan *centroid* sebagai pusat kelompok. Metode ini efektif untuk data dengan distribusi yang jelas dan tanpa *outlier*. Sebaliknya, *DBSCAN* mampu mengelompokkan data dengan bentuk distribusi yang kompleks dan lebih tahan terhadap *outlier* karena pendekatannya berbasis kepadatan [5]. Kedua metode ini memiliki keunggulan dan kekurangan, sehingga perlu dilakukan perbandingan untuk menentukan metode yang lebih sesuai dalam pengelompokan saham indeks LQ45. Penelitian ini menggunakan *silhouette score* untuk mengevaluasi kualitas kluster yang dihasilkan oleh metode *K-Means* dan *DBSCAN*. Semakin tinggi nilai *silhouette score*, semakin baik pemisahan antar kluster yang terbentuk [6].

Penelitian ini bertujuan untuk membandingkan performa metode *K-Means* dan *DBSCAN* dalam mengelompokkan saham yang terdaftar pada indeks LQ45 berdasarkan rasio keuangan seperti *Return on Equity* (ROE), *Net Profit Margin* (NPM), dan *Debt to Equity Ratio* (DER). Serta secara khusus mengkaji kinerja kedua algoritma pada data saham yang memiliki karakteristik data kompleks dan mengandung *outlier*. Hasil pengelompokan diharapkan dapat memberikan informasi yang bermanfaat bagi investor dalam memahami pola kinerja perusahaan, sehingga memudahkan pengambilan keputusan investasi. Melalui penelitian ini, diharapkan metode yang paling optimal untuk pengelompokan saham dapat diidentifikasi, memberikan kontribusi signifikan bagi literatur mengenai penerapan analisis klustering pada pasar saham.

2. TINJAUAN PUSTAKA

2.1 *K-Means*

Metode *K-Means* adalah teknik pengelompokan berbasis jarak, di mana semakin dekat jarak antar sampel, semakin tinggi tingkat kemiripannya. Sebaliknya, objek dalam kluster yang berbeda memiliki tingkat kemiripan karakteristik yang sangat rendah. Metode ini membagi sampel

menjadi k kelompok berdasarkan fungsi jarak, dengan menggunakan jarak *Euclidean* untuk menghitung kedekatan antar sampel. Jarak *Euclidean* membantu menentukan data mana yang akan dimasukkan ke dalam kluster dengan *centroid* terdekat. Untuk menentukan jumlah kluster yang optimal, digunakan metode *elbow* dan *silhouette*. Metode *elbow* dilakukan dengan mengamati titik di mana grafik jumlah kluster membentuk siku [7]. Jika penurunan terbesar terjadi antara kluster pertama dan kedua, maka jumlah kluster tersebut dianggap paling optimal. Semakin banyak kluster yang digunakan, semakin kecil varians dalam masing-masing kluster. Sedangkan, Metode *silhouette* digunakan untuk menilai kualitas dan kekuatan kluster, dengan mengevaluasi seberapa baik sebuah objek ditempatkan dalam kluster tertentu [8]. Nilai *silhouette* berkisar antara -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan bahwa data berada di kluster yang tepat dan terpisah jauh dari kluster lainnya. Metode ini mengukur jarak antara titik data dengan kluster lain dibandingkan dengan jaraknya ke kluster tempat data tersebut berada. Untuk menilai sejauh mana pengelompokan kluster dilakukan dengan baik, digunakan rumus *Average Silhouette Width* (ASW) yang dijelaskan sebagai berikut.

$$A(i)_i = \frac{1}{|n_k|-1} \sum_{k=1}^q d(X_i, X_r) \quad (1)$$

$$B(i)_i = \min_{k \neq i} \frac{1}{|\mu_k|-1} \sum_{k=1}^q d(X_i, X_r) \quad (2)$$

$$ASW = \frac{1}{n} \sum_{i=1}^n \frac{B(i)_i - A(i)_i}{\max \{A(i)_i, B(i)_i\}} \quad (3)$$

Keterangan:

- i = Jumlah observasi ($i = 1, 2, \dots, n$)
- k = Jumlah kluster optimum ($k = 1, 2, \dots, q$)
- l = Pusat kluster (*centroid*) ($l = 1, 2, \dots, p$)
- $A(i)_i$ = Rata-rata jarak antara titik ke- i dengan semua titik data lain dalam kluster
- $B(i)_i$ = Jarak minimum rata-rata antara titik ke- i dengan titik data dalam kluster lain
- n_k = Jumlah observasi dalam kluster k
- $d(X_i, X_r)$ = Jarak *Euclidean* antara data ke- i dalam kluster l

Metode *K-Means* mengelompokkan objek berdasarkan jarak terdekat ke pusat kluster. Proses klastering dianggap selesai ketika pusat kluster tidak lagi mengalami perubahan. Langkah-langkah dalam metode *K-Means* adalah sebagai berikut.

1. Menentukan jumlah kluster optimum, yaitu nilai k , yang dapat ditentukan menggunakan metode *elbow* atau metode *silhouette*.
2. Memilih nilai acak sebagai *centroid* awal dan menghitung jarak setiap titik data ke masing-masing *centroid* menggunakan rumus jarak *euclidean* yang ditunjukkan dalam bentuk persamaan sebagai berikut.

$$d(X_i, C_l) = \sqrt{\sum_{l=1}^p \sum_{i=1}^n (X_{ijk} - C_{lk})^2} \quad (4)$$

Keterangan:

- i = Jumlah observasi ($i = 1, 2, \dots, n$)
 - j = Jumlah variabel yang digunakan ($j = 1, 2, \dots, m$)
 - k = Jumlah kluster optimum ($k = 1, 2, \dots, q$)
 - l = Pusat kluster (*centroid*) ($l = 1, 2, \dots, p$)
 - $d(X_i, C_l)$ = jarak *Euclidean* antara titik X_i dan pusat kluster C_l
 - X_{ijk} = Titik data ke- i dan variabel ke- j dalam kluster k
 - C_{lk} = Pusat kluster ke- l dalam kluster k
3. Mengelompokkan setiap data ke kluster berdasarkan jarak terdekat dengan *centroid*.

4. Memperbarui *centroid* baru dengan menghitung rata-rata data dalam masing-masing kluster menggunakan rumus sebagai berikut.

$$C_{jl} = \frac{1}{n_k} \sum_{i=1}^n X_{ijl} \quad (5)$$

Keterangan:

- i = Jumlah observasi ($i = 1, 2, \dots, n$)
 - j = Jumlah variabel yang digunakan ($j = 1, 2, \dots, m$)
 - l = Pusat kluster (*centroid*) ($l = 1, 2, \dots, p$)
 - C_{jl} = Pusat kluster ke- l dan variabel ke- j
 - X_{ijl} = Titik data ke- i dan variabel ke- j dalam pusat kluster l
 - n_k = Jumlah observasi dalam kluster k
5. Jika komposisi data dalam kluster masih berubah, ulangi langkah 2 hingga 4 sampai anggota setiap kluster stabil dan tidak berubah lagi.

2.2 DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

Algoritma *DBSCAN* (*Density-Based Spatial Clustering of Applications with Noise*) adalah metode pengelompokan data berbasis kepadatan yang populer dan sering digunakan dalam berbagai literatur ilmiah merupakan salah satu algoritma yang digunakan untuk pengelompokan data yang dimana pengelompokan data algoritma berbasis kepadatan [9]. *DBSCAN* membentuk kluster dengan mengidentifikasi area yang memiliki kepadatan tinggi berdasarkan konektivitas densitas [10]. Algoritma *DBSCAN* membutuhkan parameter dua parameter penting yaitu radius (ϵ) dan jumlah minimum titik (MinObj) untuk membentuk kluster. Untuk menentukan ϵ dan MinObj yang optimal, dapat digunakan grafik *k-dist* dengan mengamati titik perubahan tajam pada nilai ϵ , sementara k diambil sebagai MinObj [11]. Adapun prinsip-prinsip dasar dari *DBSCAN* dapat dijelaskan sebagai berikut.

1. *Neighborhood* dalam radius ϵ disebut dengan ϵ -neighborhood dari sebuah objek.
2. Objek dalam ϵ -neighborhood yang memiliki jarak terkecil dalam sebuah kluster tersebut disebut *core object*.
3. Objek p disebut *density reachable* dari objek q (dengan ϵ dan MinObj) jika terdapat suatu rantai objek $p_1, p_2, \dots, p_n$, di mana $p_1 = q$ dan $p_n = p$, dan setiap p_{i+1} *density reachable* secara langsung dari p_i dengan respek ke ϵ dan MinObj, untuk $1 \leq i \leq n$, p_i anggota D .
4. Suatu objek p adalah *density connected* ke objek q dengan respek ke ϵ dan MinObj dalam suatu set objek D jika terdapat suatu objek o anggota D dengan kedua p dan q adalah *density reachable* dari o dengan respek ke ϵ dan MinObj.

2.3 Indeks LQ45

Indeks saham adalah indikator statistik yang mencerminkan pergerakan harga secara keseluruhan dari sekelompok saham yang dipilih berdasarkan kriteria dan metodologi tertentu, serta dievaluasi secara berkala [12]. Indeks LQ45 adalah indeks yang mengukur kinerja harga dari 45 saham yang memiliki likuiditas tinggi dan kapitalisasi pasar besar serta didukung oleh fundamental perusahaan yang baik. Indeks LQ45 direview setiap tiga bulan untuk memastikan saham-saham yang masuk masih memenuhi kriteria yang telah ditetapkan. Pergantian saham dalam indeks dilakukan setiap enam bulan, yaitu pada bulan Februari dan Agustus.

3. METODE PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian adalah data sekunder yang merupakan rekapitulasi laporan keuangan dari perusahaan atau emiten BEI (Bursa Efek Indonesia) yang terdaftar dalam indeks LQ45 pada periode bulan Februari sampai dengan Juli 2023. Data yang digunakan adalah data ROE, DAR, DER, UP (ukuran perusahaan), NPM, dan VOL (volume perdagangan saham) tahun 2022 yang selanjutnya disebut indikator kinerja keuangan perusahaan. Data diperoleh dari

IDX Company Fact Sheet LQ45 pada website www.idx.co.id. Namun, terdapat satu emiten, yaitu GOTO yang tidak memenuhi kriteria kelengkapan data. Oleh karena itu, yang digunakan dalam penelitian ini menjadi 44 emiten

3.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian terdiri atas variabel prediktor yang disimbolkan dengan X. Variabel penelitian dapat dituliskan dalam tabel sebagai berikut.

Tabel 1. Variabel Penelitian			
Variabel	Keterangan	Skala	Satuan
X ₁	ROE (Return on Equity)	Rasio	Persen (%)
X ₂	DAR (Debt to Assets Ratio)	Rasio	Kali
X ₃	DER (Debt to Equity Ratio)	Rasio	Kali
X ₄	UP (Ukuran Perusahaan)	Rasio	-
X ₅	NPM (Net Profit Margin)	Rasio	Persen (%)
X ₆	VOL (Volume Perdagangan Saham)	Rasio	Juta Saham

3.3 Langkah Analisis

Langkah analisis yang digunakan untuk menyelesaikan permasalahan dalam penelitian ini adalah sebagai berikut.

1. Melakukan perencanaan penelitian
 - a) Mengidentifikasi permasalahan penelitian. Permasalahan pada penelitian ini adalah bagaimana menentukan metode klustering yang optimal antara *DBSCAN* dan *K-Means* untuk pengelompokan saham indeks LQ45.
 - b) Merumuskan tujuan penelitian, yaitu membandingkan performa metode *K-Means* dan *DBSCAN* dalam pengelompokan saham indeks LQ45 serta menentukan kluster optimal dengan *Silhouette Score*.
 - c) Melakukan kajian literatur terkait metode klustering, validasi kluster menggunakan *Silhouette Score*, dan pengelompokan saham.
 - d) Menentukan sumber data yang akan digunakan, seperti laporan keuangan perusahaan di indeks LQ45 tahun 2023, dan memilih variabel seperti ROE, NPM, dan DER.
2. Mengumpulkan data laporan keuangan dari perusahaan yang termasuk dalam indeks LQ45.
3. Melakukan *pre-processing* data, dengan normalisasi data.
4. Menganalisis saham menggunakan metode klustering *K-Means* dan *DBSCAN*.
 - a) Metode *K-Means*
 - Melakukan klustering menggunakan algoritma *K-Means* dengan jumlah kluster tertentu.
 - Memvariasikan jumlah kluster (*k*) untuk menentukan hasil klustering yang optimal.
 - b) Metode *DBSCAN*
 - Melakukan klustering menggunakan algoritma *DBSCAN* dengan berbagai parameter.
 - Mengidentifikasi hasil klustering, termasuk jumlah kluster yang terbentuk dan data yang terklasifikasi sebagai *noise*.
5. Menganalisis validasi pada kluster yang terpilih menggunakan *Silhouette Score* (SI) dan *Index Dunn* sebagai metrik untuk menilai kualitas kluster.
6. Membandingkan performa kedua metode berdasarkan *Silhouette Score*.
7. Menginterpretasikan hasil klustering untuk melihat karakteristik setiap kluster yang terbentuk.
8. Menarik kesimpulan dan saran.

4. PEMBAHASAN

4.1 Pre-Processing Data

Tahap ini merupakan langkah awal sebelum proses clusterisasi. Proses ini mencakup penghapusan fitur yang tidak relevan, serta proses pembersihan data bertujuan menghasilkan dataset yang lebih bersih dan konsisten. Setelah itu dilakukan tahap normalisasi data ini

menggunakan metode standarisasi *Z-score scaling*, melalui pendekatan ini, seluruh atribut memiliki skala yang seragam, sehingga dapat meminimalkan bias akibat perbedaan rentang nilai dan meningkatkan kualitas hasil analisis.

4.2 Karakteristik Data Kinerja Perusahaan

Bagian ini menjelaskan karakteristik variabel pembentuk klaster sebelum dinormalisasi yang terdiri dari, *Return on Equity* (ROE), *Debt to Assets Ratio* (DAR), *Net Profit Margin* (NPM), *Debt to Equity Ratio* (DER), Ukuran Perusahaan, dan Volume Perdagangan Saham. Dijelaskan sebagai berikut.

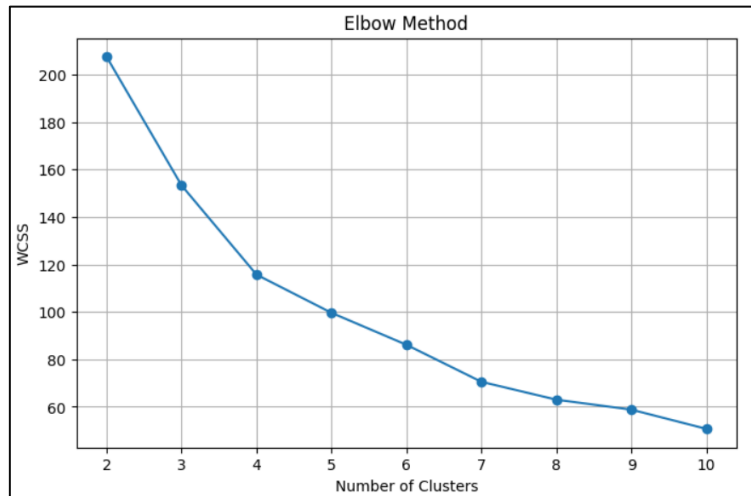
Tabel 2. Karakteristik Data

Variabel	Mean	Minimum	Maksimum
ROE	16,08	-3,91	46,01
DAR	0,47	0,06	0,87
NPM	21,30	-5,70	139,72
DER	1,91	0,07	13,56
Ukuran Perusahaan	18,08	15,22	21,41
Volume Perdagangan Saham	1249	71	5768

Tabel 2 menunjukkan bahwa karakteristik data pada perusahaan yang terdaftar dalam indeks saham LQ45 tahun 2023. ROE pada perusahaan yang terdaftar dalam indeks saham LQ45 tahun 2023 memiliki rata-rata sebesar 16,08. Perusahaan dengan ROE terkecil adalah PT Chandra Asri Pacific Tbk (TPIA) perusahaan yang bergerak dalam bisnis usaha pengolahan dan manufaktur petrokimia dengan nilai ROE sebesar -3,91, sedangkan perusahaan dengan ROE terbesar adalah PT Indo Tambangraya Megah Tbk (ITMG) perusahaan tersebut bergerak dalam penambangan dan penjualan batu bara dengan nilai ROE sebesar 46,01. DAR pada perusahaan indeks saham LQ45 tahun 2023 memiliki rata-rata sebesar 0,47. Perusahaan dengan DAR terkecil adalah PT Bukalapak.com Tbk (BUKA) perusahaan yang bergerak di sektor *e-commerce* Indonesia dengan nilai DAR sebesar 0,06, sedangkan perusahaan dengan DAR terbesar adalah PT. Bank Tabungan Negara (Persero) Tbk (BBTN) yang bergerak pada sektor perbankan dengan nilai DAR sebesar 0,87. NPM pada perusahaan indeks saham LQ45 tahun 2023 memiliki rata-rata sebesar 21,30 dan. Perusahaan dengan NPM terkecil adalah TPIA dengan NPM sebesar -5,70, sedangkan perusahaan dengan NPM terbesar adalah BUKA dengan NPM sebesar 139.

DER pada perusahaan indeks saham LQ45 tahun 2023 memiliki rata-rata sebesar 1,91. Perusahaan dengan DER terkecil adalah BUKA dengan DER sebesar 0,07, sedangkan perusahaan dengan DER terbesar adalah BBTN dengan DER sebesar 13,56. Ukuran perusahaan pada perusahaan indeks saham LQ45 tahun 2023 memiliki rata-rata sebesar 18,08. Perusahaan dengan ukuran perusahaan terkecil adalah PT. Industri Jamu dan Farmasi Sido Muncul Tbk (SIDO) yang bergerak pada industri jamu dan farmasi dengan ukuran perusahaan sebesar 18, sedangkan perusahaan dengan ukuran perusahaan terbesar adalah PT. Bank Mandiri (Persero) Tbk (BMRI) yang bergerak di sektor perbankan dengan ukuran perusahaan sebesar 21,41. Volume perdagangan saham pada perusahaan indeks saham LQ45 tahun 2023 memiliki rata-rata sebesar 1249. Perusahaan dengan volume perdagangan saham terkecil adalah PT Indocement Tunggul Prakarsa Tbk (INTP), salah satu produsen semen terbesar di Indonesia dengan volume perdagangan saham sebesar 71, sedangkan perusahaan dengan volume perdagangan saham terbesar adalah BUKA dengan volume perdagangan saham sebesar 5768.

4.3 Menentukan Jumlah Kluster Optimum *K-Means*

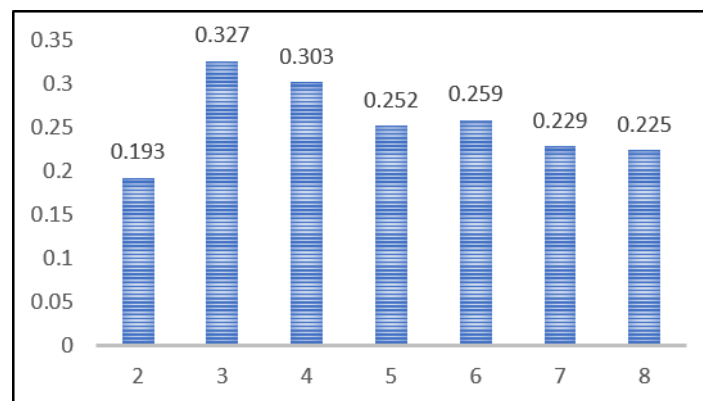


Gambar 1. Kluster optimal

Gambar 1 metode *Elbow* menunjukkan bahwa pada $k=3$ mengalami penurunan yang paling signifikan bila dibandingkan dengan jumlah k yang lain. Jumlah kluster selain $k=3$ mengalami penurunan yang tidak signifikan atau bisa dikatakan stabil. Sehingga dapat disimpulkan bahwa jumlah kluster optimal yang terbentuk adalah sejumlah 3 kluster. Untuk validitas jumlah kluster optimal lebih lanjut dilakukan dengan melihat nilai *silhouette score*.

4.4 Silhouette Score *K-Means*

Pemilihan jumlah optimal kluster dengan melakukan percobaan jumlah luster antara 2-8, setelah itu dilihat nilai *Silhouette Score* terbesar berdasarkan algoritma *K-Means* dapat dilihat Gambar 4 berikut.



Gambar 2. *Silhouette score K-Means*

Gambar 2 menunjukkan bahwa *silhouette score* terbesar terdapat pada jumlah kluster sebanyak 3 atau $k=3$, sehingga jumlah kluster terbaik yang terbentuk pada algoritma *K-Means* dengan membagi sebanyak 3 kluster. Dengan nilai *silhouette score* sebesar 0,327. Hasil klastering dengan nilai $k=3$ dapat ditunjukkan pada tabel 2 berikut.

4.5 *K-Means* Kluster

Pengelompokan perusahaan LQ45 tahun 2023 berdasarkan kinerja keuangan perusahaan menggunakan algoritma *K-Means* dilakukan dengan jumlah kluster optimal, yaitu sebanyak 3. Hasil pengelompokan perusahaan LQ45 tahun 2023 dapat ditunjukkan dalam tabel sebagai berikut.

Tabel 3. Kluster *K-Means* perusahaan

Kluster	Jumlah Perusahaan	Mean					
		ROE	DAR	DER	NPM	UP	VOL
Kluster 1	8	33,16	0,23	0,33	49	17,14	1464
Kluster 2	29	11,03	0,43	1,32	10,80	17,82	958
Kluster 3	7	17,45	0,82	6,13	33,17	20,11	2213

Tabel 3 menunjukkan Kluster 1 menunjukkan perusahaan dengan kinerja keuangan terbaik, ditandai oleh profitabilitas tertinggi (ROE 33,16 dan NPM 49,00) serta tingkat utang yang rendah (DAR 0,23 dan DER 0,33). Kondisi ini mencerminkan perusahaan yang sehat dan berisiko keuangan rendah. Kluster 2 terdiri dari perusahaan dengan kinerja keuangan terendah, terlihat dari nilai ROE (11,03) dan NPM (10,80) yang paling rendah. Meskipun tingkat utangnya masih moderat, kemampuan perusahaan dalam menghasilkan laba relatif kurang optimal. Kluster 3 merupakan perusahaan berukuran besar dengan aktivitas perdagangan saham tertinggi (UP 20,11 dan VOL 2.213). Profitabilitasnya cukup baik (ROE 17,45 dan NPM 33,17), tetapi tingkat utangnya paling tinggi (DAR 0,82 dan DER 6,13), sehingga memiliki risiko keuangan yang lebih besar. Secara umum, kluster 1 menggambarkan perusahaan dengan fundamental paling kuat, kluster 2 menunjukkan perusahaan dengan kinerja relatif rendah, sedangkan kluster 3 merepresentasikan perusahaan besar dan aktif diperdagangkan dengan potensi keuntungan yang baik namun berisiko lebih tinggi.

4.6 Menentukan Jumlah Kluster Optimal *DBSCAN*

Kombinasi percobaan dengan parameter epsilon antara 0,5-2 dan parameter min_sample antara 2-4. Dari kombinasi parameter tersebut, setelah itu dicari hasil *silhouette score* terbaik yang dijadikan sebagai jumlah kluster optimal. Untuk penjelasannya dijelaskan pada Tabel 5 sebagai berikut.

Tabel 4. Kluster optimal *DBSCAN*

Epsilon	Min sample	Jumlah Kluster	<i>Silhouette Score</i>	Noise
0,5	2	2	-0,159	39
0,5	3	1	-	41
0,5	4	0	-	44
0,5	5	0	-	44
1	2	7	-0,005	19
1	3	3	0,007	27
1	4	1	-	36
1	5	1	-	38
1,5	2	5	0,211	6
1,5	3	3	0,269	10
1,5	4	1	-	17
1,5	5	1	-	17
2	2	2	0,332	4
2	3	2	0,343	3
2	4	2	0,346	3
2	5	1	-	7

Pemilihan parameter *DBSCAN* dengan epsilon sebesar 2 dan min_sample sebesar 4 dianggap optimal karena menghasilkan 2 kluster dengan *Silhouette Score* sebesar 0,346, yang merupakan skor tertinggi dibandingkan kombinasi parameter lainnya. Nilai *Silhouette Score* yang positif menunjukkan bahwa kluster yang dihasilkan memiliki kualitas yang baik, dengan data dalam satu kluster lebih dekat satu sama lain dibandingkan dengan data di kluster lain. Selain itu, jumlah *noise* yang dihasilkan hanya sebanyak 3 titik, yang merupakan salah satu jumlah *noise* terendah, menandakan bahwa sebagian besar data berhasil dimasukkan ke dalam kluster.

4.7 DBSCAN Klaster

Pengelompokan perusahaan LQ45 tahun 2023 berdasarkan kinerja keuangan perusahaan menggunakan *DBSCAN* dilakukan dengan menggunakan jumlah 2 klaster yaitu tinggi dan rendah. Hasil pengelompokan perusahaan LQ45 tahun 2023 dapat ditunjukkan dalam Tabel 6 sebagai berikut.

Tabel 5. Klaster *DBSCAN* perusahaan

Klaster	Jumlah Perusahaan	Mean					
		ROE	DAR	DER	NPM	UP	VOL
Klaster 1	37	16,63	0,42	0,98	16,67	17,75	1100
Klaster 2	4	16,59	0,83	5,63	41,02	21,13	1956
Noise	3	-	-	-	-	-	-

Tabel 6 menunjukkan bahwa klaster 1 terdiri dari 37 perusahaan dengan profitabilitas yang cukup baik (ROE 16,63 dan NPM 16,67) serta tingkat utang yang relatif rendah (DAR 0,42 dan DER 0,98), sehingga menunjukkan kondisi keuangan yang cukup stabil. Klaster 2 berisi 4 perusahaan dengan profitabilitas sangat tinggi, terutama NPM sebesar 41,02, serta tingkat utang yang rendah (DAR 0,83 dan DER 5,63). Selain itu, klaster ini memiliki ukuran perusahaan dan volume perdagangan yang lebih tinggi dibanding klaster 1. *Noise* terdiri dari 3 perusahaan yang tidak termasuk dalam klaster mana pun karena memiliki karakteristik yang berbeda dari kelompok lainnya. Secara umum, Klaster 1 merepresentasikan perusahaan dengan kinerja stabil dan risiko relatif rendah, sedangkan Klaster 2 menunjukkan perusahaan yang lebih besar dan lebih menguntungkan. Sementara itu, kelompok *noise* merupakan perusahaan dengan pola yang tidak seragam sehingga tidak tergolong ke dalam klaster tertentu.

4.8 Perbandingan Klaster *K-Means* dan *DBSCAN*

Tabel 6. Perbandingan *K-Means* dan *DBSCAN*

Klaster	<i>Silhouette Score</i>	<i>Index Dunn</i>
<i>K-Means</i>	0,327	0,208
<i>DBSCAN</i>	0,346	0,502

Tabel 8 menjelaskan bahwa perbandingan hasil klaster terbaik antara algoritma *K-Means* dengan *DBSCAN* yang dilihat menggunakan *silhouette score* dan *index dunn*. *DBSCAN* lebih unggul dibandingkan *K-Means* dalam hal *Silhouette Score*, meskipun perbedaannya tidak terlalu signifikan. Ini mengindikasikan bahwa *DBSCAN* mampu menangkap struktur data dengan lebih baik, terutama jika data memiliki outlier atau bentuk klaster yang tidak teratur. *DBSCAN* juga menunjukkan hasil yang jauh lebih baik dalam hal *Index Dunn*, yang berarti klaster *DBSCAN* lebih terpisah satu sama lain, karena *DBSCAN* tidak memaksakan pembentukan klaster dari data yang tidak memenuhi kriteria. *DBSCAN* memiliki *Silhouette Score* dan *Index Dunn* yang lebih tinggi, menunjukkan bahwa *DBSCAN* lebih mampu menangkap struktur data yang lebih kompleks dan memiliki klaster yang lebih terpisah

5. KESIMPULAN

Jumlah klaster optimal yang terbentuk dalam pengelompokan perusahaan LQ45 menggunakan metode *K-Means* sebanyak 3 klaster, dengan anggota klaster 1 sebanyak 8 perusahaan, klaster 2 sebanyak 29 perusahaan, dan klaster 3 sebanyak 7 perusahaan, sedangkan pengelompokan menggunakan metode *DBSCAN* menghasilkan jumlah klaster optimal sebanyak 2 dengan anggota klaster 1 sebanyak 37 perusahaan dan klaster 2 sebanyak 4 perusahaan, serta terdapat *noise* sebanyak 3 perusahaan. Metode *DBSCAN* lebih baik dalam mengelompokkan perusahaan karena memiliki nilai *Silhouette Score* dan *Index Dunn* yang lebih tinggi dibanding dengan metode *K-Means*.

DAFTAR PUSTAKA

- [1] A. Hidayah, "Investor Pasar Modal Tembus 10 Juta, Jebolan SMU Mendominasi," CNBC Indonesia. Accessed: Jul. 23, 2026. [Online]. Available: <https://www.cnbcindonesia.com/market/20221121103752-17-389764/investor-pasar-modal-tembus-10-juta-jebolan-smu-mendominasi>
- [2] S. H. Winarno, "Penilaian Kinerja Keuangan Perusahaan Melalui Analisis Profitabilitas," *Jurnal Moneter*, pp. 106–112, 2017.
- [3] C. Seniati and H. Firmansyah, "Indonesian Stock Movement Using K-Means Clustering Method," *Jurnal Sistem Informasi dan Teknologi Digital (SITEDI)*, vol. 2, no. 4, pp. 132–135, Dec. 2025, doi: 10.70888/sitedi.v2i4.70.
- [4] R. A. Hasibuan, S. W. Rizki, and H. Perdana, "Perbandingan Metode Fuzzy Clustering Means dan Single Linkage pada Pengelompokan Saham LQ45," *Buletin Ilmiah Math. Stat. dan Terapannya (BIMASTER)*, vol. 10, no. 3, pp. 361–368, 2021.
- [5] M. Shodikin and N. Taniasari, "Analisis Komparasi Algoritma DBSCAN dan K-Means dalam Pemetaan Segmentasi Pasien Rawat Inap Menggunakan Model RFMT," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 4, pp. 1231–1241, Oct. 2025, doi: 10.57152/malcom.v5i4.2218.
- [6] S. Sulistina, P. E. P. Utomo, and B. F. Hutabarat, "Klasterisasi Wilayah Rawan Kriminalitas di Kota Jambi (2022-2024) Menggunakan Algoritma K-Means," *Jurnal INSTEK (Informatika Sains dan Teknologi)*, vol. 10, no. 2, pp. 546–560, Oct. 2025, doi: 10.24252/instek.v10i2.62051.
- [7] N. P. E. Merliana, Ernawati, and Alb. J. Santoso, "Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering," in *Proceeding SENDI_U*, 2015, pp. 978–990.
- [8] T. M. Kodinariya and P. R. Makwana, "Review on Determining Number of Cluster in K-Means Clustering," *International Journal of Advance Research in Computer Science and Management Studies*, vol. 1, no. 6, pp. 90–95, 2013.
- [9] R. Virgatama, A. Suprayogi, and H. S. Firdaus, "Identifikasi Pengaruh Sistem Keamanan Lingkungan Terhadap Tingkat Kejahatan Pencurian di Kota Surakarta dengan Metode Sistem Informasi Geografis," *Jurnal Geodesi Undip*, vol. 8, no. 1, pp. 398–407, 2019.
- [10] F. Setiawan, F. Azmi, and C. Setianingsih, "Clustering pada Data Sentimen Transportasi Online Menggunakan Algoritma DBSCAN," *e-Proceeding of Engineering*, vol. 8, no. 5, pp. 6089–6096, 2021.
- [11] U. Y. Purwanto, B. Barus, and H. A. Adrianto, *Spatial Hotspots Clustering of Forest and Land Fires Using DBSCAN and ST-DBSCAN*. Bogor: Repository IPB, 2012.
- [12] I. Djajadi, *IDX Stock Index Handbook v1.2*. Jakarta: Indonesia Stock Exchange, 2021.

Biodata Penulis

Iqbal Ali Mansyah Wardana, lahir di Surabaya pada 18 Desember 2002. Saat ini menjadi mahasiswa aktif D4 Statistika Bisnis di Institut Teknologi Sepuluh Nopember Surabaya. Bidang minat *machine learning* dan *finance*. Selain itu penulis juga menekuni di bidang optimasi portofolio.

Finda Nur Lailiyah, lahir di Surabaya pada tanggal 17 Februari 2003. Penulis merupakan mahasiswa aktif D4 Statistika Bisnis di Institut Teknologi Sepuluh Nopember Surabaya dan sedang menggeluti di bidang keuangan dan bidang optimasi portofolio.

Noviyanti Santoso, penulis mempunyai latar pendidikan bidang Statistika dan menekuni bidang Komputasi. Saat ini penulis menjadi staf pengajar di salah satu perguruan tinggi negeri di Surabaya. Selain itu penulis juga aktif dalam berbagai penelitian meliputi topik *machine learning*, *text mining*, dan *big data analysis*.