Volume 5 Issue 1 (2025) Pages 57-68 P-ISSN:2798-2696 E-ISSN:2798-0057
JOSTECH
Journal of Science and Technology
Website: https://ejournal.uinib.ac.id/jurnal/index.php/jostech Email: jostech@uinib.ac.id

Algoritma DBSCAN dan *K-Means Clustering* dalam Pemilihan Saham Berdasarkan *Feature Engineering*

Ratna Tri Aulia[1]✉, Arisman Adnan[2], Ihda Hasbiyati[3]

[1,2,3]*Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Riau, Pekanbaru, Indonesia*
email: ratna.tri6886@grad.unri.ac.id[1] arisman.adnan@lecturer.unri.ac.id[2]
ihda.hasbiyati@lecturer.unri.ac.id[3]

Received 16 November 2024, Accepted 21 Maret 2025, Published 31 Maret 2025

Abstrak

Saham adalah salah satu aset investasi di pasar modal yang memiliki daya tarik tertentu bagi para investor. Berbagai indeks mencerminkan karakteristik saham yang terdaftar di Bursa Efek Indonesia, salah satunya adalah LQ45. Indeks saham ini menunjukkan bahwa saham komposit ini merupakan saham berkapitalisasi besar dengan likuiditas tinggi dan fundamental yang baik. Memilih aset untuk alokasi portofolio adalah faktor penting dalam berinvestasi, dengan tujuan untuk memaksimalkan imbal hasil dan meminimalkan risiko. Penelitian ini bertujuan untuk mengklasifikasikan saham LQ45 berdasarkan karakteristik tertentu seperti volatilitas, likuiditas, dan kapitalisasi pasar, yang diekstraksi melalui *feature engineering*. Fitur-fitur ini digunakan untuk mengekstrak informasi dari data dan dilakukan selama tahap pra-pemrosesan data. Teknik *clustering* yang digunakan adalah algoritma *K-Means* dengan bantuan algoritma DBSCAN untuk mendeteksi *outlier* dan metode *elbow* untuk menentukan jumlah *cluster* yang optimal. Proses analisis *clustering* dilakukan tanpa data *outlier* dan menggunakan *software* RStudio. Berdasarkan penelitian ini, diperoleh *Silhouette Coefficient* (SC) sebesar 0,7004 atau 70,04%, yang menunjukkan struktur pengelompokan yang kuat karena nilainya berada dalam interval 0,7 hingga 1. Selain itu, terbentuk total 5 *cluster*.

Kata Kunci: DBSCAN; *feature engineering*; *K-Means clustering*; *Silhouette coefficient*; saham.

Abstract

Stocks is one of the investment assets in the capital market that hold a particular appeal for investors. Various indexes represent the characteristics of stocks listed on the Indonesia Stock Exchange, one of which is LQ45. This stock indexes that these composite stocks are large-cap stocks with high liquidity and good fundamentals. Choosing assets for portfolio allocation is a important factor in investing, with the goal being to maximize returns and minimize risks. This study aims to classify LQ45 stocks based on specific characteristics such as volatility, liquidity, and market capitalization, extracted through feature engineering. This future are used to extract information from the data and is conducted during the data preprocessing stage. The clustering method used is the K-Means clustering algorithm with the assistance of the DBSCAN algorithm to detect outliers and the elbow method to determine the optimal number of clusters. The clustering analysis process is carried out without outlier data and

using Rstudio software. Based on this study, a Silhouette Score (SC) of 0.7004 or 70.04% was obtained, indicating a strong clustering structure as the value falls within the interval of 0.7 to 1. Additionally, a total of 5 clusters were formed.

Keywords: DBSCAN; feature engineering; K-Means clustering; Silhouette coefficient; stock.
✉ Corresponding author

PENDAHULUAN

Salah satu aset finansial yang paling populer di pasar modal yang bisa digunakan dalam jangka panjang adalah saham [1]. Kelebihan saham dibandingkan dengan instrumen keuangan lainnya terletak pada tingkat likuiditasnya yang tinggi, memungkinkan para pemegang saham untuk dengan mudah melakukan transaksi jual beli di pasar saham [2], [3]. Investasi saham dapat dikategorikan investasi dengan risiko yang cukup tinggi. Dalam konteks penerapan teori keuangan, risiko diartikan sebagai tingkat volatilitas dari harga saham [3], [4]. Investor umumnya menerapkan strategi diversifikasi untuk mengurangi risiko melalui pembentukan portofolio investasi. Diversifikasi ini bertujuan untuk menyebarkan investasi ke berbagai saham, sehingga risiko yang dihadapi dapat terdiversifikasi dan diminimalkan [5], [6].

Permasalahan yang sering terjadi dalam melakukan portofolio investasi saham adalah bagaimana menetapkan kombinasi saham agar diversifikasi portofolio dapat berjalan dengan baik [7]. Pilihan kombinasi saham yang dibuat oleh investor berdampak pada kinerja portofolio investasi [8]. Pembentukan portofolio dengan strategi pemilihan kombinasi saham telah banyak diterapkan menggunakan teknik clustering untuk mengelompokkan saham yang memiliki karakteristik tertentu. Salah satu metode yang umum digunakan untuk mengelompokkan objek berdasarkan atributnya adalah melalui algoritma *K-Means clustering* [9], [10]. Algoritma ini merupakan algoritma *clustering* non-hierarki dimana proses pengelompokannya dimulai dengan menentukan jumlah *cluster* yang akan dibentuk, kemudian mengelompokkan data ke dalam sejumlah *cluster* sebanyak mungkin berdasarkan kedekatan objek dengan pusat *cluster* [11]. Tujuan dari proses *clustering* ini adalah untuk mengurangi variasi di dalam suatu kelompok dan secara simultan memaksimalkan variasi antara kelompok-kelompok tersebut [12].

Terdapat beberapa penelitian sebelumnya yang membahas tentang teknik *clustering* yang menggunakan *K-means clustering* pada portofolio investasi. Gubu, *et al*. [13] melakukan perbandingan klasifikasi data harian harga saham indeks LQ45 menggunakan metode *K-Means* dan *K-Medoids clustering*, yang selanjutnya diikuti dengan penentuan portofolio optimal menggunakan model Markowitz. Hasil analisis menunjukkan bahwa *K-Means clustering* memiliki kinerja yang baik dalam mengelompokkan saham dengan toleransi risiko < 15, dan menghasilkan portofolio optimal dengan performa yang memuaskan. Sementara itu, Kumari, *et al*. [14] mengungkapkan bahwa *K-Means clustering* dapat meningkatkan efisiensi waktu dalam menyeleksi saham yang sejenis untuk pembentukan portofolio yang optimal. Keseluruhan penelitian tersebut membuktikan bahwa teknik *clustering*, seperti *K-Means*

clustering dapat menjadi pendekatan yang bermanfaat dalam membentuk portofolio saham dengan mempertimbangkan karakteristik dan performa yang diinginkan. Namun, penggunaan *K-Means clustering* memiliki kekurangan yaitu sensitif terhadap *outlier*, sensitif terhadap skala data, dan tidak efektif untuk berbagai *cluster* [3],[15]. Penelitian Iriany, *et al*. [16] mengatakan bahwa metode elbow dan algoritma *DBSCAN clustering* dapat digunakan untuk mengatasi kelemahan *K-Means clustering* sehingga efektif dalam mendeteksi *outlier* dan hasil interpretasi *clustering* lebih baik.

Penelitian sebelumnya umumnya hanya menggunakan salah satu metode *clustering*, yaitu *K-Means* atau *DBSCAN* dalam pengelompokan saham. Selain itu, penelitian-penelitian tersebut lebih banyak berfokus pada analisis berbasis harga saham tanpa mempertimbangkan faktor lain yang relevan dalam pemilihan saham. Dalam proses investasi, investor perlu memperhatikan berbagai aspek fundamental dan teknikal dari perusahaan penerbit saham untuk membuat keputusan yang lebih akurat. Oleh karena itu, penelitian ini menghadirkan pendekatan baru dengan mengombinasikan metode *DBSCAN* dan *K-Means clustering* guna mengoptimalkan proses pengelompokan saham, serta menambahkan *feature engineering* dalam tahap praproses data. *Feature engineering* ini dilakukan untuk mengekstrak informasi penting dari data saham, yaitu volatilitas, likuiditas, dan kapitalisasi pasar, yang dianggap sebagai faktor utama dalam penentuan saham yang layak diinvestasikan.

Dengan pendekatan ini, penelitian ini bertujuan untuk mengatasi keterbatasan dari penelitian sebelumnya dengan menghasilkan pengelompokan saham yang lebih akurat dan representatif. Kombinasi algoritma *DBSCAN* dan *K-Means clustering* diharapkan dapat mengoptimalkan keunggulan masing-masing metode dalam membentuk *cluster* yang lebih stabil, sementara penambahan fitur keuangan dapat meningkatkan relevansi hasil *clustering* dalam konteks pemilihan saham. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan strategi pemilihan saham yang lebih efektif dan berbasis data yang lebih komprehensif.

METODOLOGI

1. Sumber Data

Penelitian ini menggunakan data sekunder yaitu data saham yang tergabung dalam indeks saham LQ45 yang terdiri dari 45 saham. Data yang digunakan diantaranya berupa data profil perusahaan dan data pergerakan harga saham yang diambil berdasarkan data harian penutupan harga saham dimulai dari April 2023 – Maret 2024. Data profil perusahaan diperoleh dengan cara melakukan *scraping* dari *website* resmi Bursa Efek Indonesia (BEI). Sedangkan data penutupan harga saham diperoleh melalui *website* https://finance.yahoo.com/.

2. Metode Penelitian

Metode yang digunakan dalam penelitian ini yaitu *K-Means clustering* yang digunakan untuk membentuk *cluster* dari 45 saham. Hal ini didukung dengan penggunaan algoritma

DBSCAN clustering untuk mengatasi outlier pada proses *clustering*, dan juga dilengkapi dengan metode elbow untuk menentukan jumlah *cluster* yang optimal. Proses *clustering* dilakukan dengan menggunakan *software* R. Adapun langkah analisis yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Pengumpulan data
   Data yang digunakan berupa data profil perusahaan masing-masing saham yang disimpan dalam Ms. Excell atau dalam format (.xlsx), serta data harian harga penutupan saham diperoleh secara online dengan menggunakan *function 'tq_get()'* dari *packages tidyquant* yang terdapat di *software* R.
2. Penambahan *feature engineering*
   Feature engineering adalah proses menggunakan pengetahuan dari para ahli di bidang tertentu untuk memperoleh wawasan dari data, dengan tujuan meningkatkan hasil pembelajaran mesin [17]. Tiga karakteristik akan diambil dari data tersebut, yaitu volatilitas, likuiditas dan kapitalisasi saham. Adapun rumus untuk menghitung volatilitas dan likuiditas berturut-turut sebagai berikut [18]:

$$Volatility = \sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(R_i - \bar{R})^2}, \quad (1)$$

$$Liquidity = \frac{1}{n}\sum_{i=1}^{n}\frac{|R_i|}{Volume_i}, \quad (2)$$

   dengan, R_i adalah *return* saham ke-*i, n* adalah jumlah pengamatan, $\bar{R}$ adalah rata-rata *return* saham, dan $Volume_i$ adalah volume perdagangan pada waktu ke-*i.*
3. Penggabungan data
   Setelah memutuskan fitur mana yang akan digunakan dalam proses *clustering*, langkah selanjutnya adalah menggabungkan semua fitur tersebut ke dalam *data frame*.
4. Transformasi data
   Clustering belum dapat dilakukan karena data saham yang diperoleh dari fungsi '*td_qet()*' masih dalam format dengan lebih dari satu baris per saham. Oleh karena itu, data perlu ditransformasi sehingga setiap saham diwakili oleh satu baris. Data kemudian diubah menjadi format *wide data frame*.
5. Normalisasi data
   Data yang akan digunakan dalam proses *clustering* perlu disesuaikan skalanya. Hal ini penting dilakukan karena metode *K-Means clustering* bergantung pada perhitungan jarak antar data (jarak Euclidean), sehingga rentang data harus seragam. Penskalaan dilakukan menggunakan metode *z-score*, yang hanya mengubah skala data tanpa mengubah distribusi data aslinya [19].
6. Deteksi *outlier* menggunakan algoritma *DBSCAN clustering*
   Algortima *DBSCAN clustering* melibatkan dua parameter input yang digunakan untuk menilai kepadatan setiap sampel data, sehingga dapat mengelompkkan dan mengidentifikasi *outlier* dengan efektif. Parameter tersebut diantaranya radius lingkungan (*eps*) dan ambang batas (*MinPts*) [20]. Algoritma *DBSCAN* melibatkan langkah-langkah berikut [8].

(1) Inisialisasi parameter *MinPts* dan parameter *eps*.
(2) Tentukan nilai eps dengan menghitung jarak euclidean antara titik *p* dan semua titik yang memiliki kepadatan atau densitas yang sesuai. Jarak euclidean dapat dihitung menggunakan persamaan berikut [21].

$$d = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad ; \quad i = 1,2,\dots,n, \tag{3}$$

dimana, x_i adalah data sampel ke-i dan y_i adalah titik pusat dari *k cluster*.
(3) Input nilai MinPts dan eps terbaik pada model *DBSCAN*.
(4) Simpan label yang telah dibentuk oleh *DBSCAN*.
(5) Tentukan titik mana yang dapat membentuk titik inti.
(6) Hitung jumlah *cluster* dan nilai silhouette coefficient dengan menggunakan persamaan (5).

7. Menentukan *k* optimal menggunakan metode elbow
Metode elbow adalah sebuah pendekatan dalam analisis *cluster* yang digunakan untuk menentukan jumlah *cluster* yang optimal, dengan fokus pada interpretasi dan pengujian konsistensi kinerja [22]. Pendekatan ini memeriksa *Sum of Squared Errors* (SSE) sebagai fungsi dari jumlah *cluster*. Nilai SSE dapat dicari dengan menggunakan persamaan berikut [23].

$$SSE = \sum_{k=1}^{K}\sum_{x \in C_k}\|x - C_k\|^2 \quad , k = 1,2,\dots,K \tag{4}$$

dengan, C_k adalah pusat *cluster* ke-k dan x adalah titik data dalam *cluster* C_k.

8. Melakukan *clustering* menggunakan *K-Means clustering*
K-Means clustering adalah metode pengelompokan non-hierarkis yang bertujuan untuk mengelompokkan data ke dalam *cluster* berdasarkan jumlah cluster yang telah ditentukan sebelumnya dengan menentukan posisi awal *centroid* [24]. Algoritma *K-Means clustering* menggunakan pendekatan iteratif untuk membentuk *cluster* dalam basis data. *K-Means clustering* akan terus memperbarui titik pusat pada setiap iterasi [25]. Setelah proses iterasi *K-Means clustering* selesai, setiap objek dalam dataset diklasifikasikan ke dalam salah satu *cluster*. Proses pengelompokan *K-Means clustering* dapat diuraikan melalui langkah-langkah berikut ini [13].
(1) Tentukan jumlah *k cluster* yang akan dibentuk.
(2) Pilih data *k* sebagai pusat awal (*centeroid*) dari tiap *cluster*.
(3) Kelompokkan data menjadi *k cluster* berdasarkan titik pusat yang ditentukan.
(4) Update pusat *cluster* dan ulangi langkah 3 hingga nilai titik centeroid tidak berubah lagi.
(5) Hitung nilai *silhouette coefficient* menggunakan persamaan (5).

9. Uji validasi algoritma *K-Means clustering*

Validasi yang digunakan dalam penelitian ini didasarkan pada indeks *silhouette* tertinggi dari hasil *cluster* yang telah ditentukan. Adapun rumus yang digunakan untuk menghitung nilai *silhouette coefficient* adalah.

$$S[i] = \frac{b[i]-a[i]}{\max\ (a[i],b[i])} \tag{5}$$

dengan, $b[i]$ adalah rata-rata ketakmiripan terendah pada *cluster* ke-*i* dengan *cluster* yang lain, dan $a[i]$ adalah rata-rata ketakmiripan dari semua data pada *cluster* ke-*i*. Adapun kriteria untuk nilai *silhouette coefficient* dilampirkan pada Tabel 1 berikut [21].

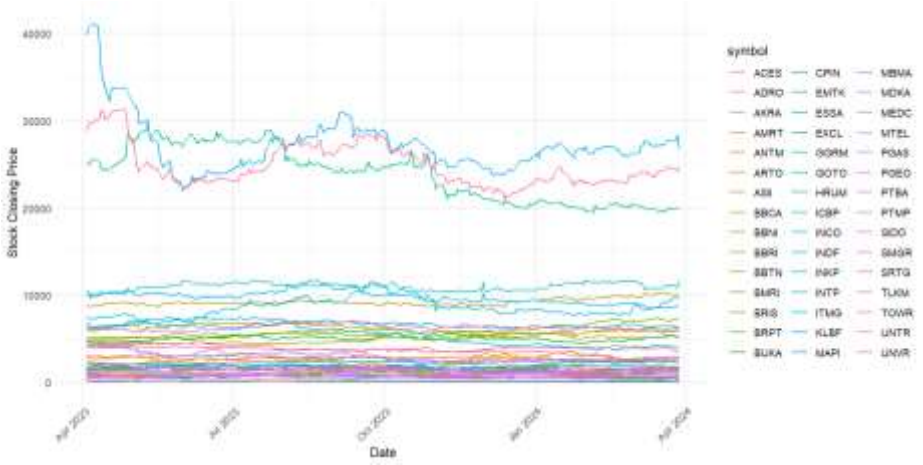
Tabel 1. Nilai *Silhouette Coefficient*

Nilai *Silhouette Coefficient* (SC)	Kriteria
$0{,}7 < SC \leq 1$	Struktur kuat
$0{,}5 < SC \leq 0{,}7$	Struktur medium
$0{,}25 < SC \leq 0{,}5$	Struktur lemah
$SC \leq 0{,}25$	Tidak ada struktur

HASIL DAN PEMBAHASAN

1. Analisis Deskriptif

Data yang digunakan pada penelitian ini adalah data pergerakan harga penutupan saham dari 45 perusahaan selama periode dari April 2023 hingga Maret 2024 yang disajikan dalam bentuk grafik pada Gambar 1.



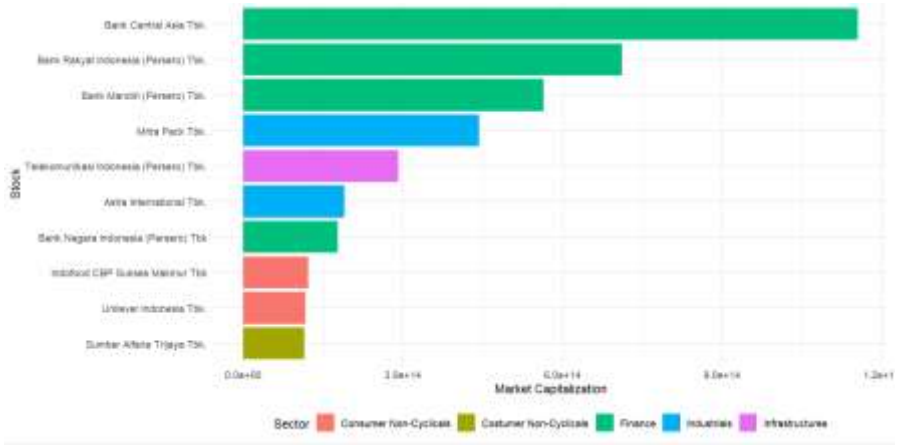
Gambar 1. Grafik harga penutupan saham LQ45 2023-2024

Dari pengamatan umum pada Gambar 1, terlihat bahwa ada beberapa saham dengan harga penutupan yang jauh lebih tinggi dibandingkan saham lainnya. Saham-saham ini

ditandai dengan garis berwarna hijau, biru, dan merah muda, menunjukkan performa yang lebih tinggi secara keseluruhan. Beberapa saham menunjukkan fluktuasi harga yang signifikan, terutama pada awal periode sekitar April 2023, dengan harga yang naik turun secara drastis. Di sisi lain, mayoritas saham lainnya cenderung stabil dengan harga penutupan yang berada pada kisaran yang lebih rendah dan tanpa fluktuasi besar.

2. Feature Engineering

Sebelum memulai analisis *clustering*, langkah awal yang penting adalah melakukan rekayasa fitur (*feature engineering*). Proses ini membantu dalam mengenali saham-saham mana yang memiliki dampak terbesar dan karakteristik yang berbeda, yang nantinya akan menjadi dasar dalam pengelompokan saham berdasarkan berbagai metrik kinerja.



Gambar 2. 10 saham teratas berdasarkan kapitalisasi pasar

Gambar 2, menampilkan 10 saham teratas berdasarkan kapitalisasi pasar, memberikan gambaran tentang dominasi sektor-sektor tertentu di pasar saham Indonesia. Selanjutnya Gambar 3, menunjukkan *data frame* yang digunakan untuk *clustering*, mencakup berbagai variabel seperti likuiditas (*medvol*), volatilitas (*sdclose*), dan kapitalisasi pasar (*marketcap*) untuk beberapa saham. Data ini akan menjadi input utama dalam analisis *clustering* untuk mengidentifikasi kelompok saham dengan karakteristik yang serupa.

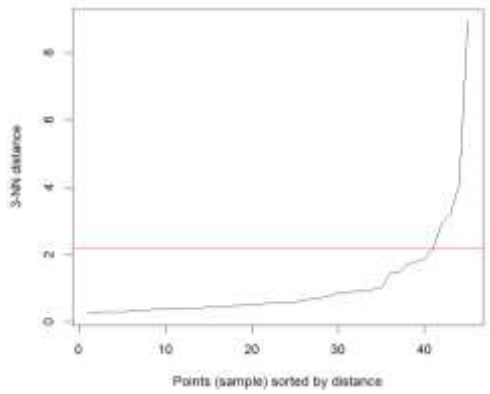
	...1	medvol_2023	medvol_2024	sdclose_2023	sdclose_2024	marketcap
1	ACES	0.37093003	0.21941020	0.03837941	0.023773141	1.37000e+13
2	ADRO	0.13542597	0.11079642	0.17759202	0.015049593	8.89400e+13
3	AKRA	0.15770663	0.12800749	0.03702446	0.013657999	3.09900e+13
4	AMRT	0.04549579	0.05680225	0.05536768	0.014225896	1.14610e+14
5	ANTM	0.14287560	0.17315866	0.02630268	0.020579450	3.76100e+13
6	ARTO	0.13828278	0.10181563	0.06175074	0.032206197	3.25600e+13
7	ASII	0.08973891	0.15730018	0.09489107	0.014899375	1.90270e+14
8	BBCA	0.05157376	0.05540063	0.05412846	0.009991848	1.15570e+15
9	BBNI	0.13010767	0.12345596	0.04201818	0.015185871	1.77670e+14
10	BBRI	0.07232820	0.07815375	0.01932814	0.015281973	7.11734e+14
11	BBTN	0.11010717	0.24844358	0.06136967	0.026225318	1.73320e+13
12	BMRI	0.08036331	0.09829071	0.17958308	0.014326525	5.64666e+14
13	BRIS	0.04190144	0.09222172	0.06071487	0.026439349	1.07481e+14
14	BRPT	0.07897071	0.10360451	0.06934026	0.035009735	1.10871e+14
15	BUKA	0.12751094	0.10717214	0.06113036	0.023164241	1.34040e+13
16	CPIN	0.03609282	0.04237834	2.30840474	0.016315004	8.56800e+13
17	EMTK	0.04040560	0.03410340	0.07024268	0.026136059	2.69940e+13
18	ESSA	0.17656088	0.17409179	0.09845253	0.031523735	1.40400e+13
19	EXCL	0.12687807	0.14425677	0.15030969	0.024346681	3.22880e+13
20	GGRM	0.07228360	0.03790367	0.76247025	0.012247625	3.67620e+13
21	GOTO	0.26403874	0.19279760	0.09068175	0.040859558	7.06570e+13
22	HRUM	0.10442666	0.08204518	1.54273304	0.023961775	1.77110e+13
23	ICBP	0.04158325	0.03919256	0.48046870	0.018406306	1.22158e+14
24	INCO	0.06984665	0.13331973	0.03632891	0.028417175	4.86080e+13

Showing 1 to 24 of 45 entries, 6 total columns

Gambar 3. Data frame untuk *clustering*

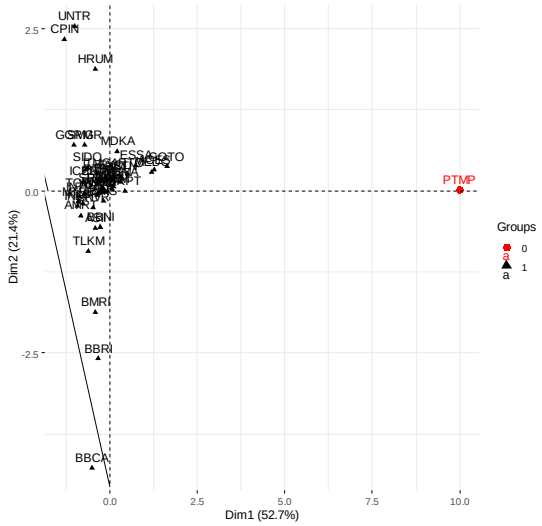
3. Implementasi Algoritma *DBSCAN Clustering*

Pada tahap ini, algoritma *DBSCAN clustering* digunakan untuk mendeteksi *outlier*. Sebelum melakukan identifikasi *outlier* dari data saham, penting untuk terlebih dahulu mengidentifikasi parameter optimal untuk *outlier detection* yang akan digunakan. Gambar 4 menunjukkan langkah awal dalam proses ini dengan melakukan identifikasi *outlier* menggunakan metode kepadatan, selanjutnya hasil *outlier* ditunjukkan pada Gambar 5.



Gambar 4. Analisis *outlier* menggunakan metode density

Gambar 4, menunjukkan analisis *outlier* menggunakan algoritma *DBSCAN*, di mana titik-titik dengan jarak 3-NN (tetangga terdekat ke-3) yang besar diidentifikasi sebagai *outlier*, dengan garis merah menandai ambang batasnya. Hasil parameter optimal untuk algoritma *DBSCAN*, yaitu eps = 2,2 dan minPts = 3, yang digunakan untuk menentukan tetangga dan membentuk klaster dalam deteksi *outlier*. Berdasarkan hasil analisis diperoleh skor silhouette sebesar 0,7957 atau 79.57% menunjukkan kriteria dengan struktur kuat.

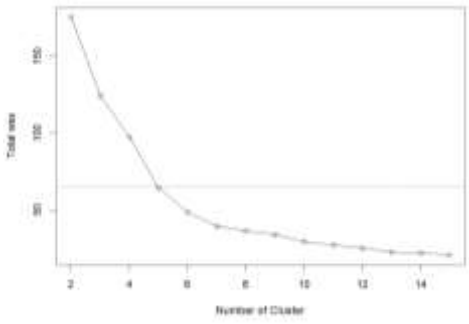


Gambar 5. *Principal component analysis (PCA)*

Gambar 5, menunjukkan hasil *PCA* yang mengidentifikasi dua kelompok ditandai dengan simbol-simbol saham yang dikelompokkan berdasarkan kesamaannya. Sebagian besar simbol saham berkumpul di pusat, menunjukkan bahwa mereka memiliki karakteristik yang mirip dalam dimensi yang diproyeksikan oleh komponen utama. Namun, terdapat simbol saham seperti "PTMP" yang terpisah dari kelompok utama yang menunjukkan bahwa saham tersebut memiliki karakteristik yang berbeda secara signifikan yang dapat dikatakan sebagai data *outlier*.

4. Implementasi Metode Elbow

Sebelum menerapkan algoritma *clustering* untuk mengelompokkan data saham, langkah penting yang harus dilakukan adalah menentukan jumlah *cluster* optimal dengan menggunakan metode elbow. Hasil dari penerapan metode dapat dilihat pada Gambar 6.

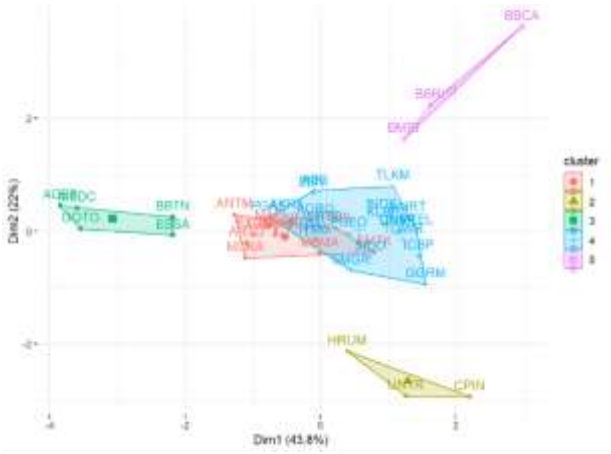


Gambar 6. Metode Elbow

Gambar 6, menunjukkan hasil dari metode elbow, yang digunakan untuk menentukan jumlah *cluster* optimal dalam analisis clustering. Pada grafik tersebut menunjukkan penurunan tajam pada WSS saat jumlah klaster bertambah dari 2 hingga 5, setelah itu penurunan melambat dan menjadi kurang signifikan. Titik "elbow" yang terlihat pada $k = 5$ menunjukkan jumlah klaster optimal. Artinya pada $k = 5$, penurunan WSS menjadi lebih kecil dengan penambahan *cluster* lebih lanjut yang menandakan bahwa 5 *cluster* memberikan pengelompokan yang cukup baik dengan variasi yang minimal dalam setiap klaster.

5. Implementasi Algoritma K-Means Clustering

Setelah menentukan jumlah klaster optimal menggunakan metode elbow, langkah berikutnya adalah menerapkan algoritma clustering untuk mengelompokkan data saham dengan menggunakan algoritma K-Means clustering. Jumlah klaster yang telah ditentukan akan digunakan pada tahap ini sehingga dapat memvisualisasikan hasil clustering ini untuk memahami bagaimana saham-saham tersebut dikelompokkan. Hasil cluster dapat disajikan pada Gambar 7.



Gambar 7. *K-Means cluster*

Gambar 7, menunjukkan hasil implementasi algoritma *K-Means clustering* menggunakan jumlah *cluster* didasarkan pada hasil metode elbow yaitu berjumlah 5 *cluster*. Dari grafik tersebut, dapat dilihat bahwa saham-saham terkelompok dengan baik ke dalam beberapa *cluster* yang berbeda, dengan masing-masing *cluster* menunjukkan kesamaan karakteristik di antara saham-saham yang termasuk di dalamnya. *Cluster* 1 terdiri dari 12 anggota, termasuk saham-saham seperti ANTM, ARTO, EMTK, dan lainnya. *cluster* ini memiliki likuiditas yang menengah pada tahun 2023 dan sedikit meningkat pada tahun 2024, dengan volatilitas yang lebih rendah pada tahun 2023 dan meningkat tajam pada tahun 2024, serta kapitalisasi pasar yang sedikit negatif.

Cluster 2, yang terdiri dari 3 anggota seperti CPIN, HRUM, dan UNTR, ditandai dengan penurunan likuiditas pada kedua tahun dan kenaikan volatilitas yang signifikan pada tahun 2023, serta penurunan kecil dalam kapitalisasi pasar. *Cluster* 3, dengan 5 anggota, menunjukkan likuiditas yang sangat tinggi pada kedua tahun dengan volatilitas menurun pada tahun 2023 dan meningkat pada tahun 2024, serta kapitalisasi pasar yang negatif. *Cluster* 4 adalah yang terbesar, terdiri dari 21 anggota, dan memiliki likuiditas yang menurun pada kedua tahun dengan sedikit perubahan dalam volatilitas dan kapitalisasi pasar yang negatif. Terakhir, *Cluster* 5, dengan 3 anggota, menunjukkan penurunan likuiditas dan volatilitas pada kedua tahun dengan kapitalisasi pasar yang sangat tinggi.

SIMPULAN

Secara keseluruhan, hasil *clustering* ini mencerminkan karakteristik yang berbeda dari saham-saham yang dikelompokkan berdasarkan likuiditas, volatilitas, dan kapitalisasi pasar. Selain itu, hasil ini juga menunjukkan kualitas *clustering* dengan kriteria kuat yang ditunjukkan oleh skor silhouette sebesar 0,7004 atau 70,04%, sehingga menandakan bahwa *cluster-cluster* yang terbentuk memiliki pemisahan yang jelas dan kohesivitas internal yang tinggi. Untuk penelitian lebih lanjut, disarankan untuk mengamati pergeseran *cluster* seiring waktu dan mengimplementasikan pembentukan portofolio berdasarkan *cluster* yang telah ditemukan. Penelitian ini juga dapat diperluas dengan menambahkan fitur lain yang signifikan dalam menentukan portofolio saham.

DAFTAR PUSTAKA

[1] E. Endri, F. Fauzi, and M. S. Effendi, "Integration of the Indonesian Stock Market with Eight Major Trading Partners' Stock Markets," *Economies*, vol. 12, no. 12, pp. 1–22, 2024.

[2] T. Febiyola, R. S. Utari, B. T. Panggabean, and R. Agustina, "Analisis Surat Berharga Sebagai Alat Investasi," *Publ. Ilmu Huk.*, vol. 2, no. 3, pp. 75–86, 2024.

[3] A. F. Ridwan, H. Napitupulu, and Sukono, "Decision-making in formation of mean-VaR optimal portfolio by selecting stocks using K-means and average linkage clustering," *Decis. Sci. Lett.*, vol. 11, pp. 431–442, 2022.

[4] T. Andika, I. Fahmi, and T. Andati, "the Macroeconomic Surprise Effects on Lq45 Stock Return Volatility," *J. Apl. Manaj.*, vol. 17, no. 2, pp. 235–243, 2019.

[5] S. Amaroh and C. Nasichah, "Risk-Return Analysis on Optimum Portfolio Selection of Islamic Stocks," *Equilib. J. Ekon. Syariah*, vol. 9, no. 1, pp. 65–82, 2021.

[6] M. Lekovic, "Investment Diversification as a Strategy for Reducing Investment Risk," *Ekon. horizonti*, vol. 20, no. 2, pp. 169–184, 2018.

[7] A. Srimurdianti, Sukamto, W. Setiawan, E. Esyudha, and Pratama, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Data Mining untuk Pengelompokan Saham pada Sektor Energi dengan Metode K-Means," *J. Edukasi dan Penelit. Inform.*, vol. 9, no. 1, pp. 76–81, 2023.

[8] S. Liu and X. Liu, "Research on Density-Based K-means Clustering Algorithm," *J. Phys. Conf. Ser.*, vol. 2137, no. 012071, 2021.

[9] B. Siregar and Y. Yosia, "Implementation of K-means Clustering Algorithm for the Indonesian Stock Exchange," *J. Sisfotek Glob.*, vol. 14, no. 1, pp. 49–56, 2024.

[10] A. A. N. DEVI, K. DHARMAWAN, and N. K. T. TASTRAWATI, "Pengelompokan Saham Menggunakan K-Means Dalam Pembentukan Portofolio Optimal," *E-Jurnal Mat.*, vol. 12, no. 4, pp. 302–308, 2023.

[11] Z. Cebeci and F. Yildiz, "Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures," *J. Agric. Informatics*, vol. 6, no. 3, pp. 13–23, 2015.

[12] D. Wu, X. Wang, and S. Wu, "Construction of stock portfolios based on k-means clustering of continuous trend features," *Knowledge-Based Syst.*, vol. 252, p. 109358, 2022.

[13] L. Gubu, E. Cahyono, H. Budiman, and M. Kabil Djafar, "Cluster Analysis for Mean-Variance Portfolio Selection: a Comparison Between K-Means and K-Medoids Clustering," *J. Ris. Ap. Mat*, vol. 07, no. 02, pp. 104–115, 2023.

[14] S. K. Kumari, P. Kumar, J. Priya, S. Surya, and A. K. Bhurjee, "Mean-value at risk portfolio selection problem using clustering technique: A case study," *AIP Conf. Proc.*, vol. 2112, no. 020178, 2019.

[15] B. Aslam, R. A. Bhuiyan, and C. Zhang, "Portfolio Construction with K-Means Clustering Algorithm Based on Three Factors," *MATEC Web Conf.*, vol. 377, 2023.

[16] A. Iriany, H. R. A. Putri, and H. M. Tua, "K-means Nonhierarchical Cluster and Dbscan Outlier Detection In the Grouping of Stock Issuers," *Proof*, vol. 3, no. 4, pp. 21–28, 2023.

[17] H. Firmansyah and D. Rosadi, "Construction of stock portfolios based on k-means clustering of continuous trend features," *Knowledge-Based Syst.*, vol. 20, no. 2, pp. 149–172, 2024.

[18] I. Nathania and S. Sumani, "Volatility and Liquidity Comparison of Indonesian and Singapore Stock Market in COVID-19 Mobility Restrictions Era," *Binus Bus. Rev.*, vol. 14, no. 3, pp. 235–245, 2023.

[19] D. Virmani, S. Taneja, and G. Malhotra, "Normalization based K means Clustering Algorithm," *Int. J. Adv. Eng. Res. Sci.*, vol. 2, no. 2, pp. 36–40, 2015.

[20] H. Simanjutak, . Sawaluddin, and M. Zarlis, "Analysis of DBSCAN Algorithm in Determining Epsilon Parameters Numerical Data Clustering," in *Proceedings ofthe International Conference on Natural Resources andTechnology*, 2019, pp. 220–222.

[21] M. Fuad, E. M. S. Rochman, and A. Rachmad, "Salt Commodity Data Clustering Using Fuzzy C-Means," *J. Phys. Conf. Ser.*, vol. 2406, no. 1, pp. 0–9, 2022.

[22] K. S. Pranata, A. A. S. Gunawan, and F. L. Gaol, "Development clustering system IDX company with k-means algorithm and DBSCAN based on fundamental indicator and ESG," *Procedia Comput. Sci.*, vol. 216, pp. 319–327, 2023.

[23] N. A. Maori and E. Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 277–287, 2023.

[24] P. Dziuba, D. Glukhova, and K. Shtogrin, "Risk, Return and International Portfolio Diversification: K-Means Clustering Data," *Balt. J. Econ. Stud.*, vol. 8, no. 3, pp. 53–64, 2022.

[25] A. Fawaid Ridwan and S. Supian, "IDX30 Stocks Clustering with K-Means Algorithm based on Expected Return and Value at Risk," *Int. J. Quant. Res. Model.*, vol. 2, no. 4, pp. 201–208, 2021.