

Winsorization for Outliers in Clustering Non-Cyclical Stocks with K-Means and K-Medoids*

Winsorization untuk Penanganan Pencilan dalam Penggerombolan Saham Sektor Consumer Non-Cyclical dengan K-Means dan K-Medoids

Naura Tirza Ardhani¹, Khairil Anwar Notodiputro^{1‡}, Sachnaz Desta Oktarina¹

¹ Study Program on Statistics and Data Science, IPB University, Indonesia

[‡]corresponding author: khairil@apps.ipb.ac.id

Copyright © 2025 Naura Tirza Ardhani, Khairil Anwar Notodiputro, and Sachnaz Desta Oktarina. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Non-cyclical consumer sector stocks are often chosen by investors because the products in this sector are essential products that always in demand by society. Therefore, the demand for these products tends to be stable and defensive or less affected by economic shocks. However, it does not guarantee that every stock in this sector has good performance, thus it is necessary to group stocks based on their fundamental indicators in the form of financial ratios. This research aims to identify the best method by considering outliers and determining the clusters with the best fundamental performance as a recommendation for investors to make the right investment decisions. The data used in this study is secondary data with observations in the form of 50 non-cyclical consumer sector stocks. The variables used are Earning per Share, Return on Equity, Return on Assets, Debt to Equity Ratio, Price to Earnings Ratio, and Price to Book Value. The clustering results indicated that K-Medoids is the best clustering method, both on the data before and after handling extreme outliers with winsorization approach. However, the optimum number of clusters before and after winsorization are different, with 3 and 6 clusters. Considering the influence of extreme outliers and to get a more informative clustering result, the clustering result after the application of winsorization technique was chosen, which resulted in 6 clusters. Cluster 1, which consists of AALI, GGRM, INDF, and SGRO can be recommended because it has excellent fundamental performance, especially in terms of Earning per Share in 2022.

Keywords: fundamentals, K-Medoids, outliers, stocks, winsorization.

* Received: Jul 2024d; Reviewed: Aug 2024; Published: Jun 2025

1. Pendahuluan

Penggerombolan adalah suatu metode untuk mengelompokkan objek ke dalam subset tertentu sehingga karakteristik objek di dalam setiap subset memiliki kesamaan yang tinggi, sementara karakteristik antar-subset memiliki perbedaan yang signifikan (Xu & Wunsch, 2008). Metode ini terbagi menjadi dua jenis, yaitu hierarki dan non-hierarki. Perbedaan dari kedua jenis tersebut adalah banyaknya gerombol pada metode berhierarki ditentukan setelah proses penggerombolan sedangkan banyaknya gerombol pada metode non-hierarki ditentukan sebelum proses penggerombolan. Metode non-hierarki yang sering digunakan ialah metode *K-Means* dan *K-Medoids*. Saat ini telah banyak dikembangkan metode penggerombolan oleh para ahli dan diaplikasikan dalam berbagai bidang. Salah satu bidang yang dapat memanfaatkan adanya metode untuk penggerombolan objek adalah bidang investasi (Ristiyanti, 2017).

Data dari Kustodian Sentral Efek Indonesia (KSEI, n.d.) menunjukkan bahwa jumlah investor pasar modal meningkat sebesar 17,95% menjadi 12,16 juta investor pada Desember 2023 dibandingkan dengan tahun 2022. Investasi saham, terutama di sektor consumer non-cyclical, banyak diminati karena produk di sektor ini merupakan kebutuhan pokok yang stabil dan tidak terlalu terpengaruh oleh fluktuasi ekonomi. Investasi jangka panjang menjadi strategi yang disarankan untuk mendapatkan keuntungan lebih besar, meskipun diiringi dengan risiko tinggi akibat fluktuasi harga saham (Aufa Hasibuan et al., 2021). Untuk meminimalisir risiko, penting bagi investor untuk menganalisis fundamental saham berdasarkan laporan keuangan perusahaan. Analisis ini membantu memahami kinerja perusahaan melalui rasio keuangan. Namun, proses ini memakan waktu jika harus membaca laporan keuangan perusahaan satu per satu. Oleh karena itu, penggerombolan saham berdasarkan indikator fundamental berupa rasio keuangan diharapkan dapat mempermudah investor dalam memilih saham sesuai kriteria yang diinginkan.

Data rasio keuangan pada saham tidak terlepas dari adanya nilai pencilan. Pencilan atau outlier didefinisikan sebagai pengamatan yang terletak jauh dari sebaran data lainnya. Keberadaan pencilan pada data memiliki dampak yang signifikan terhadap hasil analisis dan berpotensi menghasilkan kesimpulan yang tidak akurat. Dengan demikian, pada penelitian ini diperlukan suatu metode untuk menangani data yang bersifat pencilan sehingga hasil analisis dapat lebih akurat dan interpretatif. Salah satu metode yang dapat digunakan adalah dengan menggunakan pendekatan *winsorization*.

Sejumlah penelitian relevan telah dilakukan sebelumnya. Penelitian yang dilakukan oleh Khairunnisa (2018) menunjukkan bahwa algoritma pengelompokan terbaik untuk mengelompokkan data yang mengandung pencilan adalah dengan algoritma *K-Medoids*. Soepriyono dan Triayudi (2023) melakukan penelitian mengenai perbandingan algoritma clustering untuk memprediksi harga saham, didapatkan bahwa setiap data saham memiliki kluster optimal berdasarkan proses pengelompokan dengan algoritma *K-Means*. Kemudian, pada kasus lain, hasil penelitian yang dilakukan oleh (Prabowo, 2023) mengenai perbandingan metode *K-Means*, *K-Medoids*, dan Hierarki menunjukkan bahwa metode analisis kluster optimal yang diperoleh adalah *K-Medoids* dengan 4 gerombol. Berdasarkan latar belakang yang telah diuraikan, penelitian ini bertujuan mengidentifikasi metode terbaik dalam menggerombolan saham sektor *consumer non-cyclical* dengan mempertimbangkan pencilan dan menentukan gerombol dengan performa fundamental terbaik sebagai rekomendasi bagi investor untuk pengambilan keputusan investasi yang tepat.

2. Metodologi

2.1 Bahan dan Data

Penelitian ini menggunakan data indikator fundamental berupa rasio keuangan saham berdasarkan hasil laporan keuangan yang dibukukan pada akhir tahun 2022. Data yang digunakan pada penelitian ini merupakan data sekunder berupa 50 saham sektor consumer non-cyclical yang termasuk ke dalam papan perdagangan utama. Data ini bersumber dari laman Bursa Efek Indonesia (BEI) dan situs <https://www.stockbit.com>. Peubah yang digunakan pada penelitian ini adalah sebanyak enam peubah yang tertera pada Tabel 1.

Tabel 1: Peubah-peubah yang digunakan dalam penelitian

Peubah	Satuan	Rujukan
<i>Earnings per Share</i> (EPS)	Rp (rupiah)	(Hervita Nenobais et al., 2022)
<i>Return on Equity</i> (ROE)	% (persen)	(Yana & Agustiningsih, 2022)
<i>Return on Assets</i> (ROA)	% (persen)	(Nenobais et al. 2022)
<i>Debt to Equity Ratio</i> (DER)	-	(Fitriani 2016)
<i>Price to Earnings Ratio</i> (PER)	-	(Yana & Agustiningsih, 2022)
<i>Price to Book Value</i> (PBV)	-	(Fitriani, 2016)

2.2 Winsorization

Dalam suatu data, pencilan dapat disebabkan oleh variabilitas alami atau kesalahan dalam pengumpulan data. Pencilan dapat sangat memengaruhi hasil analisis penggerombolan. Keberadaan pencilan dapat menggeser *centroid* gerombol, menyebabkan hasil penggerombolan yang tidak akurat dan menyesatkan (Gentle et al., 1991). Oleh karena itu, penanganan pencilan sangat penting dalam analisis data karena dapat memengaruhi hasil analisis dan interpretasi. Tukey (1962) menyarankan bahwa *winsorization* dapat membantu dalam menstabilkan rata-rata dan keragaman dalam *dataset* yang mengandung pencilan. *Winsorization* merupakan pendekatan transformasi statistik dengan membatasi nilai ekstrem dalam data statistik untuk mengurangi efek kemungkinan pencilan (Wilcox 2021). Pendekatan ini dapat dilakukan dengan memodifikasi data ekstrem menjadi nilai pada persentil tertentu misal pada persentil ke-5 dan persentil ke-95 sehingga data yang dianggap outlier dapat diubah menjadi nilai yang lebih representatif.

2.3 Penggerombolan *K-Means*

K-Means merupakan metode pada analisis gerombol yang mengalokasikan objek secara tegas sebagai anggota suatu gerombol sehingga suatu objek hanya berada pada satu gerombol tertentu. Menurut Wu (2012) metode ini bertujuan untuk mengelompokkan data ke dalam *k* gerombol dengan masing-masing gerombol diwakilkan oleh rata-rata dari setiap objek dalam gerombol tersebut. Metode ini memiliki kelebihan, yaitu sederhana dan sangat efisien. Sementara itu, kekurangan dari metode ini adalah sensitif terhadap pencilan dan membutuhkan inisiasi jumlah gerombol oleh peneliti (Oded Maimon, 2005).

Tahapan dengan menggunakan algoritma *K-Means* adalah sebagai berikut (Wu 2012):

1. Menentukan jumlah gerombol sebanyak k dengan mencoba $k = 2, \dots, 6$.
2. Menentukan nilai *centroid* awal secara acak sebanyak k .
3. Menghitung jarak antara titik *centroid* dengan titik tiap objek menggunakan persamaan ukuran jarak *Euclidean*.
4. Tandai jarak terdekat objek ke *centroid* awal dan kelompokkan menjadi cluster awal.
5. Menentukan nilai *centroid* baru dari masing-masing gerombol. *Centroid* baru didapatkan dari nilai rata-rata objek dalam gerombol.
6. Menghitung jarak antara titik *centroid* dengan titik tiap objek menggunakan persamaan ukuran jarak *Euclidean*.
7. Tandai jarak terdekat objek ke *centroid* baru dan kelompokkan menjadi gerombol baru.
8. Kembali ke langkah 5, lakukan pengulangan hingga nilai *centroid* yang dihasilkan tetap dan anggota gerombol tidak berpindah ke gerombol lain.

2.3 Penggerombolan *K-Medoids*

K-Medoids merupakan salah satu metode gerombol non-hierarki yang dikembangkan dari metode *K-Means*. Metode *K-Medoids* menggunakan data yang berada di tengah gerombol (median) sehingga metode ini lebih kekar terhadap pencilon dibandingkan dengan metode *K-means*. Tahapan dengan menggunakan algoritma *K-Medoids* adalah sebagai berikut (Gentle et al., 2009):

1. Menentukan k sebagai banyaknya gerombol dengan mencoba $k = 2, \dots, 6$.
2. Mengalokasikan setiap data ke gerombol terdekat menggunakan persamaan jarak *Euclidean*.
3. Memilih secara acak objek pada masing-masing gerombol sebagai kandidat *medoid* baru.
4. Menghitung jarak setiap objek ke *medoid* baru dengan menggunakan persamaan jarak *Euclidean*.
5. Menghitung total simpangan (S), dengan menghitung selisih nilai total jarak baru dengan total jarak lama. Apabila $S < 0$, maka tukar objek dengan data klaster untuk membentuk set objek baru sebanyak k sebagai *medoid*.
6. Ulangi langkah-langkah 3 hingga 5 sampai tidak ada perubahan pada *medoid* sehingga diperoleh gerombol dan anggota dari setiap gerombol.

2.4 Koefisien *Silhouette*

Koefisien *Silhouette* adalah ukuran berbasis internal yang digunakan untuk mengevaluasi kualitas penggerombolan. Koefisien *Silhouette* menggabungkan metode kohesi untuk mengukur kedekatan hubungan antara objek dalam sebuah gerombol dan metode separasi yang berfungsi untuk mengukur sejauh mana gerombol tersebut terpisah dengan gerombol lainnya (Rousseeuw 1987). Nilai koefisien *Silhouette* berkisar antara -1 hingga 1. Nilai 1 menunjukkan objek yang dikelompokkan dengan baik, 0 menunjukkan objek berada di perbatasan gerombol, dan -1 menunjukkan objek yang kemungkinan berada di gerombol yang tidak tepat. Menurut Kauffman dan Rousseeuw (2009), nilai koefisien *Silhouette* dapat dihitung menggunakan persamaan sebagai berikut:

$$SC = \frac{b_i - a_i}{\max \{a_i, b_i\}} \quad (1)$$

a_i adalah rata-rata objek ke- i dengan semua objek lainnya di dalam gerombol dan b_i adalah jarak minimum objek dari rata-rata jarak objek ke- i dengan semua objek lainnya pada gerombol berbeda.

2.5 Rasio Simpangan Baku

Rasio simpangan baku kerap dijadikan metode untuk melakukan validasi gerombol. Rasio ini didapat dengan menghitung nilai keragaman dalam gerombol (S_w) dan nilai keragaman antargerombol (S_b). Gerombol dengan nilai rata-rata rasio S_w dan S_b terkecil dipilih sebagai gerombol terbaik. Rumus untuk menghitung rasio tersebut adalah sebagai berikut (Gorunescu, 2022).

$$S_w = \frac{1}{c} \sum_{i=1}^c S_{ij} \quad (2)$$

c adalah ukuran gerombol yang digunakan dan S_{ij} adalah simpangan baku gerombol ke- i pada peubah ke- j .

$$S_b = \left[\frac{1}{c-1} \sum_{i=1}^c (v_{ij} - \bar{v}_j)^2 \right]^{\frac{1}{2}} \quad (3)$$

v_{ij} adalah pusat gerombol ke- i pada peubah ke- j dan \bar{v}_j adalah rata-rata keseluruhan gerombol pada peubah ke- j .

2.2 Metode Penelitian

Analisis penggerombolan saham dilakukan dengan menggunakan perangkat lunak R dengan package *cluster*, *factoextra*, dan *DescTools*. Tahapan analisis yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Melakukan praproses data dengan memilih saham yang memenuhi kriteria tertentu untuk disertakan dalam analisis, yaitu:
 - a. Saham-saham pada sektor *consumer non-cyclical* yang termasuk ke dalam papan perdagangan utama.
 - b. Saham yang memiliki batas IPO (*Initial Public Offering*) pada akhir tahun 2021.
2. Melakukan analisis penggerombolan menggunakan data sebelum penerapan pendekatan *Winsorization* dan dengan sesudah penerapan pendekatan *winsorization*.
 - a. Analisis gerombol sebelum penerapan pendekatan *winsorization*:
 1. Melakukan eksplorasi data dengan matriks korelasi, adapun untuk identifikasi pencilan menggunakan *boxplot*.
 2. Melakukan standarisasi data untuk menyeragamkan peubah dengan skala yang berbeda dengan menggunakan *z-score*.
 3. Melakukan Analisis Komponen Utama jika terdapat korelasi yang cukup kuat.
 4. Melakukan penggerombolan dengan metode *K-Means* dan *K-Medoids* dengan banyaknya gerombol sebesar 2 sampai 6.

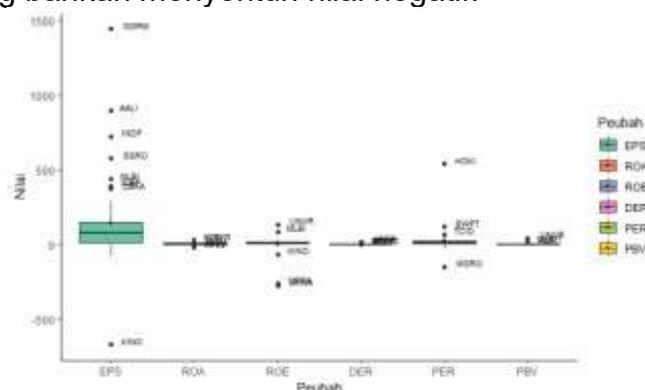
5. Mengevaluasi banyaknya gerombol optimum berdasarkan koefisien *Silhouette* pada setiap metode.
 6. Pemilihan metode terbaik dengan melakukan validasi hasil gerombol optimum berdasarkan nilai rasio simpangan baku.
- b. Analisis gerombol sesudah penerapan pendekatan *winsorization*:
1. Melakukan penerapan penanganan pencilan menggunakan pendekatan *winsorization* sebesar 5%.
 2. Melakukan pengulangan langkah 1(a) – 6(a).
 3. Melakukan pemilihan hasil penggerombolan terbaik.
 4. Menginterpretasikan karakteristik masing-masing gerombol pada hasil penggerombolan terbaik yang dapat digunakan sebagai rekomendasi untuk para investor berdasarkan rasio keuangan.

3. Hasil dan Pembahasan

3.1 Analisis Gerombol Sebelum Penerapan Pendekatan *Winsorization*

3.1.1 Eksplorasi Data

Gambar 1 menunjukkan bahwa terdapat beberapa emiten yang memiliki nilai rasio keuangan sangat berbeda dari mayoritas, tercermin dari adanya pencilan yang sangat signifikan pada masing-masing peubah. Peubah EPS memiliki jumlah pencilan yang mendominasi dibandingkan dengan peubah lainnya. Terlihat bahwa terdapat pencilan ekstrem yaitu emiten GGRM yang merupakan pencilan atas dan emiten KINO sebagai pencilan bawah pada peubah EPS. Kemudian, terlihat bahwa emiten UNVR menjadi amatan pencilan atas pada tiga peubah yang berbeda, yaitu ROE, ROA, dan PBV. Lalu pada peubah ROE, didapatkan tiga pencilan bawah yang terdiri atas emiten KINO, MPPA, dan JAWA yang bahkan menyentuh nilai negatif.



Gambar 1: Boxplot masing-masing peubah

Selanjutnya, untuk melihat intensitas hubungan antarpeubah, digunakan matriks korelasi. Gambar 2 menunjukkan bahwa terdapat hubungan positif dan negatif yang kuat antarpeubah. Hubungan positif yang kuat dimiliki oleh peubah ROE dan ROA. Korelasi positif yang kuat sebesar 0,68 menunjukkan bahwa ROE dan ROA cenderung bergerak searah dengan hubungan yang signifikan. Kemudian, korelasi negatif yang kuat sebesar -0,78 menunjukkan bahwa semakin meningkat nilai DER atau utang suatu perusahaan maka cenderung akan menurunkan ROE atau laba bersih dari perusahaan tersebut.



Gambar 2: Matriks korelasi antarpeubah

Dikarenakan adanya korelasi yang kuat antarpeubah, maka selanjutnya dilakukan analisis komponen utama (AKU) untuk meningkatkan akurasi dari hasil penggerombolan dengan data yang sudah distandardisasi terlebih dahulu. Menurut James *et al.* (2021), batas keragaman kumulatif yang sering digunakan berada pada rentang 70% hingga 90%. Selain itu, komponen utama juga dapat dipilih jika nilai eigen atau akar ciri lebih besar dari 1. Dengan demikian, dipilih 3 komponen utama (KU) karena dengan 3 KU tersebut sudah dapat menjelaskan keragaman sejumlah 77,969% dengan nilai akar ciri sebesar 1,050.

3.1.2 Penggerombolan Menggunakan Metode *K-Means*

Tabel 2 menunjukkan evaluasi banyaknya gerombol optimum pada metode *K-Means*. Pemilihan nilai k yang optimal ditentukan berdasarkan nilai koefisien *Silhouette* yang semakin besar. Berdasarkan Tabel 2, didapatkan koefisien *Silhouette* tertinggi ketika gerombol berjumlah 2. Oleh karena itu, pada metode *K-Means* dipilih hasil penggerombolan dengan jumlah 2 gerombol.

Tabel 2: Evaluasi banyaknya gerombol optimum pada *K-Means* dengan koefisien *Silhouette*

Metode	Banyak Gerombol	Koefisien <i>Silhouette</i>
<i>K-Means</i>	2	0,686
	3	0,684
	4	0,399
	5	0,401
	6	0,369

3.1.3 Hasil Penggerombolan Menggunakan Metode *K-Medoids*

Penggerombolan selanjutnya dilakukan menggunakan metode *K-Medoids*. Banyaknya gerombol yang digunakan, yaitu $k = 2$ sampai dengan $k = 6$. Berdasarkan Tabel 3, koefisien *Silhouette* tertinggi diperoleh ketika gerombol berjumlah 3. Oleh karena itu, pada metode *K-Medoids* dipilih hasil penggerombolan dengan jumlah 3 gerombol.

Tabel 3: Evaluasi penggerombolan *K-Medoids* dengan koefisien *Silhouette*

Metode	Banyak Gerombol	Koefisien <i>Silhouette</i>
<i>K-Medoids</i>	2	0,686
	3	0,689
	4	0,289
	5	0,287
	6	0,359

3.1.4 Pemilihan Metode Terbaik dengan Validasi Gerombol

Pemilihan metode terbaik dilakukan berdasarkan hasil penggerombolan optimum kedua metode yang telah dilakukan sebelumnya. Hasil penggerombolan dari metode *K-Means* dan *K-Medoids* masing-masing memperoleh banyak gerombol optimum dan anggota gerombol yang berbeda. Dengan demikian, diperlukan validasi hasil penggerombolan menggunakan rasio S_w / S_b . Metode yang memiliki nilai koefisien *Silhouette* tertinggi serta rasio S_w / S_b terendah akan dipilih sebagai metode terbaik yang dapat menggerombolkan saham sektor *consumer non-cylical* berdasarkan rasio keuangan selama tahun 2022.

Tabel 4: Perbandingan nilai koefisien *Silhouette* dan rasio S_w / S_b

Metode	Banyak Gerombol	Koefisien <i>Silhouette</i>	S_w / S_b
<i>K-Means</i>	2	0,686	0,322
<i>K-Medoids</i>	3	0,689	0,268

Berdasarkan Tabel 4, metode *K-Medoids* dengan jumlah 3 gerombol terpilih sebagai metode terbaik karena menghasilkan nilai koefisien *Silhouette* yang lebih tinggi dan rasio S_w / S_b lebih rendah dibandingkan dengan metode *K-Means*. Selanjutnya, ditampilkan hasil penggerombolan yang berisikan tiap anggota dari setiap gerombol pada Tabel 5.

Tabel 5: Anggota gerombol hasil penggerombolan metode *K-Medoids*

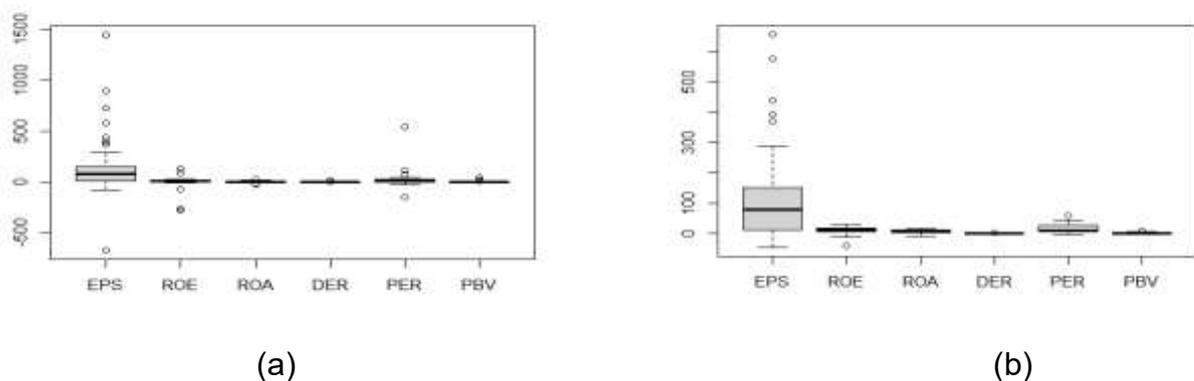
Gerombol	Banyak Anggota	Kode Saham
1	46	AALI, AISA, AMRT, ANJT, BISI, BUDI, BWPT, CAMP, CEKA, CLEO, CMRY, CPIN, CSRA, DLTA, DSFI, DSNG, GGRM, GOOD, GZCO, HMSP, HOKI, ICBP, INDF, JPFA, LSIP, MAIN, MBTO, MGRO, MLBI, MLPL, MRAT, MYOR, PMMP, PSGO, ROTI, SDPC, SGRO, SIMP, SSMS, TAPG, TBLA, TCID, UCID, ULTJ, VICI, WIIM
2	3	JAWA, KINO, MPPA
3	1	UNVR

3.2 Penggerombolan dengan Pendekatan *Winsorization*

Penggerombolan sebelumnya dilakukan dengan menggunakan data asli yang memiliki banyak *outlier* atau pencilan ekstrem yang signifikan. Sementara itu, keberadaan pencilan pada data memiliki dampak yang signifikan terhadap hasil analisis dan berpotensi menghasilkan kesimpulan yang tidak akurat. Oleh karena itu, untuk meningkatkan kualitas hasil analisis dan interpretabilitas maka penggerombolan selanjutnya akan dilakukan dengan menggunakan pendekatan *winsorization* terlebih dahulu. *Winsorization* pada tingkat 5% adalah praktik umum dalam statistik karena dapat meminimalkan pengaruh pencilan tanpa mengubah struktur data secara signifikan. Maka dari itu, pada penelitian ini, data awal akan dilakukan menggunakan proses *winsorization* terlebih dahulu sebesar 5%. Artinya, nilai-nilai di bawah persentil ke-5 dan di atas persentil ke-95 akan digantikan dengan nilai pada persentil ke-5 dan ke-95 masing-masing.

3.2.1 Eksplorasi Data Hasil *Winsorization*

Eksplorasi data akan dilakukan dengan melihat perbandingan boxplot pada data sebelum dan sesudah dilakukan *winsorization* sebesar 5%. Hal ini bertujuan untuk melihat perubahan terkait keberadaan pencilan. Gambar 3 menunjukkan bahwa penerapan pendekatan *winsorization* membantu untuk menghilangkan atau mengurangi pencilan yang ekstrem pada sebagian besar peubah, terutama pada peubah EPS, ROE, dan PER sehingga data menjadi lebih terfokus dan menyebar dalam rentang nilai yang lebih sempit.



Gambar 3: Boxplot sebelum (a) dan sesudah (b) dilakukan *winsorization*

Eksplorasi selanjutnya dilakukan dengan menggunakan matriks korelasi untuk mengidentifikasi hubungan antarpeubah. Gambar 4 memperlihatkan bahwa terdapat korelasi positif yang sangat kuat antara peubah ROE dan ROA, yaitu sebesar 0,87. Artinya, perubahan dalam ROE biasanya diikuti oleh perubahan serupa dalam ROA dengan tingkat yang signifikan. Korelasi negatif yang kuat juga dimiliki oleh peubah DER dengan ROA yang menunjukkan korelasi sebesar -0,61. Hal ini menunjukkan bahwa perubahan dalam DER biasanya diikuti oleh perubahan terbalik yang signifikan dalam ROA. Sementara itu, korelasi paling rendah dimiliki oleh peubah PBV dan DER yang hanya sebesar 0,05. Hal tersebut dapat menunjukkan bahwa peningkatan atau penurunan dalam DER tidak secara signifikan berkaitan dengan perubahan dalam PBV, begitupun sebaliknya.



Gambar 4: Matriks korelasi antarpeubah setelah dilakukan *winsorization*

Dengan adanya beberapa nilai korelasi yang menunjukkan hubungan kuat antarpeubah, maka selanjutnya akan dilakukan analisis komponen utama dengan menggunakan data yang sudah distandardisasi. Penentuan jumlah komponen utama didasarkan pada nilai proporsi keragaman dengan rentang 70% hingga 90% (Jolliffe 2002). Selain itu, batas nilai akar ciri yang baik adalah 0,7. Dengan demikian, komponen utama yang dipilih adalah 3 KU karena dapat menjelaskan keragaman data sebesar 80,359% serta akar ciri yang bernilai 0,796.

3.2.2 Hasil Penggerombolan Menggunakan Metode *K-Means*

Tabel 6 menunjukkan bahwa pada penggerombolan dengan metode *K-Means*, nilai koefisien *Silhouette* tertinggi didapatkan ketika gerombol berjumlah 5. Oleh karena itu, pada metode *K-Means* dipilih hasil penggerombolan dengan jumlah 5 gerombol.

Tabel 6: Evaluasi penggerombolan *K-Means* dengan koefisien *Silhouette*

Metode	Banyak Gerombol	Koefisien <i>Silhouette</i>
<i>K-Means</i>	2	0,308
	3	0,343
	4	0,283
	5	0,359
	6	0,305

3.2.3 Hasil Penggerombolan Menggunakan Metode *K-Medoids*

Selanjutnya, penggerombolan dilakukan dengan metode *K-Medoids*. Jumlah gerombol yang digunakan adalah antara $k = 2$ dan $k = 6$. Nilai koefisien *Silhouette* digunakan untuk menentukan jumlah gerombol yang optimal. Tabel 7 menunjukkan evaluasi gerombol menggunakan nilai koefisien *Silhouette* dari metode *K-Medoids*. Berdasarkan Tabel 7, koefisien *Silhouette* tertinggi, yaitu sebesar 0,374 diperoleh ketika gerombol berjumlah 6. Oleh karena itu, dipilih hasil penggerombolan dengan jumlah 6 gerombol pada metode *K-Medoids*.

Tabel 7: Evaluasi penggerombolan *K-Medoids* dengan koefisien *Silhouette*

Metode	Banyak Gerombol	Koefisien <i>Silhouette</i>
<i>K-Medoids</i>	2	0,285
	3	0,297
	4	0,345
	5	0,334
	6	0,374

3.2.4 Pemilihan Metode Terbaik dengan Validasi Gerombol

Hasil penggerombolan dari metode *K-Means* dan *K-Medoids* masing-masing memperoleh banyak gerombol optimum dan anggota gerombol yang berbeda. Maka dari itu, diperlukan validasi hasil penggerombolan menggunakan rasio S_w / S_b .

Tabel 8: Perbandingan nilai koefisien *Silhouette* dan rasio S_w / S_b tiap metode

Metode	Banyak Gerombol	Koefisien <i>Silhouette</i>	S_w / S_b
<i>K-Means</i>	5	0,359	0.468
<i>K-Medoids</i>	6	0,374	0.427

Tabel 8 memperlihatkan bahwa nilai koefisien *Silhouette* tertinggi serta rasio S_w / S_b terendah dimiliki oleh metode penggerombolan *K-Medoids* dengan banyaknya gerombol optimum sebesar 6. Oleh karena itu, metode *K-Medoids* dipilih sebagai metode terbaik pada data yang telah diterapkan menggunakan pendekatan *winsorization*. Selanjutnya, ditampilkan hasil penggerombolan yang berisikan tiap anggota dari setiap gerombol pada Tabel 9.

Tabel 9: Anggota gerombol hasil penggerombolan metode *K-Medoids*

Gerombol	Banyaknya Anggota	Kode Saham
1	4	AALI, GGRM, INDF, SGRO
2	14	AISA, ANJT, BUDI, DSFI, GZCO, JPFA, MBTO, MLPL, PMMP, PSGO, SDPC, SIMP, TBLA, UCID
3	10	AMRT, CLEO, CMRY, CPIN, MLBI, GOOD, MLBI, MYOR, UNVR, VICI
4	14	BISI, CAMP, CEKA, CSRA, DLTA, DSNG, HMSP, ICBP, LSIP, MRAT, ROTI, SSMS, ULTJ, WIIM
5	4	BWPT, HOKI, MAIN, TCID
6	4	JAWA, KINO, MGRO, MPPA

3.3 Pemilihan dan Interpretasi Hasil Penggerombolan Terbaik

Analisis penggerombolan yang telah dilakukan sebelumnya menunjukkan bahwa pada kedua kondisi yang berbeda, metode penggerombolan *K-Medoids* dipilih sebagai metode terbaik. Meskipun demikian, dengan mempertimbangkan pengaruh pencilan ekstrem serta untuk mendapatkan hasil penggerombolan yang lebih informatif, dipilihlah hasil penggerombolan *K-Medoids* setelah penerapan pendekatan *winsorization* yang menghasilkan 6 gerombol. Untuk menginterpretasikan karakteristik dari setiap gerombol, disajikan nilai rata-rata semua peubah untuk setiap gerombol yang dapat dilihat pada Tabel 10.

Tabel 10: Rataan setiap gerombol untuk masing-masing peubah

Gerombol	Rataan					
	EPS	ROE	ROA	DER	PER	PBV
1	637,88	11,29	5,70	0,56	8,59	0,79
2	43,79	5,92	2,34	0,88	7,83	0,85
3	121,08	19,33	11,78	0,32	33,25	6,26
4	164,25	18,47	11,98	0,29	10,58	1,85
5	26,08	0,85	0,40	1,16	54,12	0,92
6	-35,7	-31,3	-5,63	2,66	-2,26	4,16

Interpretasi karakteristik dari masing-masing gerombol yang dihasilkan adalah sebagai berikut:

1. Gerombol 1 terdiri atas emiten-emiten yang menunjukkan performa keuangan yang sangat baik dengan rata-rata EPS yang sangat tinggi dibandingkan dengan gerombol lainnya. Rata-rata ROE dan ROA yang dimiliki masing-masing sebesar 11,29% dan 5,7% yang menandakan bahwa emiten-emiten ini sangat efisien dalam menghasilkan laba baik dari ekuitas maupun dari total aset. DER dengan angka di bawah 1 menunjukkan bahwa penggunaan utang yang tidak terlalu besar sehingga masih tergolong pada kondisi finansial perusahaan yang sehat. PER sebesar 8,58x dan PBV sebesar 0,79x menunjukkan bahwa saham-saham dalam gerombol ini dapat dikatakan *undervalued* atau memiliki potensi pertumbuhan yang stabil. Saham yang *undervalued* biasanya dihargai di bawah nilai intrinsiknya.
2. Gerombol 2 terdiri atas emiten-emiten dengan rata-rata EPS, ROE, dan ROA positif yang masih tergolong baik tetapi cukup rendah. DER sebesar 0,88x menunjukkan bahwa emiten-emiten ini menggunakan sedikit lebih banyak utang dibandingkan dengan emiten pada gerombol 1. PER sebesar 7,83x dan PBV sebesar 0,85x memungkinkan saham-saham ini bersifat *undervalued*.
3. Gerombol 3 terdiri atas perusahaan-perusahaan dengan rata-rata nilai EPS sebesar Rp121.08 menunjukkan laba per saham yang cukup tinggi. Rata-rata ROE dan ROA yang dimiliki pada perusahaan-perusahaan ini besar, menandakan bahwa perusahaan-perusahaan ini memiliki tingkat efisiensi yang tinggi dalam penggunaan ekuitas dan aset untuk menghasilkan laba. DER yang cukup rendah juga menunjukkan penggunaan utang yang hanya sedikit. Namun, rata-rata PER dan PBV yang sangat tinggi menunjukkan bahwa saham-saham ini sangat memungkinkan untuk berpotensi overvaluation.

Artinya, saham-saham ini mungkin diperdagangkan pada harga yang sudah sangat tinggi atau bahkan dianggap dijual pada harga yang kurang wajar.

4. Gerombol 4 terdiri atas perusahaan-perusahaan yang memiliki rata-rata EPS cukup besar, yaitu Rp164.25. ROE dan ROA masing-masing sebesar 18.47% dan 11,98% menunjukkan efisiensi yang tinggi dalam menghasilkan laba dari ekuitas dan aset. DER yang rendah di angka 0,29x menunjukkan perusahaan-perusahaan ini hanya memiliki sedikit utang. Sementara itu, nilai PER dan PBV yang tinggi dapat memungkinkan bahwa saham-saham ini relatif dianggap sedikit overvalued.
5. Gerombol 5 terdiri atas perusahaan-perusahaan yang memiliki performa keuangan cukup lemah dengan rata-rata EPS yang hanya sebesar Rp26.08. ROE dan ROA masing-masing yang hanya sebesar 0.85% dan 0.4% menandakan efisiensi yang sangat rendah dalam penggunaan ekuitas dan aset dalam menghasilkan laba. Kemudian, DER yang tinggi dengan angka lebih dari 1 menunjukkan bahwa perusahaan ini bergantung relatif besar terhadap utang. PER yang sangat tinggi, yaitu sebesar 54,13x dapat terjadi dikarenakan laba perusahaan yang dimiliki sangat rendah atau mendekati nol. PBV sebesar 0,93x menunjukkan valuasi yang sedikit di atas nilai buku.
6. Gerombol 6 terdiri atas perusahaan-perusahaan yang mengalami performa keuangan sangat buruk pada tahun 2022. Hal ini ditunjukkan dengan rata-rata EPS, ROE, ROA, dan PER yang menyentuh angka negatif. DER yang sangat tinggi, yaitu 2,66x menandakan bahwa perusahaan-perusahaan ini ketergantungan sangat besar terhadap utang sehingga dapat meningkatkan risiko keuangan perusahaan. PBV yang sangat tinggi, yaitu 4,16x juga menunjukkan bahwa saham-saham ini dianggap overvaluation.

4. Simpulan dan Saran

4.1 Simpulan

Hasil penggerombolan saham sektor *consumer non-cyclical* berdasarkan rasio keuangan tahun 2022 menunjukkan bahwa metode *K-Medoids* merupakan metode terbaik, baik pada data sebelum maupun sesudah penanganan pencilan ekstrem dengan pendekatan *winsorization*. Meskipun demikian, jumlah gerombol optimum yang dihasilkan berbeda, yaitu 3 dan 6 gerombol. Gerombol dengan jumlah 3 memiliki persebaran anggota berturut-turut sebesar 46, 3, dan 1 kode saham, sedangkan gerombol dengan jumlah 6 memiliki persebaran anggota sebesar 4, 14, 10, 14, 4, dan 4 kode saham. Dengan mempertimbangkan pengaruh pencilan ekstrem serta untuk mendapatkan hasil penggerombolan yang lebih informatif, dipilihlah hasil penggerombolan setelah penerapan pendekatan *winsorization* yang menghasilkan 6 gerombol. Berdasarkan nilai rata-rata secara keseluruhan, gerombol 1 dipilih sebagai gerombol terbaik karena terdiri dari saham-saham dengan performa fundamental sangat baik pada tahun 2022. Gerombol ini terdiri atas empat kode saham, yaitu AALI, GGRM, INDF, dan SGRO. Oleh karena itu, keempat saham ini diduga dapat menjadi rekomendasi potensial bagi para investor untuk mengambil keputusan investasi yang tepat. Namun, penting bagi para investor untuk tetap melihat kembali pertumbuhan finansial tiap saham dari tahun ke tahun serta kondisi terkini.

4.2 Saran

Pada penelitian selanjutnya, diharapkan dapat melakukan penggerombolan dengan menggunakan data panel yang memiliki periode lebih panjang. Hal ini berfungsi untuk mengamati perubahan dan tren jangka panjang dalam performa saham dan stabilitas penggerombolan. Selain itu, analisis selanjutnya dapat

mempertimbangkan untuk menambah peubah indikator fundamental lainnya, seperti *dividen yield* dan *net profit margin* (NPM) atau pun dapat menambahkan peubah dengan faktor teknikal sehingga dapat memberikan gambaran yang lebih komprehensif mengenai performa saham. Kemudian, penggerombolan saham juga dapat dilakukan pada sektor lainnya.

Daftar Pustaka

- Aufa Hasibuan, R., Wira Rizki, S., & Perdana INTISARI, H. (2021). Perbandingan Metode Fuzzy Clustering Means Dan Single Linkage Pada Pengelompokan Saham Lq45. *Buletin Ilmiah Math. Stat. Dan Terapannya (Bimaster)*, 10(3): 361–368.
- Fitriani, R. (2016). Pengaruh NPM, PBV, dan DER Terhadap Harga Saham pada Perusahaan Sub Sektor Makanan dan Minuman di Bursa Efek Indonesia.
- Gentle, J. E., Kaufman, L., & Rousseeuw, P. J. (1991). Finding Groups in Data: An Introduction to Cluster Analysis. *Biometrics*, 47(2): 788. <https://doi.org/10.2307/2532178>
- Gentle, J. E., Kaufman, L., & Rousseeuw, P. J. (2009). Finding Groups in Data: An Introduction to Cluster Analysis. *Biometrics*, 47(2): 788. <https://doi.org/10.2307/2532178>
- Gorunescu, F. (2022). Data Mining Concepts, Models and Techniques. In *Mining of Massive Datasets* (Vol. 2). Berlin: Springer.
- Hervita Nenobais, A., Sia Niha, S., & Manafe, H. A. (2022). Pengaruh Return on Asset (ROA), Return on Equity (ROE), Net Profit Margin (NPM) dan Earning Per Share (EPS) terhadap Harga Saham (Suatu Kajian Studi Literatur Manajemen Keuangan Perusahaan). *Jurnal Ekonomi Manajemen Sistem Informasi*, 4(1): 10–22. <https://doi.org/10.31933/JEMSI.V4I1.1146>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning*. <https://doi.org/10.1007/978-1-0716-1418-1>
- Khairunnisa, I. W. (2018). *Perbandingan Algoritma Pengelompokan Non-Hierarki K-Medians Dan K-Medoids Pada Data Yang Mengandung Pencilan*.
- KSEI. (n.d.). Masuki Usia 26 Tahun, KSEI Siap Implementasikan 3 Rencana Strategis. 2023.
- Oded Maimon, L. R. (2005). Data Mining and Knowledge Discovery Handbook. *Data Mining and Knowledge Discovery Handbook*. <https://doi.org/10.1007/B107408/COVER>
- Prabowo, A. A. (2023). *Perbandingan Metode K-Means, K-Medoids, dan Hierarki untuk Penggerombolan Kabupaten/Kota di Sumatera Utara Berdasarkan Jenis Kekerasan terhadap Perempuan*.
- Rand R. Wilcox. (2021). Introduction to Robust Estimation and Hypothesis Testing | ScienceDirect.
- Ristiyanti. (2017). *Cluster Ensemble dalam Penggerombolan Provinsi di Indonesia Berdasarkan Indikator Pelayanan Kesehatan Ibu Hamil*.
- Soepriyono, G., & Triayudi, A. (2023). Perbandingan Kinerja Algoritma Clustering Data Mining Untuk Prediksi Harga Saham Pada Reksadana dengan Davies Bouldin Index. *Jurnal Media Informatika Budidarma*, 7(4): 2061–2073. <https://doi.org/10.30865/mib.v7i4.6623>
- Tukey, J. W. (1962). The Future of Data Analysis. *The Annals of Mathematical Statistics*, 33(1): 1–67. <https://doi.org/10.1214/AOMS/1177704711>,
- Wu, J. (2012). *Advances in K-means Clustering*. <https://doi.org/10.1007/978-3-642->

29807-3

Xu, R., & Wunsch, D. (2008). *Clustering*. Piscataway (NJ): Wiley-IEEE Press.

Yana, D., & Agustiningsih, W. (2022). Pengaruh Return on Equity (RoE), Debt to Equity Ratio (DER) dan Price Earnings Ratio (Per) Terhadap Harga Saham. *AKSELERASI: Jurnal Ilmiah Nasional*, 4(2): 67–77.
<https://doi.org/10.54783/JIN.V4I2.566>