

## Energy Sector Stock Price Forecasting with Time Series Clustering Approach\*

Peramalan Harga Saham Sektor Energi Dengan Pendekatan Penggerombolan Data Deret Waktu

Linda Sakinah<sup>1‡</sup>, Rahma Anisa<sup>1</sup>, I Made Sumertajaya<sup>1</sup>

<sup>1</sup>Department of Statistics, IPB University, Indonesia

<sup>‡</sup>corresponding author: [sakinah\\_linda@apps.ipb.ac.id](mailto:sakinah_linda@apps.ipb.ac.id)

Copyright © 2024 Linda Sakinah, Rahma Anisa, and I Made Sumertajaya. 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

Stock investment promises higher returns but carries high risks because unpredictable price fluctuations. Energy sector shows potential due to its highest sectoral index growth in 2022. However, this doesn't indicate that stock price increases occur evenly among all issuers. Therefore, it's necessary to analyze clustering of issuers based on similarity of their stock price movements and used for forecasting stock prices at cluster level. This study aims to evaluate performance of clustering energy sector issuers using autocorrelation-based distance and dynamic time warping(DTW), and to forecast stock prices at cluster level. The data used consists weekly closing stock prices. The clustering used hierarchical average linkage method. Stock price forecast for each cluster used ARIMA model and its performance was evaluated using rolling-cross validation. The results showed that DTW distance had the best clustering performance. Energy sector issuers were grouped into four clusters with strong cluster category, indicated by silhouette coefficient >0.71. ARIMA models for each cluster produced MAPE values between 10-20%, categorizing them as good forecasting models. Clusters A and D were recommended for investors because have highest potential for capital gain based on forecasted stock prices. That clusters also consisted of companies with strong fundamentals and dividend policies.

**Keywords:** ARIMA, autocorrelation-based, dynamic time warping, stock prices, time series clustering.

---

\* Received: Jul 2023; Reviewed: Oct 2023; Published: Dec 2024

## 1. Pendahuluan

*Clustering* atau penggerombolan adalah metode untuk mengatur titik data ke dalam kelompok berdasarkan kesamaannya (Kaufman & Rousseeuw, 2009). Analisis gerombol data deret waktu mempunyai prosedur penggerombolan yang berbeda dengan data statis karena data deret waktu memiliki korelasi fitur dan dimensi yang tinggi. Metode penggerombolan berhirarki sangat cocok untuk data deret waktu (Aghabozorgi et al., 2014). Pautan rata-rata dianggap lebih stabil dan lebih toleran terhadap pencilan (Mattjik & Sumertajaya, 2011).

Pada kasus penggerombolan emiten saham *consumer non cyclical*, jarak *dynamic time warping* (DTW) menunjukkan hasil penggerombolan yang lebih baik dibandingkan dengan jarak *Euclidean* (Al Fatah, 2022). Jarak DTW lebih unggul karena tidak hanya memperhitungkan kedekatan nilai, namun juga memperhitungkan kemiripan pola dan pergerakan data (Sardá-Espinosa, 2017). Penggerombolan penggerombolan data deret waktu juga dapat menggunakan jarak *autocorrelation-based* (D'Urso & Maharaj, 2009). Pendekatan jarak ini memungkinkan reduksi data deret waktu yang pada dasarnya berdimensi tinggi agar prosedur penggerombolan menjadi lebih efisien.

Jumlah investor di Indonesia terus meningkat sejalan dengan perkembangan teknologi dan kesadaran masyarakat untuk berinvestasi. *Press release* yang dikeluarkan oleh Kustodian Sentral Efek Indonesia pada 21 November 2022 menunjukkan kenaikan jumlah investor pasar modal sebesar 33,53% dibandingkan tahun 2021. Saham merupakan salah satu instrumen investasi pasar modal yang banyak diminati karena menjanjikan keuntungan yang lebih tinggi. Akan tetapi, risiko investasi saham juga tinggi karena fluktuasi harga saham yang sulit diprediksi. Hal ini berakibat pada tingkat pengembalian (*return*) saham yang juga bergerak fluktuatif. Saham sektor energi merupakan salah satu pilihan investasi yang potensial. Emiten sektor energi cenderung mengalami kenaikan indeks sektoral tertinggi di sepanjang tahun 2022 (BEI, 2022). Namun, naiknya indeks sektoral ini tidak menandakan bahwa kenaikan harga saham terjadi secara merata pada semua emiten sektor energi.

Penelitian ini bertujuan untuk mengevaluasi kinerja penggerombolan harga saham mingguan sektor energi dengan menggunakan jarak *autocorrelation-based* dan jarak DTW menggunakan metode berhirarki *average linkage*. Dari kesamaan pola yang dihasilkan, akan dilakukan peramalan harga saham pada level gerombol menggunakan Model *autoregressive integrated moving average* (ARIMA) sebagai informasi tambahan dalam pemberian rekomendasi kepada investor. Peramalan harga lebih efisien apabila dilakukan penggerombolan terlebih dahulu (Ulinnuha et al., 2022). Hasil dari penelitian ini dapat digunakan investor untuk memetakan investasinya agar mendapat keuntungan optimal.

## 2. Metodologi

### 2.1 Data

Penelitian ini menggunakan data sekunder yang diperoleh dari laman BEI pada <http://www.idx.co.id/> dan Yahoo Finance pada <https://finance.yahoo.com/>. Data yang digunakan adalah harga penutupan mingguan saham sektor energi dari perusahaan terbuka yang aktif dengan periode penawaran umum saham perdana atau *Initial Public Offering* (IPO) sebelum tahun 2019. Rentang waktu yang digunakan yaitu dari 1 Januari 2019 hingga 30 Desember 2022.

## 2.2 Prosedur Analisis Data

Analisis penggerombolan saham dilakukan dengan menggunakan perangkat lunak R. Prosedur penelitian yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Melakukan praproses data, yaitu memilih emiten yang memenuhi kriteria tertentu untuk disertakan dalam analisis. Data yang tidak memenuhi kriteria adalah:
  - a. Saham yang IPO pada tahun 2019 dan setelahnya.
  - b. Saham yang tidak terbuka dan tidak aktif dalam kegiatan jual beli di BEI.
2. Eksplorasi data menggunakan plot data harga penutupan saham mingguan 1 Januari 2019 hingga 30 Desember 2022 untuk melihat perilaku data harga saham.
3. Melakukan penggerombolan emiten dengan tahapan sebagai berikut:
  - a. Memeriksa kestasioneran data menggunakan plot ACF serta uji *Augmented Dickey Fuller* (ADF) dengan persamaan sebagai berikut:

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t$$

$y_t$  adalah amatan waktu ke- $t$ ,  $p$  adalah panjang ordo AR,  $\phi$  adalah koefisien AR ordo  $p$  ( $\sum_{i=1}^p \phi_i$ ), dan  $\varepsilon_t$  adalah sisaan pada waktu ke- $t$ .

- b. Melakukan transformasi Box-Cox jika terdapat data tidak stasioner dalam ragam dan *differencing* jika terdapat data tidak stasioner dalam nilai tengah, hingga diperoleh data yang stasioner.
- c. Melakukan perhitungan ukuran ketidakmiripan dengan menggunakan jarak sebagai berikut:

- 1) Jarak *Autocorrelation-Based* atau jarak autokorelasi

Autokorelasi suatu peubah acak  $x$  pada lag  $r$  ( $r = 1, \dots, T - 1 = R$ ) didefinisikan sebagai berikut (Kardiyan & Güney, 2015):

$$\hat{\rho}_{kr} = \frac{\sum_{t=r+1}^T (x_{kt} - \bar{x}_k)(x_{k(t-r)} - \bar{x}_k)}{\sum_{t=1}^T (x_{kt} - \bar{x}_k)^2}$$

Vektor dari perkiraan ukuran autokorelasi dari setiap pasangan deret waktu ( $x_k$  dan  $x_k'$ ) yaitu  $\hat{\rho}_k = (\hat{\rho}_{k1}, \dots, \hat{\rho}_{kR})'$  dan  $\hat{\rho}_{k'} = (\hat{\rho}_{k'1}, \dots, \hat{\rho}_{k'R})'$ , perbedaan diukur dengan menggunakan ukuran jarak Euclidean kuadrat berikut:

$$d_{kk'}^2 = \sum_{r=1}^R (\hat{\rho}_{kr} - \hat{\rho}_{k'r})^2 = \|\hat{\rho}_k - \hat{\rho}_{k'}\|^2$$

$r = 1, \dots, R$  adalah *time lag* untuk deret waktu  $x_k$  dan  $x_k'$ .

- 2) *Dynamic Time Warping* (DTW).

Misalkan terdapat dua deret waktu  $Q = q_1, q_2, \dots, q_i, \dots, q_n$  dan  $C = c_1, c_2, \dots, c_j, \dots, c_m$ . *Local distance matrix* berukuran  $(n \times m)$  yang dibentuk dari dua deret waktu  $Q$  dan  $C$ . Elemen dalam matriks jarak tersebut adalah elemen  $D(i, j)$  yaitu seluruh kemungkinan jarak antar kedua deret waktu kemudian akan dipilih nilai kumulatif minimum seperti pada persamaan:

$$D_{i,j} = ED(q_i, c_j) + \min\{D_{(i-1)(j-1)}, D_{(i-1)j}, D_{i(j-1)}\}; ED(q_i, c_j) = (q_i - c_j)^2$$

$D_{i,j}$  adalah setiap elemen  $(i, j)$  dari matriks jarak lokal yang merupakan penjumlahan antara kuadrat jarak  $q_i$  dan  $c_j$  dan jarak kumulatif minimum dari ketiga elemen yang berdekatan dengan elemen  $(i, j)$ , di

mana  $0 < i \leq n$  dan  $0 < j \leq m$ . Matriks jarak  $D$  yang terbentuk akan digunakan untuk mencari *warping path* optimal. Jarak DTW diperoleh dari: (Niennattrakul & Ratanamahatana, 2007)

$$DTW_{(Q,C)} = \min_{w \in P} \left\{ \sqrt{\sum_{k=1}^K d_{wk}} \right\}$$

$P$  adalah himpunan dari seluruh *warping path* yang mungkin, dan  $w_k$  adalah  $(i, j)$  pada elemen ke- $k$  dari *warping path* dan  $K$  adalah panjang *warping path*.

- d. Melakukan penggerombolan berhirarki dengan metode pautan rata-rata (*average linkage*). Adapun perhitungan metode pautan rata-rata sebagai berikut (Rencher, 2005):

$$d_{(uv)w} = \frac{\sum_i \sum_k d_{ik}}{N_w N_{uv}}$$

$d_{ik}$  adalah jarak antara objek ke- $i$  dalam gerombol ( $UV$ ) dan objek ke- $k$  dalam gerombol ke  $W$ ,  $N_{uv}$  adalah jumlah dari anggota pada kelompok  $UV$ ,  $N_w$  adalah jumlah dari anggota pada kelompok  $W$ , dan  $d_{(uv)w}$  adalah jarak  $UV$  dengan  $W$ .

- e. Mengevaluasi ukuran ketidakmiripan dengan menghitung nilai korelasi *cophenetic*. Nilai koefisien korelasi *cophenetic* semakin besar menandakan semakin baik ukuran ketidakmiripan yang digunakan.
- f. Menentukan banyaknya gerombol optimum dengan menghitung koefisien *silhouette*. Nilai koefisien *silhouette* yang semakin tinggi menunjukkan kualitas penggerombolan yang baik. Formula koefisien *silhouette* adalah:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

$a(i)$  adalah rata-rata ketidakmiripan setiap objek ke- $i$  dengan semua objek lainnya dalam gerombol A,  $b(i)$  adalah rata-rata ketidakmiripan setiap objek ke- $i$  dengan semua objek lainnya dalam gerombol B jika gerombol A dianggap tidak ada.

Tabel 1. Kategori penggerombolan berdasarkan koefisien *silhouette*.

Koefisien <i>silhouette</i>	Kategori Penggerombolan
0,71–1,00	<i>Strong Cluster</i>
0,51–0,70	<i>Good Cluster</i>
0,26–0,50	<i>Weak Cluster</i>
0,00–0,25	<i>Bad Cluster</i>

- g. Mengidentifikasi anggota gerombol pada masing-masing  $k$  yang terpilih.
4. Pembentukan model level gerombol dengan model ARIMA dengan tahapan sebagai berikut:
    - a. Menghitung data perwakilan setiap gerombol (*prototype*) menggunakan median.
    - b. Membagi *prototype* menjadi data latih dan data uji. Data tersebut akan digunakan untuk keperluan evaluasi hasil peramalan. Peramalan dilakukan untuk setiap *fold* yang dengan rincian periode pada Tabel 2.

Tabel 2. Pembagian data latih dan data uji.

Fold	Data Latih	Data Uji
1	Januari 2019 – Desember 2021	Januari – Maret 2022
2	Januari 2019 – Maret 2022	April – Juni 2022
3	Januari 2019 – Juni 2022	Juli – September 2022
4	Januari 2019 – September 2022	Oktober – Desember 2022

- c. Pemeriksaan kestasioneran data setiap gerombol menggunakan plot ACF dan uji ADF.
- d. Melakukan transformasi Box-Cox jika terdapat data yang tidak stasioner dalam ragam dan proses *differencing* jika terdapat data yang tidak stasioner dalam nilai tengah, hingga diperoleh data yang stasioner.
- e. Identifikasi beberapa model tentatif dengan melihat plot ACF dan PACF untuk memperoleh ordo p,d, dan q mengacu pada Tabel 3 (Montgomery et al., 2015).

Tabel 3. Identifikasi ACF dan PACF.

Model	ACF	PACF
AR[p]	<i>Tails off</i>	<i>Cuts off</i> setelah lag p
MA[q]	<i>Cuts off</i> setelah lag q	<i>Tails off</i>
ARMA[p,q]	<i>Tails off</i>	<i>Tails off</i>
ARIMA[p,d,q]	<i>Tails off</i> dan <i>differencing</i>	<i>Tails off</i> dan <i>differencing</i>

Montgomery et al. (2008) menjelaskan bentuk umum model ARIMA(p,d,q) adalah:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d y_t = c + (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t$$

$y_t$  adalah amatan pada waktu ke-t,  $\phi_p$  koefisien AR ordo p,  $\theta_q$  koefisien MA ordo q,  $\varepsilon_t$  sisaan waktu ke-t,  $c$  konstanta,  $p$  panjang ordo AR,  $q$  panjang ordo MA,  $d$  panjang ordo *differencing*, dan  $B$  operator *backshift*.

- f. Pendugaan parameter menggunakan *maximum likelihood estimation (MLE)* dengan memaksimalkan fungsi *log-likelihood* bersyarat bagi model ARMA(p,q) yaitu (Wei, 2006):

$$\ln L_*(\phi, \mu, \theta, \sigma_a^2) = -\frac{n}{2} \ln 2\pi\sigma_a^2 - \frac{S_*(\phi, \mu, \theta)}{2\sigma_a^2}$$

$S_*(\phi, \mu, \theta)$  adalah fungsi jumlah kuadrat bersyarat. penduga  $\hat{\phi}$ ,  $\hat{\mu}$ , dan  $\hat{\theta}$  adalah nilai yang memaksimalkan penduga dan diperoleh dengan meminimumkan  $S_*(\phi, \mu, \theta)$ .

- g. Pengujian diagnostik sisaan. Model yang baik adalah model yang memiliki sisaan saling bebas dengan uji Ljung-Box serta menyebar normal dengan uji *Kolmogorov-Smirnov* (Montgomery et al., 2015).
- h. Melakukan *overfitting* dengan menambahkan satu parameter AR dan MA secara bergantian pada ordo hasil identifikasi.
- i. Penentuan tiga kandidat model dari seluruh model tentatif dan *overfitted* model yang mempunyai *Akaike Information Criterion (AIC)* terkecil.
- j. Evaluasi performa model pada setiap *fold* menggunakan *mean absolute percentage error (MAPE)* di setiap gerombol. Performa peramalan setiap model secara keseluruhan diperoleh dari nilai rata-rata MAPE dari seluruh *fold*. Nilai MAPE didefinisikan sebagai:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_t - \hat{x}_t}{x_t} \right| \times 100$$

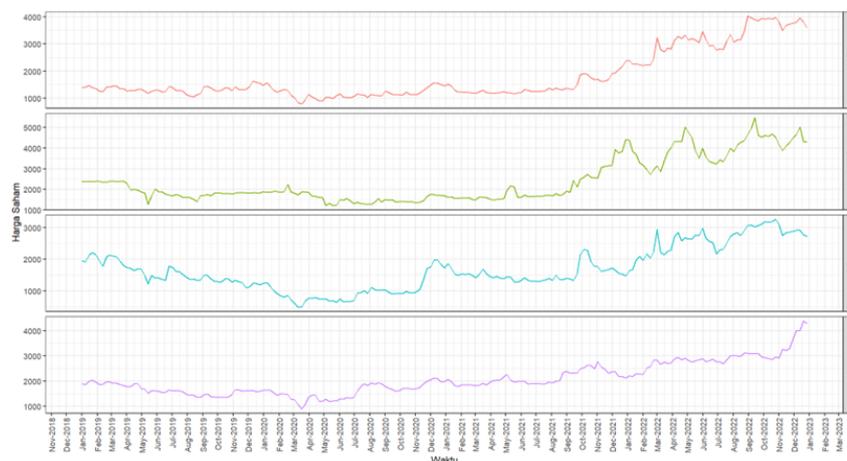
$x_t$  adalah aktual pada periode ke-t dan  $\hat{x}_t$  adalah data prediksi pada periode ke-t.

- k. Menentukan model terbaik yang akan digunakan dalam peramalan periode mendatang yaitu model yang memiliki nilai rata-rata MAPE terkecil.
  - l. Melakukan peramalan periode mendatang pada setiap gerombol menggunakan model terbaik.
5. Mendeskripsikan gerombol dan hasil peramalan yang dihasilkan.

### 3. Hasil dan Pembahasan

#### 3.1 Eksplorasi Data

Terdapat 77 emiten sektor energi yang terdaftar di BEI pada akhir tahun dan 45 emiten terpilih karena memenuhi kriteria sesuai dengan yang dipaparkan pada prosedur analisis data. Plot data deret waktu setiap emiten memiliki pergerakan harga yang berbeda-beda namun beberapa diantaranya memiliki pola pergerakan harga yang mirip seperti ADRO, BSSR, INDY, dan PTRO. Kemiripan pola tersebut menjadi dasar bahwa proses penggerombolan dapat dilakukan untuk mengelompokkan emiten yang memiliki pola pergerakan harga saham serupa. Gambar 1 menunjukkan harga empat emiten tersebut cenderung konstan pada periode Januari 2019 hingga Februari 2020, tren menurun pada periode Maret hingga Mei 2020, dan memiliki tren meningkat secara signifikan pada periode 2021 hingga akhir 2022. Penurunan harga di bulan Maret 2020 merupakan reaksi investor atas kasus COVID-19 pertama di Indonesia.



Gambar 1. Plot harga saham dari emiten ADRO, BSSR, INDY, dan PTRO periode 1 Januari 2019 hingga 30 Desember 2022.

Rata-rata harga saham sektor energi sebesar Rp1.956,06. Terdapat tiga emiten dengan harga yang sangat jauh di atas nilai rata-rata sehingga dikategorikan sebagai *high-priced stocks* yaitu DSSA dengan rata-rata harga Rp23.342,10, ITMG dengan rata-rata harga Rp18.977,39, dan TCPI dengan rata-rata harga Rp7.512,51. Hal ini menjadi pertimbangan keikutsertaan *high-priced stocks* dalam proses penggerombolan karena harga tinggi yang terlalu mendominasi menyebabkan pola harga dari 42 emiten lainnya sulit dianalisis.

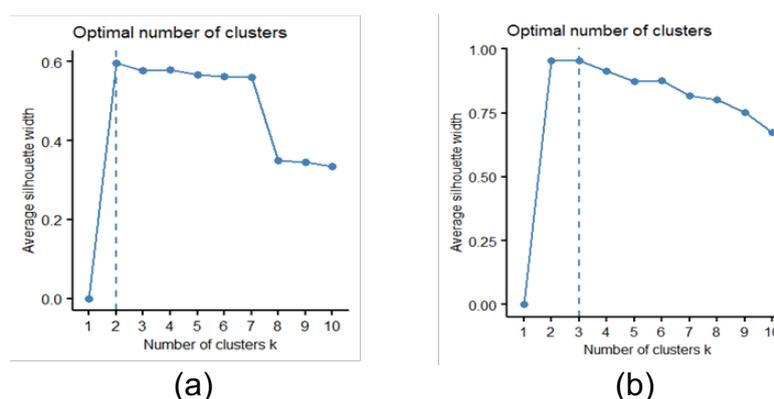
### 3.2 Penggerombolan Data Deret Waktu

Proses penggerombolan menggunakan data harga penutupan periode mingguan yang telah stasioner. Data stasioner diperoleh setelah dilakukan differencing satu kali dan dibuktikan dengan hasil uji ADF menghasilkan nilai p lebih kecil dari taraf nyata 5%. Teknik penggerombolan yang digunakan yaitu penggerombolan berhierarki dengan pautan rata-rata (*average linkage*). Penelitian ini menggunakan dua ukuran ketidakmiripan yaitu jarak autokorelasi dan jarak DTW. Perbandingan nilai korelasi *cophenetic* dari kedua jarak yang digunakan terdapat pada Tabel 4.

Tabel 4. Perbandingan nilai korelasi *cophenetic* memuat dan tanpa *high-priced stocks*.

Jarak	Memuat <i>high-priced stocks</i>	Tanpa <i>high-priced stocks</i>
<i>Autocorrelation-based</i>	0,798	0,790
<i>Dynamic time warping (DTW)</i>	0,975	0,830

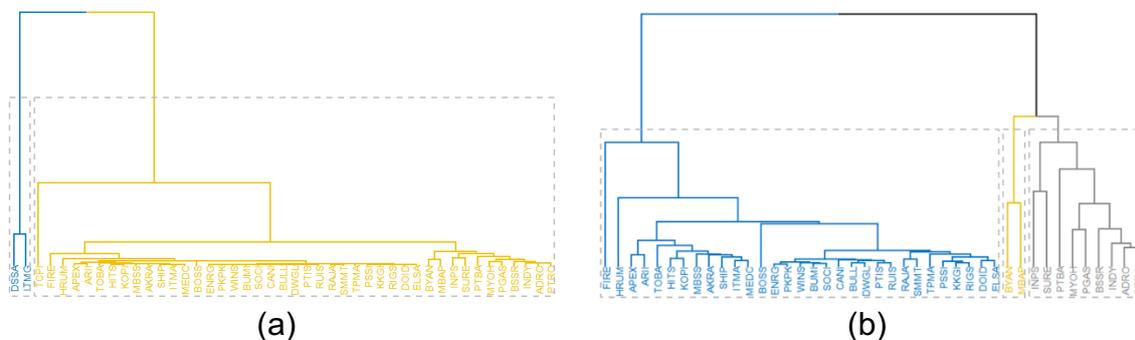
Mengacu pada hasil eksplorasi data, proses penggerombolan mempertimbangkan keikutsertaan *high-priced stocks*. Korelasi *cophenetic* dihitung untuk melihat ketepatan dan kualitas dari hasil analisis penggerombolan dan diperoleh hasil perbandingan yang disajikan pada Tabel 4. Penggerombolan menggunakan jarak DTW menghasilkan korelasi *cophenetic* lebih besar dibandingkan jarak autokorelasi baik memuat atau tanpa *high-priced stocks*. Hal tersebut mengindikasikan dendrogram yang dihasilkan oleh jarak DTW lebih baik dalam merepresentasikan setiap objek. Tahapan penggerombolan selanjutnya hanya akan menggunakan jarak yang memiliki nilai korelasi *cophenetic* terbesar yaitu jarak DTW. Jarak DTW memuat *high-priced stocks* menghasilkan nilai *cophenetic* 0,975 bernilai lebih besar dibandingkan tidak memuat *high-priced stocks* yaitu sebesar 0,830. Namun, penentuan jumlah gerombol optimal ditentukan menggunakan interpretasi subjektif dari nilai maksimum koefisien *silhouette* yang mempertimbangkan setiap kemungkinan jumlah gerombol memuat dan tanpa *high-priced stocks* (disajikan pada plot di Gambar 2).



Gambar 2. Plot koefisien *silhouette* penggerombolan dengan jarak DTW (a) kondisi memuat *high-priced stocks* (b) kondisi tanpa *high-priced stocks*.

Dendrogram dengan banyak gerombol  $k = 2$  pada Gambar 3(a) menunjukkan *high-priced stocks* cenderung terpisah membentuk satu gerombol karena harga tinggi yang menyebabkan penyebaran emiten lainnya tidak terlihat. Oleh karena itu, 42 emiten lainnya (selain *high-priced stocks*) digerombolkan lagi menjadi tiga gerombol mengacu

pada plot koefisien *silhouette* pada Gambar 2(b). Gerombol pada dendrogram Gambar 3(a) dan 3(b) memiliki urutan anggota yang sama dan sehingga penggerombolan 42 emiten lainnya dianggap dapat memperbaiki penyebaran pada gerombol. Secara keseluruhan penggerombolan emiten sektor energi menggunakan jarak DTW menghasilkan empat gerombol yang memiliki nilai rata-rata dan pola pergerakan harga yang berbeda.



Gambar 3. Perbandingan dendrogram hasil penggerombolan (a) mencakup *high-priced stocks* (b) tanpa *high-priced stocks*.

Anggota setiap gerombol terdapat pada Tabel 5. Gerombol A terdiri dari tiga *high-priced stocks* yaitu emiten DSSA, ITMG, dan TCPI terbentuk karena tiga emiten tersebut memiliki rentang harga yang jauh berbeda dengan emiten lainnya. Secara umum, harga saham gerombol A memiliki tren meningkat dalam jangka panjang. Saham kategori ini mempunyai likuiditas yang baik dan mempunyai potensi keuntungan yang lebih besar karena biasanya terkait dengan perusahaan yang memiliki performa yang baik serta memiliki potensi keuntungan yang lebih besar bagi investor. Namun, perlu biaya yang lebih tinggi dan mungkin telah mencapai puncaknya sehingga mengurangi potensi kenaikan harga.

Tabel 5. Rincian anggota setiap gerombol

Gerombol	Jumlah Anggota	Anggota
A	3	DSSA, ITMG, TCPI
B	9	ADRO, BSSR, INDY, INPS, MYOH, PGAS, PTBA, PTRO, SURE
C	31	AKRA, APEX, ARII, BOSS, BULL, BUMI, CANI, DOID, DWGL, ELSA, ENRG, FIRE, HITS, HRUM, ITMA, KEGI, KOPI, MBSS, MEDC, PKPK, PSSI, PTIS, RAJA, RIGS, RUIS, SHIP, SMMT, SOCI, TOBA, TPMA, WINS
D	2	BYAN, MBAP

Harga saham seluruh anggota gerombol B berfluktuasi di sekitar rata-rata sehingga dikategorikan sebagai *medium-priced stocks*. Sebagian besar anggota gerombol B memiliki pola pergerakan harga yang serupa, yaitu memiliki tren menurun saat sebelum pandemi hingga mencapai harga terendah saat COVID-19 pertama diumumkan di Indonesia dan berfluktuasi dengan tren meningkat pada periode setelahnya. Harga saham dari mayoritas anggota gerombol C berfluktuasi di bawah rata-rata sehingga dikategorikan sebagai *low-priced stocks*. Sedangkan gerombol D hanya berisi dua emiten yaitu BYAN dan MBAP dengan pola pergerakan harga saham

yang serupa. Harga saham dari anggota gerombol D memiliki pola harga yang berbeda dibanding anggota gerombol lainnya. Harga saham anggota gerombol D cenderung konstan berfluktuasi di sekitar rata-rata pada periode 2019 hingga akhir 2021. Namun, mempunyai tren meningkat di atas rata-rata sepanjang tahun 2022.

### 3.3 Pemodelan Level Gerombol

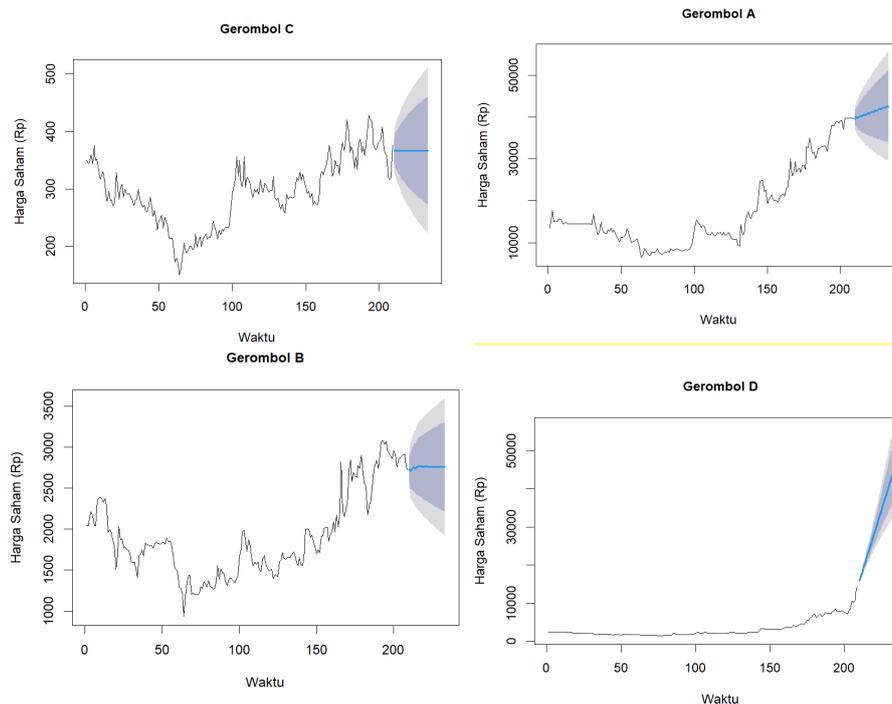
Pemodelan pada level gerombol menggunakan data perwakilan dari setiap gerombol (*prototype*). *Prototype* menggunakan nilai median karena lebih *robust* terhadap pencilan. Identifikasi model ARIMA(p,d,q) didasarkan pada plot ACF dan PACF dari *prototype* yang telah stasioner. Model pada setiap gerombol yang memiliki nilai AIC terkecil dan signifikan pada setiap dugaan parameternya dipilih untuk dilihat performa peramalannya pada setiap *fold* menggunakan nilai MAPE. Nilai MAPE diperoleh dengan menggunakan *rolling cross-validation*. Data dibagi menjadi empat *fold*, dan evaluasi performa model dilakukan pada masing-masing *fold*. MAPE pada setiap *fold* dihitung, dan MAPE untuk setiap model diperoleh dari rata-rata dari MAPE pada setiap *fold*. Semakin kecil nilai MAPE maka semakin baik performa model dalam melakukan peramalan. Setiap model dipilih untuk peramalan harga saham periode mendatang pada setiap gerombol. Rata-rata nilai MAPE dari model terbaik yang terpilih pada setiap gerombol terdapat pada tabel 6. Model terbaik tersebut telah memenuhi asumsi kebebasan sisaan berdasarkan hasil uji Ljung-box yang menunjukkan nilai p lebih besar dari taraf nyata 5%, juga memenuhi asumsi kenormalan sisaan.

Tabel 6: Nilai MAPE dari model terbaik setiap gerombol.

<i>Fold</i>	Gerombol A ARIMA(3,1,3)	Gerombol B ARIMA(6,1,3)	Gerombol C ARIMA(1,1,0)	Gerombol D ARIMA(4,2,0)
1	18,342%	9,919 %	16,576%	17,999%
2	10,372%	15,531 %	6,932%	11,175%
3	11,142%	16,555 %	10,687%	3,185%
4	1,135%	3,283 %	7,679%	14,409%
Rata-rata	10,248%	11,322%	10,469%	11,692%

### 3.4 Peramalan Periode Mendatang

Plot peramalan harga saham setiap gerombol enam bulan kedepan dapat dilihat pada Gambar 4. Hasil peramalan harga saham setiap gerombol seluruhnya menunjukkan potensi kenaikan harga untuk enam bulan kedepan namun dengan besar persentase yang berbeda. Investor diperkirakan dapat memperoleh keuntungan dengan memetakan investasinya ke anggota gerombol manapun. Meskipun demikian, terdapat kemungkinan penurunan ataupun kenaikan harga saham yang dalam selang kepercayaan 95% yang terdapat pada plot ramalan harga. Gerombol A dan D diduga dapat menghasilkan keuntungan lebih besar dari *capital gain* karena plot peramalan menunjukkan tren harga meningkat yang signifikan. Selain itu, bisa menjadi pilihan untuk investasi jangka panjang. Emiten yang memiliki harga saham yang tinggi biasanya adalah perusahaan yang memiliki performa keuangan yang baik, fundamental perusahaan yang stabil, dan prospek pertumbuhan yang cerah dalam jangka panjang. Perusahaan yang memiliki harga saham yang tinggi umumnya memiliki kebijakan dividen yang baik sehingga investor dapat memperoleh keuntungan maksimal dari investasi mereka.



Gambar 4. Plot ramalan *prototype* periode Januari – Juni 2023 setiap gerombol

#### 4. Simpulan

Jarak DTW memiliki kinerja yang lebih baik dalam proses penggerombolan emiten sektor energi dibandingkan dengan jarak *autocorrelation-based*. Melalui pendekatan analisis gerombol berhierarki dengan jarak DTW, emiten sektor energi dikelompokkan menjadi empat gerombol dengan kategori *strong cluster* karena nilai koefisien *silhouette* lebih dari 0,71. Model ARIMA level gerombol pada setiap gerombol memiliki nilai MAPE yang berkisar antara 10-20% sehingga dikategorikan mampu meramalkan harga saham sektor energi dengan baik. Secara umum, kecenderungan pola pergerakan ramalan harga saham sektor energi membentuk pola tren meningkat. Gerombol A dan D bisa menjadi rekomendasi investasi yang potensial. Gerombol A dan D merupakan gerombol dengan potensi keuntungan tertinggi dari *capital gain* berdasarkan hasil peramalan harga saham. Gerombol A dan D juga berisi perusahaan dengan fundamental perusahaan yang stabil, serta prospek pertumbuhan dan kebijakan dividen yang baik.

#### Daftar Pustaka

- Aghabozorgi, S., Ying Wah, T., Herawan, T., Jalab, H. A., Shaygan, M. A., & Jalali, A. (2014). A Hybrid Algorithm for Clustering of Time Series Data Based on Affinity Search Technique. *The Scientific World Journal*, 2014, 1–12. <https://doi.org/10.1155/2014/562194>
- Al Fatah, R. S. (2022). *Pengelompokan Emiten Saham Consumer non Cyclical Menggunakan Jarak Euclidean dan Dynamic Time Warping*. IPB University.
- [BEI]. (2022). 13 Daftar Saham Sektor Energi yang Bisa Jadi Investasi Anda [News]. *IDX Channel*. <https://www.idxchannel.com/market-news/13-daftar-saham-sektor-energi-yang-bisa-jadi-investasi-anda> [diakses pada 4 Januari 2023]

- D'Urso, P., & Maharaj, E. A. (2009). Autocorrelation-based fuzzy clustering of time series. *Fuzzy Sets and Systems*, 160(24), 3565–3589. <https://doi.org/10.1016/j.fss.2009.04.013>
- Kardiyan, F., & Güney, H. (2015). A Study on Time Series Clustering. *Gazi University Journal of Science*, 28(2), 331–347.
- Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons.
- Mattjik, A. A., & Sumertajaya, I. M. (2011). *Sidik Peubah Ganda dengan Menggunakan SAS*. IPB Press.
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons.
- Niennattrakul, V., & Ratanamahatana, C. A. (2007). On Clustering Multimedia Time Series Data Using K-Means and Dynamic Time Warping. *2007 International Conference on Multimedia and Ubiquitous Engineering (MUE'07)*, 733–738. <https://doi.org/10.1109/MUE.2007.165>
- Rencher, A. C. (2005). A Review Of “Methods of Multivariate Analysis, Second Edition.” *IIE Transactions*, 37(11), 1083–1085. <https://doi.org/10.1080/07408170500232784>
- Sardá-Espinosa, A. (2017). Comparing time-series clustering algorithms in r using the dtwclust package. *R Package Vignette*, 12, 41.
- Ulinuha, M., M Afendi, F., & Sumertajaya, I. M. (2022). Study of Clustering Time Series Forecasting Model for Provincial Grouping in Indonesia Based on Rice Price: Kajian Model Peramalan Clustering Time Series untuk Penggerombolan Provinsi Indonesia berdasarkan Harga Beras. *Indonesian Journal of Statistics and Its Applications*, 6(1), 50–62. <https://doi.org/10.29244/ijsa.v6i1p50-62>
- Wei, W. W. S. (2006). *Time series analysis: Univariate and multivariate methods* (2nd ed). Pearson Addison Wesley.