

Aplikasi Pengelompokan Data Runtun Waktu dengan Algoritma K-Medoids

Muhammad Aldani Zen¹, Sri Wahyuningsih^{1*}, dan Andrea Tri Rian Dani²

¹Laboratorium Statistika Ekonomi dan Bisnis, Program Studi Statistika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman, Samarinda, Kalimantan Timur, Indonesia

²Laboratorium Statistika Terapan, Program Studi Statistika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman, Samarinda, Kalimantan Timur, Indonesia

*Corresponding author: swahyuningsih@fmipa.unmul.ac.id

Received: 28 January 2023

Revised: 16 September 2023

Accepted: 19 September 2023

ABSTRAK – Teknologi informasi yang terus berkembang selalu diiringi dengan informasi digital yang secara terus menerus tersimpan dan terkumpul dalam jumlah yang sangat besar. Pemilihan suatu algoritma yang sesuai menjadi sangat penting dalam menangani permasalahan *big data* salah satunya dengan menggunakan analisis kluster. Analisis kluster merupakan serangkaian metode yang digunakan untuk mempartisi sekumpulan objek pengamatan ke dalam suatu himpunan. Himpunan-himpunan yang terbentuk disebut dengan kluster. Analisis kluster dapat diimplementasikan pada data runtun waktu, di mana dalam pemrosesan data sedikit berbeda dengan tipe data *cross-section*. *Time series clustering* merupakan salah satu pendekatan yang dapat digunakan dalam pemrosesan data runtun waktu multivariabel. Algoritma pengelompokan yang digunakan pada *time series clustering* adalah *K-Medoids*. Tujuan penelitian ini dilakukan untuk memperoleh nilai *K* optimal dalam menentukan banyaknya kluster berdasarkan koefisien *silhouette* dan memperoleh hasil pengelompokan menggunakan algoritma *K-Medoids*. Pada penelitian ini, ukuran kemiripan yang digunakan adalah jarak *dynamic time warping*. Data pada penelitian adalah data harga minyak goreng 34 provinsi di Indonesia dari Oktober 2017 – Oktober 2022. Berdasarkan hasil analisis, diperoleh nilai *K* optimal sebesar 2 kluster dengan 19 provinsi bergabung pada kluster 1 di mana kluster dengan harga minyak goreng dibawah kluster 2 dan 15 provinsi bergabung pada kluster 2 yang merupakan kluster dengan harga minyak goreng tertinggi.

Kata kunci– *K-Medoids*, Minyak Goreng, Koefisien *Silhouette*, *Time Series Clustering*.

ABSTRACT – The development of information technology will always be accompanied by the storage and accumulation of massive quantities of digital information. Cluster analysis is one of many data processing problems that require the selection of an appropriate algorithm when dealing with large data sets. Cluster analysis is a collection of techniques for dividing a set of observation objects into clusters. Cluster analysis is applicable to time series data, the processing of which differs slightly from that of cross-section data. Clustering time series is a technique for processing multivariable time series data. *K-Medoids* is the clustering algorithm used for time series clustering. The objective of this study is to obtain optimal *K*-values in determining the number of clusters based on silhouette coefficients and grouping outcomes using the *K-Medoids* algorithm. In this study, the dynamic time-warping distance is utilized as the similarity metric. This study provides cooking oil price data for 34 Indonesian provinces from October 2017 to October 2022. The optimal *K* value is determined for two clusters based on the results of the analysis, with 19 provinces joining cluster 1, where the cluster with cooking oil prices was below cluster 2 and 15 provinces joining cluster 2 which is the cluster with the highest cooking oil prices.

Keywords– *Cooking Oil*, *K-Medoids*, *Silhouette Coefficient*, *Time Series Clustering*.

I. PENDAHULUAN

Seiring berkembangnya kemajuan teknologi yang semakin canggih, masyarakat dituntut untuk bertransformasi menjadi serba digital. Teknologi informasi merupakan salah satu perkembangan teknologi yang cukup pesat. Informasi digital yang diterima baik terstruktur maupun tidak terstruktur akan secara terus menerus tersimpan dan terkumpul dengan jumlah yang sangat besar. Penggunaan pembelajaran *data mining* banyak digunakan dalam membantu pengelolaan data yang cukup besar.

Data mining merupakan salah satu metode untuk mendapatkan wawasan yang menarik dari kumpulan basis data dengan ukuran yang sangat besar. Proses dari *data mining* yaitu mengekstraksi dan mengidentifikasi wawasan tersembunyi yang tersimpan dalam basis data besar (*big data*) sehingga memiliki wawasan baru yang dapat dipahami dengan mudah [1]. Konsep dari *big data* cenderung memiliki sifat kompleksitas yang tinggi karena struktur dari data besar beragam dan menjadi tantangan dalam pemrosesan data [2]. *Preprocessing data* menjadi sangat penting dalam menangani permasalahan *big data*, salah satunya dengan menggunakan analisis kluster [3].

Konsep dari analisis kluster adalah mempartisi sekumpulan objek pengamatan ke dalam suatu himpunan. Himpunan-himpunan yang terbentuk disebut dengan kluster. Objek-objek di dalam kluster tersebut adalah objek yang mempunyai kemiripan pada satu kluster namun tidak ada keterkaitan dengan objek yang berada pada kluster lainnya [4]. Algoritma pengelompokan tidak hanya spesifik pada tipe data *cross-section*, data runtun waktu juga dapat diimplementasikan pada analisis kluster dengan algoritma yang sedikit berbeda dengan data *cross-section*. *Time series clustering* adalah salah satu pendekatan yang dapat digunakan dalam pemrosesan data runtun waktu multivariabel [5]. Konsep dari *time series clustering* adalah mengelompokkan berdasarkan kemiripan dari pola masing-masing runtun waktu, sehingga objek runtun waktu yang memiliki kemiripan pola yang sama akan cenderung membentuk suatu himpunan [6]. Teknik *time series clustering* dapat diimplementasikan untuk menganalisis masalah-masalah dalam bidang sektor ekonomi sebagai perancangan kebijakan pemerintah dalam mengatasi masalah-masalah yang terjadi. Ketidakstabilan harga bahan pokok merupakan salah satu permasalahan yang dihadapi saat ini dalam sektor ekonomi khususnya pada minyak goreng.

Minyak goreng merupakan salah satu komoditas pokok pangan yang paling banyak dibutuhkan di Indonesia. Tingkat konsumsi minyak goreng yang tinggi akan menyebabkan peningkatan permintaan bahan pokok tersebut setiap

tahunnya. Berdasarkan data yang disajikan oleh Pusat Informasi Harga Pangan Strategis Nasional per 26 Agustus 2022, harga minyak goreng di pasar tradisional masih tergolong tinggi hingga mencapai Rp 20.200 per liter. Puncak harga minyak goreng tertinggi pernah mencapai hingga Rp 24.650 per liter pada tanggal 22 April 2022. Terbatasnya persediaan pasar nasional dan permintaan dari pasar internasional yang semakin meningkat menyebabkan harga minyak goreng melambung tinggi. Pengendalian harga minyak goreng menjadi hal yang mendesak dalam menjaga stabilitas ekonomi di Indonesia.

Beberapa penelitian terdahulu yang mengkaji tentang pendekatan *time series clustering* diantaranya adalah [7] melakukan pengelompokan data kasus aktif COVID-19 menggunakan algoritma *agglomerative*. Penelitian yang dilakukan [8] dan [9] juga melakukan pengelompokan data kasus aktif COVID-19 dengan mengimplementasikan algoritma *K-Medoids*. Penelitian lain dilakukan oleh [10] mengelompokkan data harga minyak goreng di Indonesia menggunakan *agglomerative*.

Berdasarkan latar belakang serta pendekatan *time series clustering* yang belum pernah diaplikasikan pada data harga minyak goreng, dilakukan penelitian menggunakan pendekatan *time series clustering* pada proses pengelompokan provinsi di Indonesia yang diaplikasikan pada data harga minyak goreng dari bulan Oktober 2017 – Oktober 2022. Algoritma pengelompokan yang digunakan adalah *K-Medoids* dengan mengimplementasikan jarak *Dynamic Time Warping* (DTW). Penentuan jumlah kluster optimal menggunakan koefisien *silhouette* [6]. Tujuan penelitian ini adalah memperoleh nilai K optimal berdasarkan koefisien *silhouette* dan memperoleh hasil pengelompokan menggunakan algoritma *K-Medoids*.

II. TINJAUAN PUSTAKA

A. Analisis Kluster

Analisis kluster adalah salah satu teknik dalam *data mining* yang diimplementasikan untuk proses penggerombolan himpunan objek ke dalam suatu kluster hingga objek-objek yang memiliki persamaan karakteristik tertinggi akan bergabung di suatu kluster tersebut, namun memiliki ketidakmiripan yang tinggi dengan objek-objek yang bergabung di kluster lainnya [4]. Tujuan digunakannya analisis kluster adalah agar objek-objek yang terhimpun dalam sebuah kluster merupakan objek-objek yang ada keterkaitan satu sama lain dan tidak ada keterkaitan dengan objek di kluster yang lain [11].

Analisis kluster berdasarkan strukturnya terdapat dua algoritma pengelompokan, yaitu kluster hirarki (*hierarchical clustering*) dan kluster non-hirarki (*non-hierarchical clustering*). Perbedaan yang jelas dari kedua metode adalah proses penentuan jumlah kluster. Pada kluster non-hirarki dalam penentuan jumlah kluster ditentukan terlebih dahulu sesuai yang diinginkan, sedangkan pada kluster hirarki dilakukan diakhir dalam penentuan jumlah kelompok menyesuaikan pada tingkat yang sesuai untuk mengelompokkan objek menjadi beberapa kelompok [11].

B. Standarisasi Data

Teknik standarisasi sangat penting dalam proses analisis kluster. Nilai-nilai atribut data yang berbeda-beda rentangnya seringkali perlu dilakukan standarisasi agar proses pengelompokan tidak bias. Standarisasi akan membuat atribut memiliki bobot yang sama sehingga memperkecil perbedaan antar kelompok. Salah satu teknik standarisasi yang dapat digunakan adalah standarisasi *z-Score* di mana metode tersebut sering disebut pula dengan *zero-mean* karena dalam prosesnya atribut diubah ke dalam bentuk data standar berdasarkan nilai rata-rata sama dengan nol serta deviasi standar sama dengan satu [1].

Misalkan terdapat data runtun waktu $Z_t = Z_1, Z_2, Z_3, \dots, Z_n$ dan dilakukan proses standarisasi dengan menghitung nilai rata-rata dan deviasi standar. Formulasi dari rata-rata runtun waktu dituliskan pada Persamaan (1).

$$\bar{Z} = \left(\frac{1}{n} \right) \sum_{t=1}^n Z_t \quad (1)$$

sedangkan untuk deviasi standarnya dituliskan pada Persamaan (2).

$$S_z = \sqrt{\frac{1}{n-1} \sum_{t=1}^n (Z_t - \bar{Z})^2} \quad (2)$$

standarisasi data runtun waktu Z_t dituliskan pada Persamaan (3).

$$\tilde{Z}_t = \frac{Z_t - \bar{Z}}{S_z} \quad (3)$$

dengan Z_t adalah data runtun waktu Z untuk periode waktu ke- t , n adalah banyaknya periode waktu, \bar{Z} adalah rata-rata dari Z_t , S_z adalah deviasi standar dari Z_t , dan \tilde{Z}_t adalah standarisasi data dari Z_t .

C. Dynamic Time Warping

Dynamic time warping (DTW) adalah salah satu ukuran kemiripan yang terkenal dalam mengukur jarak antar objek runtun waktu. Kelebihan dari DTW dapat mengukur objek runtun waktu dengan periode yang tidak sama panjangnya. Konsep utama dari DTW merupakan proses yang digunakan antara data runtun waktu untuk menemukan *warping path* optimal dengan panjang yang berbeda. Algoritma ini dapat diukur kedekatan dari set data runtun waktu yang ada.

Proses DTW menerapkan teknik pemrograman yang disusun dalam matriks jarak untuk memperoleh kombinasi jalur (*warping path*) yang mungkin dan akan dipilih salah satu jalur yang mempunyai jarak terkecil antara dua *set* data runtun waktu. Elemen yang terisi dalam matriks jarak adalah jarak kumulatif dari minimal tiga tetangga terdekatnya [12].

Misalkan ingin dihitung jarak DTW antara dua *set* data runtun waktu $Z(1)_t = Z_1, Z_2, Z_3, \dots, Z_i$ dan $Z(2)_t = Z_1, Z_2, Z_3, \dots, Z_j$ di mana $i = 1, 2, 3, \dots, n$ dan $j = 1, 2, 3, \dots, m$ akan membentuk sebuah matriks yang disebut dengan matriks **A** berukuran $p \times q$ dengan elemen ke- (i, j) adalah selisih antara $Z(1)_i$ dengan $Z(2)_j$, lalu ditambah dengan tiga elemen yang berdekatan $\{a_{(i-1)(j-1)}, a_{(i-1)j}, a_{i(j-1)}\}$, dengan $0 < i \leq p$ dan $0 < j \leq q$. Matriks **A** untuk elemen ke- i dan ke- j adalah sebagai berikut:

$$a_{ij} = w_{ij} + \min \{a_{(i-1)(j-1)}, a_{(i-1)j}, a_{i(j-1)}\} \quad (4)$$

dengan w_{ij} adalah nilai mutlak dari selisih antara $Z(1)_i$ dengan $Z(2)_j$. Jarak DTW dari kedua set runtun waktu tersebut adalah jarak minimum dari *warping path* yang mungkin dan dapat dirumuskan sebagai

$$d_{DTW}(Z(1)_i, Z(2)_j) = \min_{\forall W \in P} \left(\sqrt{\sum_{i,j}^s a_{ij}} \right) \quad (5)$$

dengan P merupakan kombinasi dari seluruh kemungkinan *warping path*, a_{ij} merupakan elemen (i, j) pada *warping path*, dan s merupakan panjang dari *warping path* [13].

D. Partitioning Around Method

Partitioning Around Method (PAM) atau yang lebih sering dikenal dengan *K-Medoids* adalah metode pengelompokan yang titik pusatnya menggunakan objek representatif (*medoids*). Objek lainnya akan bergabung dengan objek *medoids* yang memiliki kemiripan tertinggi. Ketika pemilihan awal *medoids* sebanyak K telah dilakukan, algoritma ini akan terus mencoba untuk mencari *medoids* yang dapat menginterpretasikan anggotanya dengan menganalisis semua kombinasi pasangan objek sampai diperoleh satu objek yang akan menjadi *medoids* dan objek lainnya yang akan menjadi anggota pada masing-masing *medoids*. Pilihan yang tepat dari titik setiap iterasi akan dipilih sebagai *medoids* untuk menganalisis pada proses iterasi selanjutnya dengan menghitung jumlah ketidakmiripan antara objek p ke objek *medoids*. Konsep dari metode partisi adalah meminimumkan jumlah ketidakmiripan antara setiap objek p dengan objek *medoids* terdekat, yaitu menggunakan jumlah kesalahan absolut. Kesalahan absolut didefinisikan pada Persamaan (6).

$$E = \sum_{i=1}^K \sum_{p \in C_i} dist(p, o_i) \quad (6)$$

di mana E adalah jumlah kesalahan absolut untuk seluruh objek p dalam kelompok data, dan o_i adalah objek *medoids* kluster C_i [1]. Kelemahan dari algoritma *K-Means* yaitu memiliki sensitivitas yang tinggi terhadap pencilan, sehingga *K-Medoids* muncul untuk mengatasi hal tersebut. Pencilan merupakan objek data yang menyimpang secara signifikan dari objek lainnya. Suatu objek dengan nilai yang berbeda dari distribusi data secara substansial akan menunjukkan penyimpangan dari distribusi data [14].

E. Koefisien Silhouette

Pada proses pengelompokan, ukuran seberapa baik suatu data dikelompokkan dapat diketahui dengan menggunakan koefisien *silhouette*. Formulasi ini memberikan informasi sesuai dengan jumlah kluster pada *dataset*. Tujuannya adalah mengevaluasi hasil pengelompokan dengan mengetahui seberapa tepat suatu objek bergabung ke dalam sebuah kluster [4]. Berikut merupakan tahapan serta formulasi untuk memperoleh nilai koefisien *silhouette*:

1. Menghitung rata-rata jarak dari objek ke- i dengan seluruh objek yang berada pada satu kluster.
2. Menghitung rata-rata jarak dari objek ke- i dengan seluruh objek yang berada pada kluster lainnya.
3. Menghitung nilai koefisien *silhouette* dengan formula pada Persamaan (7).

$$SC(i) = \frac{b_i - a_i}{\max \{a_i, b_i\}} \quad (7)$$

dengan a_i adalah rata-rata jarak dari objek ke- i dengan seluruh objek yang berada pada satu kluster dan b_i adalah rata-rata jarak dari objek ke- i dengan seluruh objek yang berada pada kluster lainnya. Koefisien *silhouette* berada pada rentang nilai $[-1, 1]$. Hasil pengelompokan dikatakan baik ketika nilai koefisien *silhouette* yang dimiliki sebesar 1, artinya objek ke- i sudah bergabung dalam kluster yang tepat. Sebaliknya ketika nilai koefisien *silhouette* sebesar -1, artinya struktur kluster yang dihasilkan tidak baik [15].

III. METODOLOGI

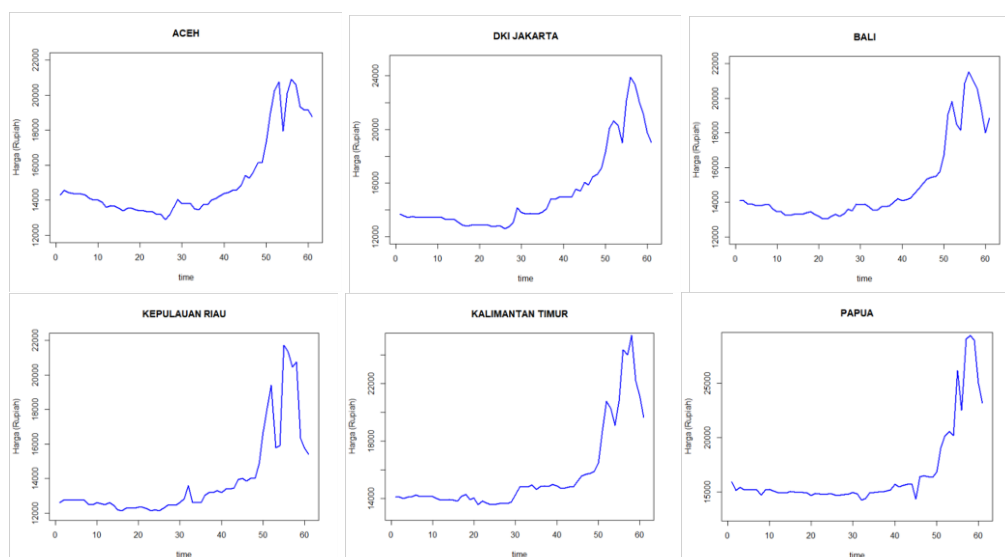
Penelitian ini menggunakan data harga minyak goreng 34 provinsi di Indonesia dari bulan Oktober 2017 sampai dengan Oktober 2022 yang disajikan pada Pusat Informasi Harga Pangan Nasional (<https://hargapangan.id/>). Tahapan-tahapan analisis data adalah sebagai berikut:

1. Eksplorasi data.
2. Melakukan standarisasi data.
3. Melakukan pengelompokan data dengan algoritma *K-Medoids* dengan tahapan sebagai berikut:

- a. Memilih sebanyak K objek sebagai *medoids*, dengan melihat nilai koefisien *silhouette* maksimum.
 - b. Menghitung jarak DTW setiap masing-masing objek ke *medoids* yang telah dipilih
 - c. Menetapkan objek ke kluster terdekat berdasarkan jarak ke *medoids* terkecil dan menghitung fungsi objektif yang merupakan total jarak dari seluruh objek ke objek *medoids* terdekat
 - d. Memilih salah satu objek dari setiap kluster sebagai kandidat *medoids* (*non-medoids*) yang baru
 - e. Menghitung kembali jarak DTW setiap masing-masing objek ke *non-medoids*.
 - f. Menetapkan setiap objek ke kluster yang sesuai berdasarkan jarak *non-medoids* terkecil dan menghitung fungsi objektif baru.
 - g. Menghitung simpangan (S) dengan menghitung selisih antara fungsi objektif baru dengan fungsi objektif lama. Jika nilai $S < 0$ dalam artian fungsi objektif baru memiliki total jarak yang lebih kecil, maka *medoids* awal diganti dengan *medoids* yang baru.
 - h. Mengulangi langkah (d) hingga langkah (g) sampai menghasilkan total jarak atau fungsi objektif yang minimum
4. Melakukan profilisasi dan menginterpretasikan hasil kluster

IV. HASIL DAN DISKUSI

Sebelum menganalisis lebih lanjut, diperlukan eksplorasi data berupa perkembangan harga minyak goreng dari bulan ke bulan untuk mengetahui karakteristik dan pola yang terbentuk dari data harga minyak goreng setiap provinsi. Adapun kondisi perkembangan harga minyak goreng di sebagian provinsi di Indonesia ditampilkan pada Gambar 1.



Gambar 1 Grafik runtun waktu harga minyak goreng

Gambar 1 menunjukkan grafik runtun waktu sebagian provinsi di Indonesia cenderung memiliki pola yang tidak berbeda yaitu membentuk pola *trend* yang ditandai dengan terjadinya peningkatan secara drastis pada periode-periode akhir runtun waktu. Wilayah Indonesia Bagian Timur contohnya seperti Papua memiliki rentang nilai yang relatif lebih tinggi dibandingkan wilayah Indonesia bagian lainnya. Hal tersebut dapat dilihat pada rentang nilai yang disajikan pada grafik Provinsi Papua yang mencapai Rp 29.000.

A. Pengelompokan Data Harga Minyak Goreng

Proses pengelompokan menggunakan *K-Medoids* dengan ukuran kemiripan yang digunakan adalah jarak DTW. Tahapan-tahapan yang dilakukan adalah sebagai berikut:

1. Standarisasi data

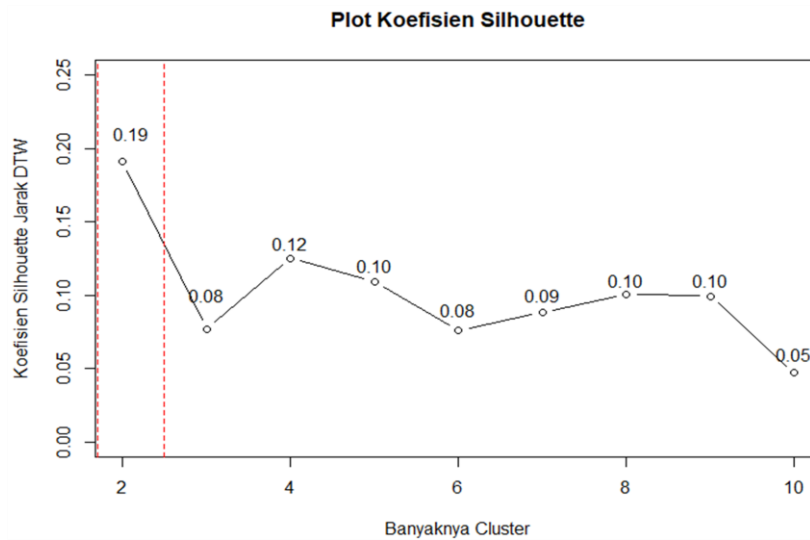
Tujuan dari standarisasi data adalah untuk membuat semua unit observasi (provinsi) memiliki nilai dalam jangkauan yang sama agar hasil pengelompokan tidak bias. Proses standarisasi data menggunakan Persamaan (1), (2), dan (3) dengan hasil standarisasi data disajikan pada Tabel 1.

Tabel 1 Hasil Standarisasi Data

t	$\tilde{Z}(1)_t$	$\tilde{Z}(2)_t$	$\tilde{Z}(3)_t$	$\tilde{Z}(4)_t$	$\tilde{Z}(5)_t$	$\tilde{Z}(6)_t$...	$\tilde{Z}(34)_t$
1	-0,35	-0,37	-0,25	-0,29	-0,52	-0,31	...	-0,35
2	-0,24	-0,38	-0,30	-0,29	-0,46	-0,28	...	-0,37
3	-0,28	-0,37	-0,36	-0,33	-0,46	-0,26	...	-0,36
4	-0,30	-0,40	-0,38	-0,36	-0,46	-0,48	...	-0,35
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
61	1,57	1,38	1,29	0,83	0,62	1,04	...	2,00

2. Penentuan Jumlah Kluster

Sebelum melakukan pengelompokan menggunakan *K-Medoids*, ditentukan jumlah kluster yang optimal dalam mengelompokkan data harga minyak goreng di Indonesia. Proses penentuan jumlah kluster optimal menggunakan koefisien *silhouette*. Grafik yang memuat nilai-nilai koefisien *silhouette* disajikan pada Gambar 2.



Gambar 2 Grafik nilai koefisien *silhouette*

Nilai *K* optimal terlihat dari koefisien *silhouette* terbesar. Berdasarkan Gambar 2, dapat dilihat bahwa jumlah kluster yang optimal dalam mengelompokkan harga minyak goreng di Indonesia sebanyak 2 kluster karena koefisien *silhouette* terbesar yang diperoleh sebesar 0,19.

3. Pengelompokan menggunakan *K-Medoids*

Sebelumnya telah diperoleh jumlah kluster optimal adalah 2 kluster. Selanjutnya adalah mengelompokkan provinsi di Indonesia berdasarkan data harga minyak goreng menggunakan *K-Medoids* dengan $K = 2$. Langkah pertama yaitu memilih 2 objek pada 34 provinsi di Indonesia sebagai objek *medoids*. Provinsi Aceh dan Provinsi Sulawesi Tengah dipilih secara acak sebagai *medoids* pada iterasi ke-1. Kemudian menghitung nilai jarak setiap objek dengan kedua *medoids* menggunakan jarak DTW.

Setelah menghitung jarak DTW untuk semua objek terhadap *medoids* berdasarkan Persamaan (5), selanjutnya adalah mengalokasikan setiap objek ke *medoids* terdekatnya dan menghitung fungsi objektif yang merupakan total jarak objek ke *medoids* terdekat, sehingga hasil nilai jarak serta informasi kluster pada iterasi pertama disajikan pada Tabel 2.

Tabel 2 Hasil Perhitungan Jarak DTW Iterasi ke-1

Provinsi	Jarak ke <i>medoids</i>		Terdekat	Kluster
	K1	K2		
Aceh	0,00	13,87	0,00	1
Sumatera Utara	6,59	12,02	6,59	1
Sumatera Barat	11,73	10,39	10,39	2
⋮	⋮	⋮	⋮	⋮
Papua	11,55	11,07	11,07	2
Papua Barat	13,26	12,46	12,46	2
Fungsi Objektif			304,62	

Berdasarkan Tabel 2, sebagai contoh untuk Provinsi Papua memperoleh jarak DTW ke *medoids* pertama sebesar 11,55 dan ke *medoids* kedua sebesar 11,07, sehingga Papua bergabung ke kluster 2 dikarenakan memiliki jarak lebih kecil ke *medoids* kedua dibandingkan ke *medoids* pertama. Setelah memperoleh fungsi objektif dari setiap objek pada iterasi ke-1, maka selanjutnya memilih salah satu objek dari setiap kluster sebagai kandidat *medoids* yang baru. Kandidat *medoids* baru untuk iterasi ke-2 adalah Provinsi DKI Jakarta dan Provinsi Nusa Tenggara Barat yang dipilih secara acak pada masing-masing kluster. Kemudian menghitung kembali nilai jarak terdekat dengan kandidat *medoids* baru (*non-medoids*) menggunakan jarak DTW dengan proses yang sama seperti pada iterasi ke-1. Setelah diperoleh jarak DTW untuk semua objek terhadap kandidat *medoids* baru, selanjutnya dihitung kembali fungsi objektif yang baru dan hasil nilai jarak DTW untuk semua objek ke kandidat *medoids* baru pada iterasi ke-2 dapat dilihat pada Tabel 3.

Tabel 3 Hasil Perhitungan Jarak DTW Iterasi ke-2

Provinsi	Jarak ke medoids		Terdekat	Klaster
	K1	K2		
Aceh	9,02	11,59	9,02	1
Sumatera Utara	6,94	10,28	6,94	1
Sumatera Barat	8,52	10,35	8,52	1
⋮	⋮	⋮	⋮	⋮
Papua	13,81	10,85	10,85	2
Papua Barat	15,95	13,41	13,41	2
Fungsi Objektif			293,66	

Fungsi objektif antara iterasi ke-1 dan iterasi ke-2 telah diperoleh, selanjutnya menghitung selisih antara fungsi objektif baru dengan fungsi objektif lama. Berdasarkan Tabel 2 diperoleh fungsi objektif lama sebesar 304,62 dan berdasarkan Tabel 3 diperoleh fungsi objektif baru sebesar 293,66. Artinya, selisih antara fungsi objektif baru dengan fungsi objektif lama menghasilkan selisih kurang dari nol ($S < 0$) yaitu sebesar -10,96, sehingga medoids awal diganti dengan medoids yang baru. Proses iterasi dilanjutkan dengan medoids yang baru sampai diperoleh fungsi objektif yang paling minimum. Penyelesaian algoritma *K-medoids* sampai diperoleh medoids yang optimal menggunakan bantuan software R. Berdasarkan software R diperoleh medoids yang optimal adalah Provinsi Banten sebagai medoids klaster pertama dan Provinsi Kalimantan Timur sebagai medoids klaster kedua dengan fungsi objektif sebesar 288,09.

4. Hasil klaster

Setelah memperoleh proses klastering yang optimal, tahap berikutnya yaitu menginterpretasikan hasil klaster berdasarkan jumlah klaster yang optimal. Pada proses peramalan harga minyak goreng di Indonesia menggunakan pendekatan *time series clustering* dengan algoritma *K-medoids* dan jarak yang digunakan adalah DTW dengan jumlah klaster optimal sebanyak 2 klaster, sehingga hasil pengelompokan yang diperoleh dapat dilihat pada Tabel 4.

Tabel 4 Hasil Klaster

Klaster	Provinsi	Jumlah		
1	Aceh, Riau, DKI Jakarta, Banten, Sumatera Utara, Sumatera Selatan, Jambi, Jawa Barat, Sumatera Barat, DI Yogyakarta, Lampung, Nusa Tenggara Barat, Jawa Tengah, Kalimantan Barat, Bali, Kalimantan Selatan, Gorontalo, Sulawesi Selatan, Jawa Timur	19		
	2		Kepulauan Riau, Kalimantan Timur, Kepulauan Bangka Belitung, Papua, Nusa Tenggara Timur, Kalimantan Utara, Sulawesi Tenggara, Kalimantan Tengah, Sulawesi Tengah, Maluku Utara, Sulawesi Utara, Sulawesi Barat, Maluku, Bengkulu, Papua Barat	15
			Total	

Visualisasi dari hasil klaster pada Tabel 4 dapat dilihat pada *spatial mapping* yang ditampilkan pada Gambar 3



Gambar 3 Spatial mapping hasil klaster

Berdasarkan Gambar 3 dapat diketahui hasil pengelompokan harga minyak goreng pada 34 provinsi di Indonesia. Pada klaster 1 terdapat 19 provinsi yang bergabung dan dapat dilihat pada wilayah yang memiliki warna lebih terang. Klaster 1 adalah klaster dengan harga minyak goreng terendah dibanding klaster 2. Selanjutnya pada klaster 2 terdapat 15 provinsi yang bergabung dan dapat dilihat pada wilayah yang memiliki warna lebih gelap. Klaster 2 adalah klaster dengan harga minyak goreng tertinggi.

V. KESIMPULAN DAN SARAN

Nilai K optimal dalam proses pengelompokan data harga minyak goreng di Indonesia menggunakan algoritma K -Medoids sebanyak 2 kluster berdasarkan koefisien *silhouette* terbesar, yaitu sebesar 0,19. Hasil pengelompokan data harga minyak goreng di Indonesia menggunakan algoritma K -Medoids dengan jarak DTW yaitu terdapat 19 provinsi yang bergabung pada kluster 1 di mana kluster dengan harga minyak goreng dibawah kluster 2 dan terdapat 15 provinsi yang bergabung pada kluster 2 yang merupakan kluster dengan harga minyak goreng tertinggi.

Koefisien *silhouette* yang dihasilkan untuk menentukan K optimal tergolong relatif kecil, sehingga penelitian selanjutnya disarankan dapat menggunakan optimasi K -Medoids Time Series Clustering dengan Principle Component Analysis (PCA).

REFERENSI

- [1] J. Han, M. Kamber, dan J. Pei, "Data Mining: Concept and Techniques Third Edition," Waltham, 2012.
- [2] N. Yudistira, "Peran Big Data dan Deep Learning untuk Menyelesaikan Permasalahan Secara Komprehensif," *EXPERT: Jurnal Manajemen Sistem Informasi dan Teknologi*, vol. 11, no. 2, hlm. 78–89, Des 2021.
- [3] M. A. A. Riyadi, K. Fithriasari, dan Dwiatmono, "Data Mining Peramalan Konsumsi Listrik dengan Pendekatan Cluster Time Series sebagai Preprocessing," *Jurnal Sains dan Seni ITS*, vol. 5, no. 1, hlm. 2337–3520, 2016.
- [4] L. Kaufman dan P. J. Rousseeuw, *Finding Groups in Data An Introduction to Cluster Analysis*. New Jersey: John Wiley & Sons Inc Publication, 1990.
- [5] M. A. A. Riyadi, D. S. Pratiwi, A. R. Irawan, dan K. Fithriasari, "Clustering stationary and non-stationary time series based on autocorrelation distance of hierarchical and K-means algorithms," *International Journal of Advances in Intelligent Informatics*, vol. 3, no. 3, hlm. 154–160, Nov 2017.
- [6] A. T. R. Dani, S. Wahyuningsih, dan N. A. Rizki, "Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu," *Jambura Journal of Mathematics*, vol. 1, no. 2, hlm. 64–78, 2019.
- [7] R. Novidianto dan A. T. R. Dani, "Analisis Kluster Kasus Aktif COVID-19 Menurut Provinsi di Indonesia Berdasarkan Data Deret Waktu," *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 12, hlm. 15–24, 2020.
- [8] M. I. Rizki, T. A. Taqqiyuddin, dan J. J. Cerelia, "K-Medoids Clustering dengan Jarak Dynamic Time Warping dalam Mengelompokkan Provinsi di Indonesia Berdasarkan Kasus Aktif Covid-19," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 5, hlm. 685–692, 2021.
- [9] I. Gunawan, G. Anggraeni, S. Rini, Y. M. Putri, dan Y. K. Zikri, "Klasterisasi provinsi di Indonesia berbasis perkembangan kasus Covid-19 menggunakan metode K-Medoids," *Seminar Nasional Matematika Dan Pendidikan Matematika (5th SENATIK)*, hlm. 301–306, 2020.
- [10] M. A. Zen, S. Wahyuningsih, dan A. T. R. Dani, "Aplikasi Pendekatan Agglomerative Hierarchical Time Series Clustering untuk Peramalan Data Harga Minyak Goreng di Indonesia," *Prosiding Seminar Nasional Official Statistics 2022*, hlm. 293–302, 2022.
- [11] E. Prasetyo, *Konsep dan Aplikasi menggunakan MATLAB*. Yogyakarta: Andi Offset, 2012.
- [12] P. Montero dan J. A. Vilar, "TSclust: An R Package for Time Series Clustering," 2014.
- [13] V. N. Chotirat dan A. Ratanamahatana, "On Clustering Multimedia Time Series Data Using K-Means and Dynamic Time Warping," *2007 International Conference on Multimedia and Ubiquitous Engineering*, 2007.
- [14] B. Wira, A. Endy Budianto, dan A. Sartika Wiguna, "Implementasi Metode K-Medoids Clustering untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 di Universitas Kanjuruhan Malang," *Jurnal Terapan Sains & Teknologi*, vol. 1, no. 3, hlm. 54–69, 2019.
- [15] J. Kim dan J. Kim, "Comparison of time series clustering methods and application to power consumption pattern clustering," *Commun Stat Appl Methods*, vol. 27, no. 6, hlm. 589–602, 2020.



© 2023 by the authors. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (<http://creativecommons.org/licenses/by-sa/4.0/>).