

# Implementasi *K-Means* dan *Hierarchical Clustering* Pada Penentuan Tingkatan *Smart City* Tahun 2022 Berdasarkan *Motion Index*

## *Implementation of K-Means and Hierarchical Clustering in Determining Levels of Smart City 2022 Based on Motion Index*

<sup>1</sup>Nissa Shahadah Qur'ani\*, <sup>2</sup>Arie Wahyu Wijayanto

<sup>1</sup>Statistika Ekonomi, Statistika, Politeknik Statistika STIS  
Jalan Otto Iskandardinata No. 64C, Jatinegara, Jakarta, Indonesia

\*e-mail: [211910910@stis.ac.id](mailto:211910910@stis.ac.id)

(received: 12 Desember 2022, revised: 27 April 2023, accepted: 4 Mei 2023)

### Abstrak

*Smart City* adalah kota dengan konsep pembangunan yang inovatif. Namun, tidak semua *Smart City* di negara-negara mempunyai standar yang sama karena mereka cukup heterogen. Maka, dilakukan analisis kluster untuk mengelompokkan *Smart City*. Hasil yang diperoleh adalah *Smart City* terbagi menjadi dua level yaitu level tinggi dan rendah. Metode yang digunakan adalah *k-means* dan *hierarchical clustering*. Dasar dari pengelompokan ini menerapkan *motion index* yaitu indikator sosial, ekonomi, lingkungan, mobilisasi dan transportasi, serta profil internasional yang diwakilkan oleh berbagai variabel. Harapannya adalah *Smart City* pada level tertentu dapat dibandingkan dengan level lainnya, agar ada perbaikan serta saling mempelajari mengenai *Smart City* pada level tinggi. Hal ini pun juga dapat mendorong kota lainnya dalam berproses menuju *Smart City*.

**Kata kunci:** *Hierarchical Clustering, K-Means Clustering, Motion Index, Smart City.*

### Abstract

*Smart City* is a city with an innovative development concept. However, not all *Smart Cities* in countries have the same standard because they are quite heterogeneous. Thus, a cluster analysis was carried out to classify *Smart City*. The result shows that *Smart City* is divided into two levels, those are high and low. *k-means* and *hierarchical clustering* is used for the method of this research. The grouping is based on the *motion index*, which consists of economic, environmental, mobilization and transportation indicators, and also international profiles represented by various variables. This research expects that *Smart City* at a certain level can be compared with other levels, in order to there are improvements and mutual learning about *Smart City* at a high level. This can also encourage other cities in the process towards *Smart City*.

**Keywords:** *Hierarchical Clustering, K-Means Clustering, Motion Index, Smart City.*

## 1 Pendahuluan

Sudah menjadi tujuan bagi setiap kota di negara-negara untuk menciptakan kondisi yang nyaman bagi penduduknya seperti adanya pertumbuhan daya saing dan pembangunan berkelanjutan. Hal inilah yang menjadi suatu inisiasi bahwa kota dapat dijadikan *Smart City*. Oleh karena itu, semakin banyak kota yang beralih konsep ke pembangunan inovatif. Tidak terkecuali dengan negara-negara berkembang yang berupaya dalam meningkatkan potensi intelektual dan memastikan pertumbuhan wilayah yang berkelanjutan. *Smart City* memungkinkan untuk meningkatkan kualitas hidup penduduk, mengurangi kesenjangan sosial-ekonomi, dan membuat manajemen kota lebih efektif. Sebelum memetakan strategi pengembangan *Smart City*, penting untuk melihat hal yang sudah ada dan meningkatkannya. Maka, lebih baik memanfaatkan sumber daya *Smart City* yang ada terlebih dahulu daripada fokus pada yang belum ada [1]. Adanya kondisi khusus yang muncul secara tidak merata di beberapa wilayah menciptakan adanya *Smart City* itu sendiri pada kota tertentu. Analisis kluster memiliki keunggulan yang jelas dibandingkan jenis metode lainnya. Pendekatan ini memungkinkan untuk mengidentifikasi suatu area, seperti memungkinkan untuk menentukan tempat paling sukses sebagai lokasi *Smart City*.

## 2 Tinjauan Literatur

*Smart City* akan meningkatkan efisiensi penempatan dan penggunaan sumber daya serta meningkatkan kualitas hidup penduduk dengan meningkatkan keamanan, meningkatkan proses, dan menyediakan layanan baru. Istilah '*Smart City*' muncul secara progresif pada 1990-an. Selama 20 tahun terakhir, konsep *Smart City* memiliki banyak definisi dan menjadi tempat di mana teknologi informasi digabungkan dengan infrastruktur untuk mengatasi berbagai masalah seperti lingkungan, ekonomi, dan sosial [2]. Leydesdorff dan Deakin [3] menekankan bahwa *Smart City* adalah proses rekonstruksi budaya yang didukung oleh kebijakan, kepemimpinan, dan strategi perusahaan dalam kepemimpinannya. Juga dicatat bahwa selama dua dekade terakhir, kota-kota besar di seluruh dunia telah terlibat dalam inisiatif untuk meningkatkan infrastruktur dan layanan perkotaan yang bertujuan untuk memperbaiki lingkungan, dan kondisi sosial dan ekonomi, meningkatkan daya tarik dan daya saing kota [4]. Visi "*Smart Cities*" adalah pusat perkotaan masa depan, aman, berlingkungan hijau, serta efisien karena dirancang dengan memanfaatkan kemajuan, bahan terintegrasi, dan sensor, elektronik yang terdiri dari database, pelacakan, dan algoritma pengambilan keputusan baik transportasi, listrik, dan air dan lainnya [5]. Secara umum, *Smart City* banyak menggunakan teknologi informasi dan komunikasi untuk membantu kota besar membangun keunggulan kompetitifnya [1]. Terlepas dari berbagai konsep dan teori, tidak ada definisi yang jelas tentang istilah "*Smart City*" [6].

Penelitian Villessuzanne, Weigel, dan Blain [7] adalah mengelompokkan 40 kota di Eropa menggunakan *Principal Component Analysis* (PCA) dengan menggunakan 7 dimensi yaitu *smart economy*, *smart architecture and technology*, *smart mobility*, *smart citizens*, *smart government*, *smart environment*, serta *quality of life*, didapatkan hasil bahwa terdapat klasifikasi meningkat secara hierarkis yang mengidentifikasi tiga klaster/kelompok kota yang relevan. Hasilnya, klaster pertama mewakili kota-kota dengan strategi cerdas yang muncul. Klaster kedua mengelompokkan kota-kota metropolitan internasional yang memiliki strategi berorientasi pada teknologi. Klaster ketiga mewakili kota-kota Eropa yang berukuran sedang dengan kualitas hidup yang baik. Berdasarkan penelitian Digel et al [8], ia mengelompokkan 17 wilayah di Kazakhstan menggunakan metode *agglomerative clustering* menggunakan *single linkage*, *complete linkage*, dan metode *Ward's clustering*. Dihadirkan bahwa terdapat ketimpangan regional mendalam yang memengaruhi potensi untuk berhasil mengembangkan dan mengelola *Smart City*. Ada tiga wilayah dengan potensi tertinggi yaitu wilayah Atyrau, kota Almaty, dan kota Astana yang memiliki banyak modal, aktivitas litbang dan inovasi yang tinggi, tingkat lapangan kerja, serta produktivitas tenaga kerja untuk mendukung pengembangan *Smart City*. Daerah yang tersisa akan membutuhkan lebih banyak sumber daya dan upaya untuk membangun *Smart City*. Saptadi, Chayn, dan Taga [9] menghasilkan penelitian mengenai perilaku masyarakat mengelola sampah di Kota Makassar Indonesia dalam proses menuju *Smart City*. Indikator yang digunakan adalah pendidikan, pekerjaan, dan waktu luang. Penelitian ini menggunakan metode *k-means clustering* dengan hasil akhir yaitu terbentuk 2 klaster. Pada klaster 1 diketahui bahwa masyarakat cenderung menjaga lingkungan di kecamatan tempat tinggal masyarakat. Klaster 2 memiliki rata-rata hasil yang rendah, seperti masyarakat yang tidak mengetahui undang-undang dampak, fasilitas sanitasi yang kurang memadai, dan pengelolaan sampah yang kurang baik dalam menuju tata kelola *Smart City*. Aralbaeva dan Berikbolova [10] juga menggunakan metode *k-means clustering* dalam menganalisis tingkat perkembangan inovatif wilayah di Kazakhstan, ia menggunakan indikator persentase inovasi dalam perusahaan, volume produk inovatif (barang dan jasa), biaya untuk inovasi produk, dan jumlah perusahaan yang memiliki jenis inovasi. Indikator tersebut membentuk 3 klaster yaitu tinggi, medium, dan rendah. Penelitian dengan metode selain *clustering* lainnya dilakukan oleh Kubina, Šulyová, dan Vodák [11], mereka menggunakan model survei klaster oleh Arcadis yaitu perusahaan terkemuka di dunia yang memberikan solusi desain, teknik, dan konsultasi yang berkelanjutan untuk aset alam dan bangunan. Didapati *Smart City* di Amerika Utara dan Eropa dapat diklasifikasikan menjadi empat kelas klaster yaitu klaster "*Balance innovators*", "*Post-industrial opportunists*", "*Evolutionary Cities*", dan "*Fast-growing megacities*" yang dikelompokkan berdasarkan tiga pilar, aspek sosial (manusia)—mobilitas dan kualitas hidup; planet (perlindungan lingkungan)—pengelolaan sumber daya yang terbatas.; aspek ekonomi (keuntungan)—kinerja ekonomi dan PDB (Produk Domestik Bruto).

Berdasarkan tinjauan literatur, dapat disimpulkan bahwa ada studi yang ditujukan untuk masalah pembentukan, pengelompokan, dan pengelolaan pembangunan *Smart City*. Namun, masih jarang

ditemui yang menggunakan 2 metode kemudian dibandingkan untuk didapatkan metode terbaik. Maka, dari itu penelitian ini membandingkan 2 metode dalam pengelompokan *Smart City* berdasarkan suatu indeks. Terdapat banyak dimensi dan indikator yang dapat membentuk atau mengelompokkan *Smart City*, peneliti memutuskan untuk menggunakan *motion index*. *Motion index* berfungsi untuk mengevaluasi perkembangan kota-kota di dunia. Sekolah Bisnis IESE Universitas Navarra [12] menyatakan bahwa terdapat 9 dimensi dari *motion index* yaitu ekonomi, modal manusia, teknologi, lingkungan, profil internasional, kohesi sosial, mobilitas dan transportasi, pemerintahan, serta tata kelola dan perencanaan kota. Berdasarkan indeks tersebut akan dilakukan pengelompokan kota-kota besar di dunia. Hasil pengelompokan ini akan memberikan gambaran dan informasi mengenai kota yang dapat dijadikan cerminan dalam proses menuju *Smart City* dan kota yang seharusnya dapat memperbaiki tata kelolanya. Analisis kluster pada penelitian ini digunakan untuk menentukan tingkat perkembangan wilayah dan pengelompokan kota. Penelitian ini menggunakan algoritma *k-means* dan *hierarchical clustering* untuk melakukan pengelompokan karena merupakan metode *unsupervised*. Metode *unsupervised* diperlukan karena berguna dalam tahap eksplorasi yang diperlukan untuk memperoleh pengetahuan awal tentang bidang studi. Selain itu, *k-means* dan *hierarchical clustering* adalah algoritma pengelompokan terkenal yang sering digunakan dalam studi pengelompokan [13]. Oleh karena itu, penelitian ini bertujuan untuk mengelompokkan berdasarkan suatu tingkatan atau level *Smart City* dari kota-kota besar di berbagai negara dan kemudian mengidentifikasi kota-kota dengan potensi terbaik untuk mengembangkan “*Smart City*” berdasarkan analisis kluster serta menjelaskan perbedaan variabel setiap kluster untuk dianalisis lebih jauh.

### 3 Metode Penelitian

Data yang digunakan pada penelitian ini adalah data sekunder diambil dari *website* Kaggle [14] yaitu data “*Smart Cities Indexes 2022*” yang bersumber dari Numbeo [15]. Terdapat 319 kota besar (*Smart City*) di berbagai negara sebagai observasi pada penelitian ini. Tidak digunakannya semua kota sebagai observasi dikarenakan keterbatasan ketersediaan data pada *website* Numbeo. Numbeo adalah *website* penyedia data terbesar di dunia tentang kota dan negara. Adapun variabel yang digunakan terbagi menjadi beberapa indikator. Data dari setiap variabel diolah menggunakan perangkat lunak *RStudio* dan *Microsoft Excel 2021*. Pada penelitian ini, dari 9 indikator yang ada hanya 5 indikator yang digunakan, variabel yang disertakan juga tidak dari semua variabel setiap indikator pada *motion index* dikarenakan keterbatasan jumlah dan sumber data. Hal ini dapat berakibat terjadinya kemungkinan bahwa jumlah variabel yang disertakan tidak cukup untuk menangkap kompleksitas setiap dimensi. Beberapa variabel yang digunakan dalam penelitian ini tersaji pada Tabel 1.

Tabel 1. Deskripsi Variabel

| Kode              | Variabel                    | Deskripsi Variabel   | Satuan        |
|-------------------|-----------------------------|--|---------------|
| Indikator Sosial  |                             |  |               |
| A1                | Indeks Kejahatan            | Estimasi tingkat umum kejahatan di suatu kota  | Rasio [0,100] |
| A2                | Indeks Keamanan             | Jika kota tersebut memiliki indeks keamanan yang tinggi, maka dianggap sangat aman.  | Rasio [0,100] |
| A3                | Indeks Kesehatan            | Estimasi kualitas keseluruhan sistem perawatan kesehatan, profesional perawatan kesehatan, peralatan, personel, biaya, dll   | Rasio [0,100] |
| Indikator Ekonomi |                             |  |               |
| B1                | Rasio Harga Sewa Pusat Kota | Rata-rata biaya kepemilikan dibagi dengan pendapatan sewa yang diterima (jika membeli untuk menyewakan) atau taksiran sewa yang akan dibayarkan jika menyewa (jika membeli untuk bertempat tinggal) untuk rumah di pusat kota. | Rasio         |
| B2                | Indeks Keterjangkauan       | Kebalikan dari hipotek sebagai persentase pendapatan.  | Rasio         |
| B3                | Indeks Biaya                | Indikator relatif harga barang konsumen, termasuk bahan  | Rasio         |

|   |                                |   |               |
|---|--------------------------------|---|---------------|
|   | Hidup                          | makanan, restoran, transportasi dan utilitas. Indeks biaya hidup tidak termasuk biaya akomodasi seperti sewa atau hipotek.  |               |
| B4  | Indeks Sewa                    | Perkiraan harga sewa apartemen di kota dibandingkan dengan New York City.   | Rasio         |
| B5  | Indeks Biaya Hidup Plus Sewa   | Estimasi harga barang konsumsi termasuk sewa dibandingkan dengan Kota New York.   | Rasio         |
| B6  | Indeks bahan makanan           | Perkiraan harga grosir di kota dibandingkan dengan Kota New York.   | Rasio         |
| B7  | Indeks Daya Beli Lokal         | Daya beli dalam membeli barang dan jasa di kota (berdasarkan gaji rata-rata), dibandingkan dengan penduduk Kota New York.   | Rasio         |
| <b>Indikator Lingkungan</b>                 |                                |   |               |
| C1  | Indeks Polusi                  | Perkiraan polusi keseluruhan di kota. Bobot terbesar diberikan pada polusi udara, daripada polusi/aksesibilitas air, dua faktor polusi utama. Bobot kecil diberikan untuk jenis polusi lainnya.                                   | Rasio [0,100] |
| C2  | Indeks Emisi CO <sub>2</sub> . | Indeks emisi CO <sub>2</sub> .  | Rasio (gram)  |
| <b>Indikator Mobilitas dan Transportasi</b> |                                |   |               |
| D1  | Indeks Lalu Lintas             | Indeks komposit waktu yang dikonsumsi dalam lalu lintas karena perjalanan kerja, estimasi ketidakpuasan konsumsi waktu, estimasi konsumsi CO <sub>2</sub> dalam lalu lintas dan inefisiensi keseluruhan dalam sistem lalu lintas. | Rasio         |
| D2  | Indeks Waktu                   | Rata-rata waktu sekali jalan yang diperlukan untuk bepergian.   | Rasio (Menit) |
| D3  | Indeks Inefisiensi             | Indeks perkiraan inefisiensi lalu lintas. Nilai tinggi menunjukkan inefisiensi berkendara yang tinggi, seperti waktu tempuh yang lama   | Rasio         |
| <b>Indikator Profil Internasional</b>       |                                |   |               |
| E1  | Indeks Harga Restoran          | Perbandingan harga makanan dan minuman di restoran dan bar di kota dengan harga di Kota New York.   | Rasio         |

Sumber: Numeo, 2022

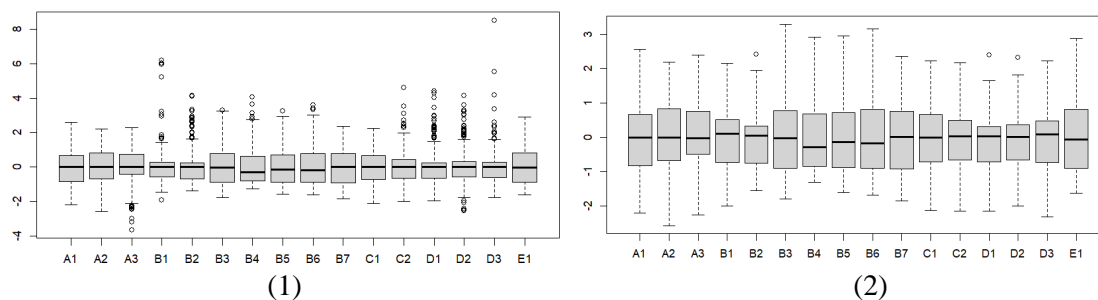
Metode analisis yang digunakan adalah *k-means* dan *hierarchical clustering*. *K-means* dan *hierarchical clustering* adalah salah satu algoritma bersifat *unsupervised learning* yang memiliki fungsi untuk mengelompokkan data ke dalam kluster sejumlah k kluster. Sehingga algoritma ini dapat menerima data tanpa adanya label kategori. Setelah didapatkan k kluster yang paling tepat dalam mengelompokkan level *Smart City* maka dilakukan analisis deskriptif mengenai setiap variabel indikator per kluster.

#### 4 Hasil dan Pembahasan

Sebelum dilakukan pengolahan kemudian analisis terhadap data, tahapan yang dilakukan adalah *preprocessing data*.

##### Tahap *Preprocessing Data*

Pada tahap ini dilakukan deteksi ada atau tidaknya *missing value* dan *outlier*. Dari hasil pendeteksian terdapat *missing value* pada data sehingga perlu dilakukan imputasi data. Imputasi dilakukan dengan menggunakan metode *sequential* dengan cara mengganti *missing value* dengan rata-rata nilai dari atribut tersebut. Hasil dari deteksi *outlier* didapatkan sebagai berikut.



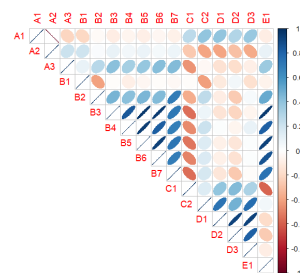
Sumber: Numbeo 2022, diolah.

**Gambar 1. Boxplot Setiap Variabel Sebelum Dinormalisasi (1) dan Setelah Dinormalisasi (2)**

Dari hasil tampilan visualisasi pada Gambar 1 (1) mengartikan bahwa ditemukan adanya *outlier* pada beberapa variabel/atribut. Proses *clustering* menjadi tidak efisien karena adanya *outlier* tersebut, sehingga diperlukan solusi untuk mengatasinya. Ghazali [16] mengungkapkan terdapat 3 cara dalam mengatasi data ekstrim karena adanya *outlier*, yaitu *transforming*, *winsorizing*, dan *treammean*. Dalam penelitian ini, data *outlier* diatasi dengan teknik *winsorizing*. Teknik *winsorizing* adalah transformasi statistik dengan mengonversi nilai titik data yang sangat tinggi ke nilai titik data tertinggi yang tidak dianggap sebagai *outlier* pada data sehingga mengurangi efek *outlier* [17]. Teknik ini dilakukan dengan mengubah nilai *outlier* menjadi 95% batas ekor atas/bawah ( $1,5 \pm IQR$ ). Selain itu, dikarenakan terdapat satuan dari variabel yang berbeda dengan yang lainnya, maka diperlukan normalisasi atau *scaling* data untuk menyamakan satuan atau skala dari setiap variabel/atribut dengan cara transformasi *z-score*. Adapun setelah *outlier* diatasi dan data dinormalisasi didapatkan hasil visualisasi seperti pada Gambar 1 (2) yaitu data sudah tidak memiliki *outlier* yang berarti dan setiap variabel dapat dibandingkan dengan yang lain karena memiliki skala yang sama.

### Principal Component Analysis

Data yang sudah tidak mengandung *missing value* dan *outlier* selanjutnya perlu dipastikan bahwa tidak adanya korelasi antar atribut atau multikolinieritas yang akan digunakan dalam *clustering*. Dalam rangka mendeteksi multikolinieritas pada data dilakukan Uji Bartlett untuk melihat plot korelasi antar atribut.



Sumber: Numbeo 2022, diolah.

**Gambar 2. Plot Korelasi Antar Variabel**

Plot Gambar 2 di atas menunjukkan bahwa adanya indikasi multikolinieritas yang ditunjukkan pada gambar berwarna merah dan biru tua. Selanjutnya dilakukan Uji Bartlett sebagai uji inferensia, hipotesis yang digunakan adalah sebagai berikut:

$H_0$ : Matriks korelasi sama dengan matriks identitas (tidak ada multikolinieritas)

$H_1$ : Matriks korelasi tidak sama dengan matriks identitas (ada multikolinieritas)

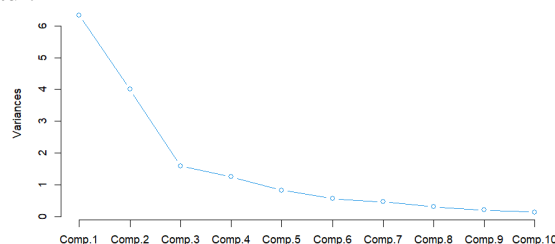
**Tabel 2. Hasil Uji Bartlett**

| Uji      | Statistik Uji | p-value                | Kesimpulan  |
|----------|---------------|------------------------|-------------|
| Bartlett | 29575,382     | $2,22 \times 10^{-16}$ | Tolak $H_0$ |

Sumber: Data sekunder, diolah

Berdasarkan uji bartlett pada Tabel 2, dengan tingkat kepercayaan 95 persen tidak terdapat cukup bukti untuk mengatakan bahwa data tidak mengandung adanya multikolinieritas sehingga antar atribut memiliki korelasi atau terjadi multikolinieritas. Langkah selanjutnya adalah mereduksi variabel dengan *Principal Component Analysis* (PCA) untuk mengatasi multikolinieritas. Sebelum

menerapkan PCA, terlebih dahulu membuang atribut yang tidak memenuhi *Kaiser Mayer Olkin* (KMO). Atribut yang tidak memenuhi yakni atribut yang memiliki *Measure of Sampling Adequacy* (MSA) kurang dari 0,5 sehingga dapat dibuang dan tidak diikuti dalam analisis lanjutan. Hasil pengolahan menunjukkan bahwa tidak ada nilai MSA yang kurang dari 0,5. Maka dari itu, tidak ada atribut yang dikeluarkan, sehingga atribut dapat dilanjutkan pada PCA untuk mengatasi multikolinearitas yang terjadi.



Sumber: Numbeo 2022, diolah.

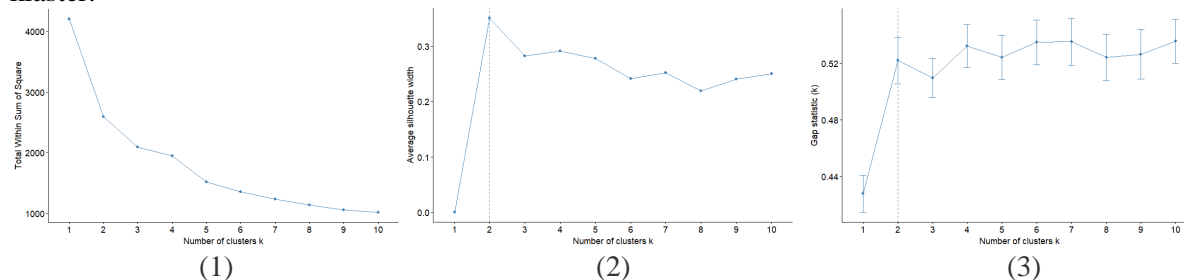
**Gambar 3. Scree Plot**

Hasil dari PCA ditampilkan dalam *scree plot* Gambar 3. Pada *scree plot* tersebut, terlihat bahwa kurva mulai melandai pada komponen utama yang ketiga. Namun berdasarkan hasil dari olahan *R Studio* bahwa nilai proporsi kumulatif varians komponen ketiga adalah 74,79 persen. Sedangkan menurut Johnson [18], dasar menentukan banyaknya komponen utama adalah proporsi kumulatif keragaman komponen minimal 80 persen. Maka dari itu peneliti menambahkan satu komponen sehingga menjadi empat komponen dan proporsi kumulatif keragaman menjadi 82,68 persen. Sehingga, empat komponen utama akan digunakan dalam analisis kluster menggunakan metode *k-means* dan *hierarchical clustering* yang kemudian dibandingkan untuk dipilih sebagai metode terbaik.

### **K-Means Clustering**

Teknik *unsupervised learning* digunakan untuk melakukan proses pembelajaran pada data yang tidak memiliki informasi label dari objek yang akan diteliti. *K-means clustering* adalah metode yang digunakan untuk melakukan pengelompokan objek ke dalam sejumlah *k* kluster. Metode ini adalah metode pengelompokan data sederhana dan banyak digunakan dalam aplikasi *data mining*. Input dari *k-means clustering* adalah data/objek dan jumlah kluster *k* yang diinginkan. Setiap kluster direpresentasikan oleh sebuah titik pusat (*centroid*) [19]. Secara umum, cara kerja algoritma *k-means clustering* adalah menentukan jumlah kluster; memilih *k* buah titik *centroid* secara acak; mengelompokkan objek-objek ke dalam sejumlah *k* kluster, di mana kluster yang terpilih adalah yang jarak *centroid*-nya dengan objek yang akan dikelompokkan paling kecil; setelah terbentuk *k* kluster, perbarui nilai *centroid*; ulangi hingga nilai *centroid* tidak berubah lagi.

Penentuan jumlah kluster yang optimal dilakukan dengan melihat *within sum of square*, *silhouette*, dan *gap statistic* pada Gambar 4. Dari total *within sum of square*, kurva melandai ketika *k* adalah 5, artinya jumlah kluster yang disarankan adalah sebanyak 5 kluster, selain itu juga karena penambahan jumlah kluster tidak membawa pengaruh banyak atas variasi di dalam kluster tersebut. Pada rata-rata lebar *silhouette*, kurva mencapai titik maksimum ketika *k* berjumlah 2, artinya jumlah kluster yang diajukan adalah sebanyak 2 kluster. Nilai *gap statistic* paling besar adalah ketika jumlah kluster adalah 2, artinya jumlah kluster yang dapat diaplikasikan untuk data ini adalah sebanyak 2 kluster.



Sumber: Numbeo 2022, diolah.

**Gambar 4. Total Within Sum of Square (1), Rata-Rata Silhouette Width (2), Gap Statistic (3) Untuk Setiap Jumlah Kluster**

Dalam meyakinkan jumlah kluster yang akan digunakan maka dilakukan validasi kluster. Dari hasil validasi kluster didapatkan nilai *connectivity* yang menunjukkan angka paling kecil ketika *k*

<http://sistemasi.ftik.unisi.ac.id>

adalah 2, nilai *dunn* paling besar saat *k* adalah 8, serta *silhouette* menunjukkan angka paling besar ketika *k* bernilai 2. Berdasarkan hal tersebut, peneliti memutuskan 2 kluster untuk digunakan sebagai jumlah kluster yang harus dibentuk. Langkah berikutnya adalah menghitung jarak antara objek dengan titik *centroid* yang dapat dihitung menggunakan penghitungan jarak *Minkowski* dengan formula pada persamaan (1).

$$d(x_i, x_j) = (|x_{i1} - x_{j1}|^g + |x_{i2} - x_{j2}|^g + \dots + |x_{ip} - x_{jp}|^g)^{\frac{1}{g}} \quad (1)$$

Di mana:

$g = 1$ , untuk menghitung jarak Manhattan

$g = 2$ , untuk menghitung jarak Euclidean

$g = \infty$ , untuk menghitung jarak Chebychev

$x_i, x_j$  adalah dua buah data yang akan dihitung jaraknya untuk menghitung jarak Manhattan

$p$  = atribut ke- $p$  dari sebuah objek

Setelah jumlah kluster ditentukan, kemudian menetapkan nilai *initial centroid cluster* untuk setiap variabel pada masing-masing kluster. Nilai *initial centroid cluster* diberikan secara acak pada iterasi pertama. Pada iterasi selanjutnya, *initial centroid cluster* dihitung dari nilai rata-rata pada setiap klasternya. Proses iterasi akan berhenti jika nilai *centroid cluster* baru sama dengan *centroid cluster* yang lama atau tidak berubah lagi. Peneliti menggunakan jarak *Euclidean* untuk menghitung jarak antara objek dengan titik *centroid*. Data yang digunakan adalah data yang terdiri dari 4 komponen pada Tabel 3 setelah melewati tahap PCA.

Tabel 3. Data 4 Komponen Utama *Smart City*

| No  | Kota                            | KU1    | KU2    | KU3    | KU4    | Kluster |
|-----|---------------------------------|--------|--------|--------|--------|---------|
| 1   | Aalborg, Denmark                | -2,031 | -0,322 | -0,910 | -1,242 | 2       |
| 2   | Aberdeen, United Kingdom        | -1,091 | -0,047 | 0,048  | -1,493 | 2       |
| 3   | Abu Dhabi, United Arab Emirates | -2,356 | -1,480 | -0,205 | -3,380 | 2       |
| 4   | Accra, Ghana                    | 3,076  | 0,014  | 0,627  | -0,601 | 1       |
| ⋮   | ⋮                               | ⋮      | ⋮      | ⋮      | ⋮      | ⋮       |
| 318 | Zagreb, Croatia                 | 0,403  | -2,958 | -0,992 | -0,255 | 1       |
| 319 | Zurich, Switzerland             | -6,401 | -1,951 | -2,203 | 1,565  | 2       |

Sumber: Data sekunder, diolah

Adapun perhitungan jarak antar data ke-1 pada masing-masing kluster adalah sebagai berikut.

$$d(x_1, c_1) = \sqrt{(KU_{11} - KU_{1c1})^2 + (KU_{21} - KU_{2c1})^2 + (KU_{31} - KU_{3c1})^2 + (KU_{41} - KU_{4c1})^2}$$

$$= \sqrt{(-2,031 - KU_{1c1})^2 + (-0,322 - KU_{2c1})^2 + (-0,910 - KU_{3c1})^2 + (-1,242 - KU_{4c1})^2} \quad (2)$$

$$d(x_1, c_2) = \sqrt{(KU_{11} - KU_{1c2})^2 + (KU_{21} - KU_{2c2})^2 + (KU_{31} - KU_{3c2})^2 + (KU_{41} - KU_{4c2})^2}$$

$$= \sqrt{(-2,031 - KU_{1c2})^2 + (-0,322 - KU_{2c2})^2 + (-0,910 - KU_{3c2})^2 + (-1,242 - KU_{4c2})^2} \quad (3)$$

$$d(x_2, c_1) = \sqrt{(KU_{12} - KU_{1c1})^2 + (KU_{22} - KU_{2c1})^2 + (KU_{32} - KU_{3c1})^2 + (KU_{42} - KU_{4c1})^2}$$

$$= \sqrt{(-1,091 - KU_{1c1})^2 + (-0,047 - KU_{2c1})^2 + (0,048 - KU_{3c1})^2 + (-1,493 - KU_{4c1})^2} \quad (4)$$

$$d(x_2, c_2) = \sqrt{(KU_{12} - KU_{1c2})^2 + (KU_{22} - KU_{2c2})^2 + (KU_{32} - KU_{3c2})^2 + (KU_{42} - KU_{4c2})^2}$$

$$= \sqrt{(-1,091 - KU_{1c2})^2 + (-0,047 - KU_{2c2})^2 + (0,048 - KU_{3c2})^2 + (-1,493 - KU_{4c2})^2} \quad (5)$$

Dihitung seterusnya hingga observasi ke-319 untuk setiap kluster dalam mendapatkan jarak pada tiap data. Langkah selanjutnya adalah memperbarui nilai titik *centroid* di masing-masing kluster yang dapat dihitung dengan formula persamaan (6).

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \quad (6)$$

$\mu_k$  = titik *centroid* dari kluster ke- $k$ .

$N_k$  = banyaknya data pada kluster ke-k.

$x_q$  = data ke-q pada kluster ke-k.

Hasil pengelompokan menunjukkan jumlah observasi yang hampir merata antara kluster 1 yaitu 159 kota dan kluster 2 sebanyak 160 kota. Sehingga penghitungan *centroid* baru sebagai berikut.

$$\mu_{1KU1} = \frac{1}{N_1} \sum_{q=1}^{N_1} x_q = \frac{1}{159} \sum_{q=1}^{159} x_q = \frac{1}{159} (3,076 + \dots + 0,403) = 2,2549 \quad (7)$$

$$\mu_{2KU1} = \frac{1}{N_2} \sum_{q=1}^{N_2} x_q = \frac{1}{160} \sum_{q=1}^{160} x_q = \frac{1}{160} (-2,031 + \dots - 6,401) = -2,2408 \quad (8)$$

$$\mu_{1KU2} = \frac{1}{N_1} \sum_{q=1}^{N_1} x_q = \frac{1}{159} \sum_{q=1}^{159} x_q = \frac{1}{159} (0,014 + \dots - 2,958) = -0,0995 \quad (9)$$

$$\mu_{2KU2} = \frac{1}{N_2} \sum_{q=1}^{N_2} x_q = \frac{1}{160} \sum_{q=1}^{160} x_q = \frac{1}{160} (-0,322 + \dots - 1,951) = 0,0989 \quad (10)$$

$$\mu_{1KU3} = \frac{1}{N_1} \sum_{q=1}^{N_1} x_q = \frac{1}{159} \sum_{q=1}^{159} x_q = \frac{1}{159} (0,627 + \dots - 0,992) = -0,0383 \quad (11)$$

$$\mu_{2KU3} = \frac{1}{N_2} \sum_{q=1}^{N_2} x_q = \frac{1}{160} \sum_{q=1}^{160} x_q = \frac{1}{160} (-0,910 + \dots - 2,203) = 0,0380 \quad (12)$$

$$\mu_{1KU4} = \frac{1}{N_1} \sum_{q=1}^{N_1} x_q = \frac{1}{159} \sum_{q=1}^{159} x_q = \frac{1}{159} (-0,601 + \dots - 0,255) = -0,0311 \quad (13)$$

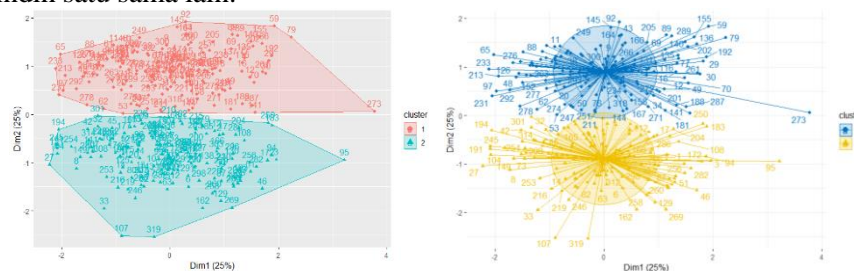
$$\mu_{2KU4} = \frac{1}{N_2} \sum_{q=1}^{N_2} x_q = \frac{1}{160} \sum_{q=1}^{160} x_q = \frac{1}{160} (-1,242 + \dots + 1,565) = 0,0309 \quad (14)$$

**Tabel 4. Nilai Centroid Baru**

| Kluster | KU1     | KU2     | KU3     | KU4     |
|---------|---------|---------|---------|---------|
| 1       | 2,2549  | -0,0995 | -0,0383 | -0,0311 |
| 2       | -2,2408 | 0,0989  | 0,0380  | 0,0309  |

Sumber: Data sekunder, diolah

Setelah nilai *centroid* baru dihitung pada Tabel 4, maka dibandingkan dengan nilai *centroid* sebelumnya. Dikarenakan berdasarkan hasil pengolahan didapatkan nilai *centroid* yang telah sama, maka proses iterasi dihentikan. Hasilnya pun didapatkan pengelompokan menjadi 2 kluster. Nilai *within cluster sum of squares* pada kluster pertama sebesar 1197,108 dan kluster 2 sebesar 1394,067. Hasil visualisasi *scatter plot* kluster ditunjukkan pada Gambar 5 yang memperlihatkan bahwa jarak antara variabel satu dengan yang lainnya dekat sehingga visualisasi antar kluster tidak tumpang tindih satu sama lain.

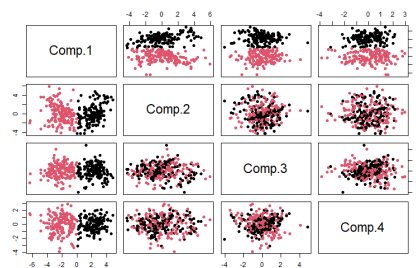


Sumber: Numbeo 2022, diolah.

**Gambar 5. Cluster Plot**

Kluster yang dapat diinterpretasikan dengan baik yaitu ketika varians *within cluster* dapat sehomogen mungkin dan varians *between cluster* seheterogen mungkin. Pada visualisasi kluster Gambar 6, antar titik observasi pada suatu kluster terlihat menggerombol dan berdekatan yang berarti varians *within cluster* yang didapatkan cukup homogen. Sementara itu, titik-titik observasi antar kluster terlihat berjauhan yang menandakan varians *between cluster* juga cukup heterogen. Maka dari itu, analisis dapat dilanjutkan dengan *cluster profiling* serta interpretasi antar kluster.





Sumber: Numbeo 2022, diolah.

**Gambar 6. Scatter Plot**

*Cluster profiling* adalah salah satu cara untuk menginterpretasikan hasil/output akhir pengklasteran dengan mengidentifikasi rata-rata atau *centroid* dari setiap atribut di dalam suatu kluster dan dibandingkan dengan *centroid* pada kluster lainnya. Pada Tabel 5 terlihat bahwa kluster 1 menunjukkan indikator sosial (A1, A2, dan A3) yang tinggi pada variabel indeks kejahatan serta rendah pada variabel indeks keamanan dan indeks kesehatan; indikator ekonomi (B1, B2, B3, B4, B5, B6, dan B7) yang rendah, indikator lingkungan (C1 dan C2) yang tinggi; indikator mobilitas dan transportasi (D1, D2, dan D3) yang tinggi, serta indikator profil internasional (E1) yang rendah. Berdasarkan indikator-indikator penyusunnya, *Smart City* yang masuk dalam kluster 1 merupakan *Smart City* dengan tingkatan yang rendah (*low*). Kluster 2 menunjukkan indikator sosial (A1, A2, dan A3) yang rendah pada variabel indeks kejahatan serta tinggi pada variabel indeks keamanan dan indeks kesehatan; indikator ekonomi (B1, B2, B3, B4, B5, B6, dan B7) yang tinggi, indikator lingkungan (C1, dan C2) yang rendah; indikator mobilitas dan transportasi (D1, D2, dan D3) yang rendah, serta indikator profil internasional (E1) yang tinggi. Sehingga ditinjau dari indikator-indikator penyusunnya, *Smart City* yang masuk ke dalam kluster 2 berdasarkan *motion index* adalah *Smart City* dengan tingkatan yang tinggi (*high*).

**Tabel 5. Cluster Profiling**

| Kluster | A1      | A2      | A3     | B1       | B2      | B3     | B4      | B5     |
|---------|---------|---------|--------|----------|---------|--------|---------|--------|
| 1       | 46,438  | 53,561  | 62,762 | 23,900   | 0,952   | 36,961 | 13,377  | 25,792 |
| 2       | 42,009  | 57,991  | 70,946 | 23,101   | 1,996   | 70,556 | 38,615  | 55,609 |
| Kluster | B6      | B7      | C1     | C2       | D1      | D2     | D3      | E1     |
| 1       | 33,6201 | 43,436  | 64,746 | 5297,660 | 160,981 | 38,220 | 176,963 | 29,129 |
| 2       | 66,066  | 100,592 | 42,060 | 5315,734 | 141,329 | 33,798 | 162,460 | 67,492 |

Sumber: Data sekunder, diolah

### Hierarchical Clustering

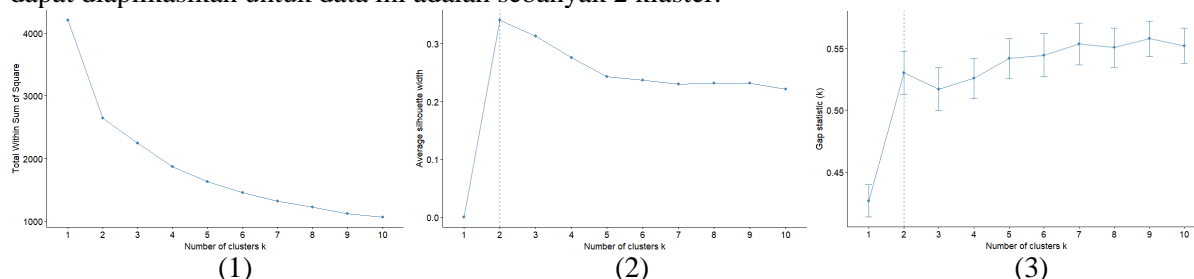
*Hierarchical Clustering* akan mengelompokkan objek-objek yang mirip pada hierarki yang berdekatan dan objek-objek yang tidak mirip pada hirarki yang berjauhan [19]. Metode ini memiliki dua macam strategi yaitu *agglomerative hierarchical clustering* dan *divisive hierarchical clustering*. Di mana *agglomerative hierarchical clustering* menggunakan cara *bottom-up* yaitu di awal objek memiliki kluster sendiri sendiri, kemudian digabung menjadi lebih besar, sedangkan *divisive hierarchical clustering* menggunakan cara *top-down* dengan konsep sebaliknya dari *bottom-up*. Dalam menghitung kemiripan, terdapat beberapa cara seperti *single linkage*, *complete linkage*, *average linkage*, *centroid*, dan metode *Ward*. Sehingga pada pada tahap ini perlu dilihat penggunaan metode yang tepat pada data yang digunakan. Dikarenakan sebelumnya tidak diketahui metode yang akan menghasilkan kluster terbaik berdasarkan *hierarchical clustering*, maka dapat menghitung koefisien *agglomerative* dari setiap metode, yaitu metrik yang mengukur kekuatan kluster, semakin dekat nilai ini dengan 1, maka kluster yang dihasilkan semakin baik. Tabel 6 memperlihatkan bahwa metode varians minimum *Ward* menghasilkan koefisien *agglomerative* tertinggi, sehingga metode tersebut digunakan untuk *hierarchical clustering*.

**Tabel 6. Koefisien Agglomerative**

| Metode                         | Average | Single  | Complete | Ward    |
|--------------------------------|---------|---------|----------|---------|
| Koefisien <i>agglomerative</i> | 0,84495 | 0,72730 | 0,92014  | 0,98201 |

Sumber: Data sekunder, diolah

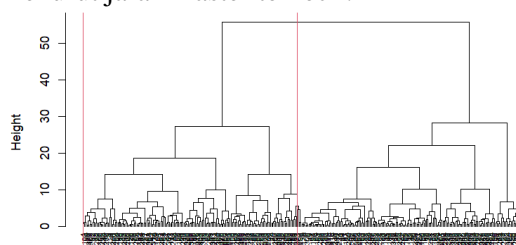
Cara menentukan jumlah kluster yang optimal pada *hierarchical clustering* sama dengan *k-means clustering* yaitu dengan melihat *within sum of square*, *silhouette*, dan *gap statistic* pada Gambar 7. Dari total *within sum of square*, kurva melandai ketika k adalah 2, artinya jumlah kluster yang dapat dipertimbangkan adalah sebanyak 2 kluster. Pada rata-rata lebar *silhouette*, kurva mencapai titik maksimum ketika k berjumlah 2, artinya jumlah kluster yang disarankan adalah sebanyak 2 kluster. Nilai *gap statistic* paling besar adalah ketika jumlah kluster adalah 2, artinya jumlah kluster yang dapat diaplikasikan untuk data ini adalah sebanyak 2 kluster.



Sumber: Numbeo 2022, diolah.

**Gambar 7. Total Within Sum of Square (1), Rata-Rata Silhouette Width (2), Gap Statistic (3) Untuk Setiap Jumlah Kluster**

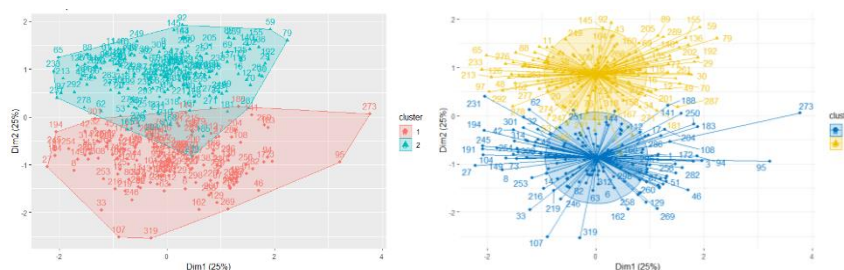
Peneliti memutuskan 2 kluster sebagai jumlah kluster yang harus dibentuk pada *hierarchical clustering* karena hasil *within sum of square*, *silhouette*, dan *gap statistic* menunjukkan bahwa kluster optimumnya adalah 2. Jarak yang digunakan dalam pengolahan data adalah jarak *Euclidean*. Peneliti menggunakan *Ward* karena selain memiliki koefisien *agglomerative* tertinggi dibandingkan dengan metode lainnya, namun juga dapat meminimalkan jumlah kesalahan dalam varian kluster. Hasil pengelompokan didapatkan menurut jarak kluster terkecil.



Sumber: Numbeo 2022, diolah.

**Gambar 8. Dendrogram**

Setiap daun di bagian bawah *dendrogram* mewakili observasi dalam kumpulan data asli. Didapatkan *dendrogram* pengelompokan menjadi 2 kluster seperti pada Gambar 8. Selain dari *dendrogram*, hasil pengelompokan juga dapat divisualisasikan melalui *cluster plot* pada Gambar 9. Hasil pengelompokan yang dilakukan menampilkan bahwa antara kluster 1 dan 2 menunjukkan jumlah observasi yang hampir merata yaitu 155 dan 164 kota. Terdapat perbedaan pengelompokan level tinggi (*high*) atau (*low*) pada 16 kota dari *k-means clustering* dan *hierarchical clustering*, yaitu kota Campinas, Jeddah, Kaohsiung, Limassol, Ljubljana, Mangalore, Manila, Nicosia, Port of Spain, Prague, Quebec City, Rome, Sharjah, Tallinn, Valencia, dan Vilnius. Hasil dari pengelompokan untuk kluster 1 memiliki karakteristik yang hampir sama dengan kluster 2 metode *k-means clustering* dan sebaliknya. Sehingga dengan metode *hierarchical Ward clustering* untuk kluster 1 merupakan level tinggi (*high*) dan kluster 2 adalah level rendah (*low*).



Sumber: Numbeo 2022, diolah.

**Gambar 9. Cluster Plot**

### Pemilihan Metode Terbaik

*K-means* dan *hierarchical Ward clustering* memang menghasilkan 2 kelompok dengan karakteristik yang mirip, namun perlu dibuktikan metode yang paling tepat digunakan dalam praktik data yang tersedia. Peneliti menggunakan 2 cara dalam memilih metode terbaik yaitu *bootstrapping* untuk menguji kecukupan klaster dan evaluasi model menggunakan *silhouette score*, *calinski-harabasz index*, dan *davies-bouldin index*. Teknik *bootstrapping* dipilih karena dapat mengambil sampel ulang secara acak dengan penggantian dari kumpulan data dan menjalankan kembali algoritma pengelompokan. Hal ini akan memperlihatkan seberapa stabil hasil penghitungan dengan metode yang digunakan. Setelah itu menggunakan ukuran Jaccard, di mana jika observasi mengelompok dengan cara yang sama hampir sepanjang waktu terlepas dari distribusi *resampling*, maka rata-rata nilai Jaccard akan tinggi. Jaccard ialah metode yang diterapkan pada dua buah objek untuk menghitung tingkat kesamaannya (*similarity*) [20]. Ketika kira-kira ada kemungkinan yang sama bahwa berbagai observasi akan mengelompok bersama, maka nilai rata-rata Jaccard akan menjadi sekitar 0,5. Maka dari itu peneliti akan menganggap klaster cukup jika rata-rata nilai Jaccard 0,6 atau lebih. Dari hasil perhitungan didapatkan nilai seperti pada Tabel 7. Berdasarkan hasil tersebut, metode *k-means clustering* adalah metode yang lebih baik untuk diimplementasikan pada data yang peneliti gunakan dalam kasus ini.

Tabel 7. Ukuran Nilai Jaccard

| <i>K-Means Clustering</i> |                   |             | <i>Hierarchical Ward Clustering</i> |                   |             |
|---------------------------|-------------------|-------------|-------------------------------------|-------------------|-------------|
| Cluster                   | Rata-rata Jaccard | Instability | Cluster                             | Rata-rata Jaccard | Instability |
| 1                         | 0,9832            | 0,006       | 1                                   | 0,6789            | 0,278       |
| 2                         | 0,9836            | 0,006       | 2                                   | 0,7048            | 0,177       |

Sumber: Data sekunder, diolah

Cara kedua dilakukan dengan cara evaluasi model menggunakan *silhouette score*, *calinski-harabasz index*, dan *davies-bouldin index* [21]. *Silhouette score* dihitung dengan cara pada persamaan (15).

$$s = \frac{b-a}{\max(a,b)} \quad (15)$$

Di mana *a* merupakan jarak rata-rata antara observasi dengan semua titik lain dalam klaster yang sama dan *b* yaitu jarak rata-rata antara observasi dengan semua titik lain dalam klaster terdekat. *Silhouette score* memiliki kisaran nilai antara -1 (pengelompokan yang salah) hingga 1 (pengelompokan yang sangat padat). Jika *silhouette score* bernilai di sekitar 0 mengartikan bahwa kluster yang tercipta kemungkinan tumpang tindih [22]. Sehingga model terbaik apabila nilai *silhouette score* bernilai lebih tinggi dan mendekati 1. Selanjutnya, *calinski-harabasz index* didapatkan melalui cara seperti persamaan (16).

$$s = \frac{tr(B_k)}{tr(W_k)} \times \frac{n-k}{k-1} \quad (16)$$

$$B_k = \sum_{q=1}^k n_q (c_q - c_x)(c_q - c_x)^T \quad (17)$$

$$W_k = \sum_{q=1}^k \sum_{x \in C_q} (x - C_q)(x - C_q)^T \quad (18)$$

Sekumpulan data *X* berukuran *n* yang telah dikelompokkan menjadi *k* kluster,  $tr(B_k)$  merupakan *trace* dari matriks disperse antar kluster,  $tr(W_k)$  adalah *trace* matriks,  $c_q$  yaitu kumpulan titik dalam kluster *q*,  $c_q$  adalah pusat kluster *q*,  $c_x$  adalah pusat data *X*, dan  $n_q$  adalah jumlah titik dalam kluster *q*. Nilai *calinski-harabasz index* yang lebih tinggi merupakan model dengan klaster yang lebih baik karena klaster yang diciptakan lebih jelas [23]. Kemudian evaluasi dengan *davies-bouldin index* yang menyatakan rata-rata *similarity* antar kluster merupakan ukuran membandingkan jarak antar kluster dengan ukuran kluster itu sendiri [24]. *Davies-bouldin index* dihitung dengan cara pada persamaan (19).

$$R_{ij} = \frac{s_i + s_j}{d_{ij}} \tag{19}$$

$$s = \frac{1}{k} \sum_{i=1}^k \max_{j=1}^k R_{ij} \tag{20}$$

Di mana  $R_{ij}$  adalah ukuran ‘kesamaan’,  $s_i$  atau diameter kluster adalah rata-rata jarak tiap titik dalam kluster  $i$  dan pusat kluster  $I$ ,  $d_{ij}$  adalah jarak antara pusat kluster  $i$  dan pusat kluster  $j$ . *Davies-bouldin index* yang mendekati 0 menunjukkan partisi yang lebih baik [25].

**Tabel 8. Evaluasi Model *K-Means Clustering* dan *Hierarchical Ward Clustering***

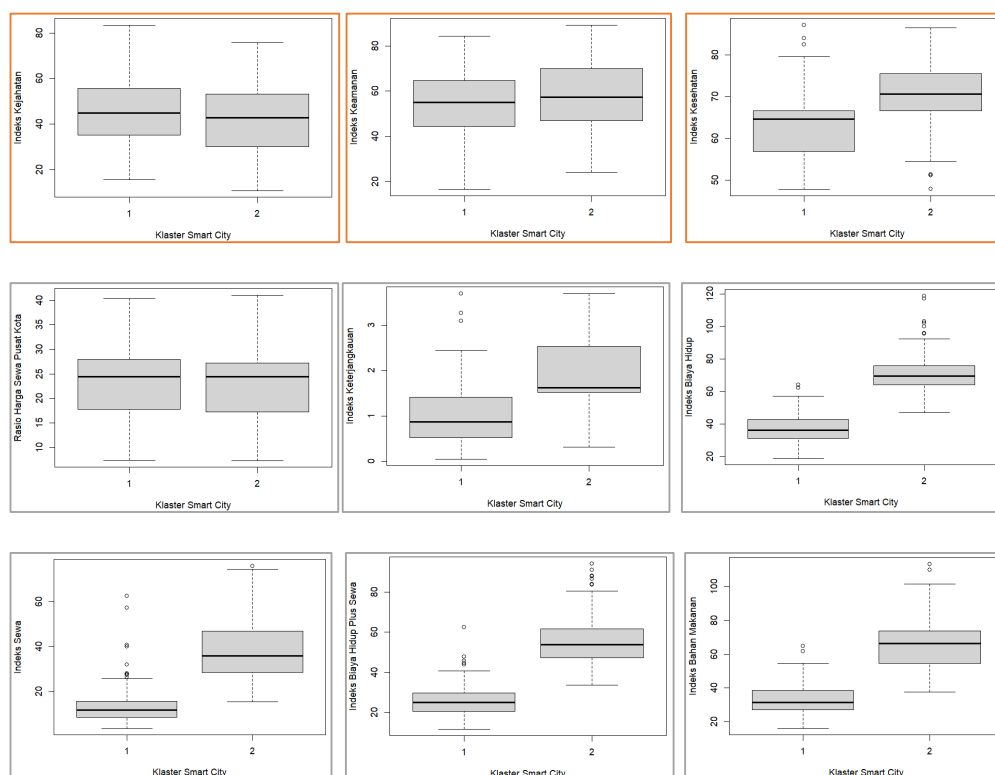
| Evaluasi Model                 | <i>K-Means Clustering</i> | <i>Hierarchical Ward Clustering</i> |
|--------------------------------|---------------------------|-------------------------------------|
| <i>Silhouette Score</i>        | 0,351                     | 0,341                               |
| <i>Calinski-Harabasz Index</i> | 197,665                   | 187,292                             |
| <i>Davies-Bouldin Index</i>    | 1,265                     | 1,301                               |

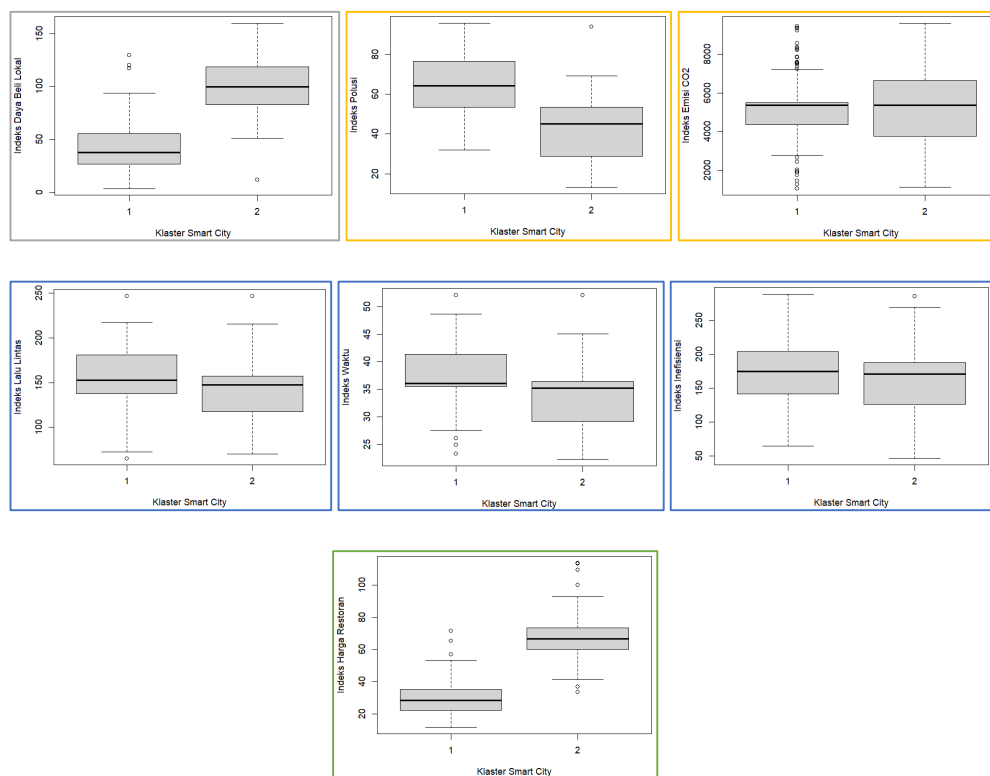
Sumber: Data sekunder, diolah

Berdasarkan Tabel 8 di atas, *silhouette score* dan *calinski-harabasz index* metode *k-means clustering* lebih tinggi dibandingkan metode *hierarchical ward clustering*. Terlebih lagi *davies-bouldin index* metode *k-means clustering* lebih kecil dibandingkan metode *hierarchical ward clustering*. Hal tersebut menunjukkan bahwa metode *k-means clustering* dengan 2 kluster optimum adalah metode terbaik. Sehingga, melalui *bootstrapping* dalam menguji kecukupan kluster dan evaluasi model menggunakan *silhouette score*, *calinski-harabasz index*, dan *davies-bouldin index* keduanya memberikan kesimpulan bahwa metode *k-means clustering* adalah yang paling baik digunakan dalam praktik kasus ini.

### Analisis Deskriptif

Setelah terpilih *k-means clustering* sebagai metode terbaik yang digunakan. Maka dilakukan analisis lebih jauh pada setiap kluster. Dalam melihat lebih detail mengenai karakteristik untuk setiap kluster maka disajikan *boxplot* untuk setiap variabel pada indikator pada penelitian.





Sumber: Numbeo 2022, diolah.

**Gambar 10. Boxplot Setiap Variabel Per Klaster**

Berdasarkan *boxplot* pada Gambar 10, kluster 2 merupakan kumpulan *Smart City* dengan level tinggi (*high*), karakteristiknya adalah pada indikator sosial rata-rata indeks kejahatan lebih rendah dari kluster 1, indeks keamanan lebih tinggi dari kluster 1 walaupun perbedaannya tidak terpaut jauh. Berbeda dengan rata-rata indeks kesehatan kluster 2 yang terpaut jauh lebih tinggi dari kluster 1, hal ini menandakan bahwa kualitas keseluruhan sistem kesehatan, peralatan, personel, biaya, dan lainnya dari kluster 2 lebih unggul dari kluster 1. Terkait dengan indikator ekonomi, hampir semua variabel dari indikator ekonomi di kluster 2 memiliki rata-rata yang lebih tinggi dari kluster 1 seperti indeks keterjangkauan, indeks biaya hidup, indeks sewa, indeks biaya hidup plus sewa, indeks bahan makanan dan indeks daya beli. Hal tersebut menunjukkan bahwa biaya hidup di *Smart City* level tinggi (*high*) jauh lebih besar dan tinggi daripada *Smart City* di level rendah (*low*), hal ini pun diimbangi dengan kemampuan penduduk yang tinggal di sana, karena memiliki indeks daya beli yang tinggi artinya mereka memiliki pekerjaan dengan gaji yang bernilai lebih tinggi sehingga mampu dalam memenuhi kebutuhannya. Ternyata untuk variabel indeks harga sewa rumah di setiap pusat kota memiliki rata-rata yang hampir sama dengan kluster 1. Sehingga dapat disimpulkan terkait harga sewa di pusat kota memiliki harga mirip secara internasional. Perbedaan terletak pada kebutuhan sandang, pangan, sekunder dan tersier lainnya. Dilihat dari indeks lingkungannya, rata-rata indeks polusi di kluster 2 lebih rendah daripada di kluster 1. Artinya *Smart City* level tinggi (*high*) dapat mengatasi atau mengurangi polusi yang dihasilkan dari kegiatan yang ada di perkotaan jauh lebih baik dari level rendah (*low*). Untuk rata-rata indeks  $CO_2$  pada kedua level *Smart City* hampir sama, dimungkinkan polusi yang ada di *Smart City* level rendah (*low*) banyak disumbangkan oleh gas selain  $CO_2$ . Mengenai indeks mobilitas dan transportasi ketiga variabel yaitu indeks lalu lintas, indeks waktu, dan indeks inefisiensi memperlihatkan rata-rata yang mirip pada kedua klaster. Hal ini menandakan adanya ketidakpuasan konsumsi waktu dan inefisiensi lalu lintas yang sama pada kedua klaster. Indikator terakhir yaitu indikator profil internasional yang diwakilkan oleh variabel harga restoran di mana pada *Smart City* level tinggi (*high*) memiliki harga restoran yang tinggi dengan perbedaan yang jauh jika dibandingkan dengan *Smart City* level rendah (*low*). Hal ini menunjukkan perbandingan harga makanan dan minuman di restoran dan bar di kota dengan harga di Kota New York sebagai kota acuan. Setelah diketahui bahwa Kota New York masuk dalam kluster 2 yaitu *Smart City* level tinggi (*high*), sehingga pada *Smart City* level rendah harga makanan dan minuman itu tidak tinggi, maka sesuai dengan harga bahan makanannya pada indikator ekonomi. Kota-kota yang masuk

<http://sistemasi.ftik.unisi.ac.id>

dalam *Smart City* level tinggi (*high*) pada pengelompokan ini seperti Aalborg, Amsterdam, Barcelona, Brazil, Beirut, Berlin, Birmingham, Brisbane, Brussels, Canberra, Chicago, Denver, Dubai, Edinburgh, Frankfurt, Hamilton, Hong Kong, Indianapolis, Las Vegas, London, Los Angeles dan lainnya. Di samping itu, kota-kota yang masuk dalam kelompok *Smart City* level rendah seperti Alexandria, Almaty, Baghdad, Bali, Bandung, Bangkok, Cape Town, Delhi, Guangzhou, Istanbul, Jakarta, Kuala Lumpur, Lviv, Malaga, Manila, Minsk, dan lainnya. Tentunya negara-negara dengan *Smart City* di level rendah (*low*) dapat meningkatkan sumber daya kota yang mereka miliki serta memaksimalkan fungsinya dan juga dapat minilik strategi pembangunan, teknologi informasi, keamanan dan kenyamanan penduduknya terhadap kota lain di kelompok *Smart City* level tinggi (*high*) dengan karakteristik yang hampir sama.

## 5 Kesimpulan

Tujuan dari *Smart City* tentunya yang paling utama adalah meningkatkan rasa aman dan nyaman bagi penduduknya, memudahkan mereka dalam menjangkau segala kebutuhan baik dari segi fasilitas maupun barang dan jasa. Namun, dikarenakan terdapat banyak perbedaan pengelolaan dari setiap kota-kota besar yang ada di dunia akan menghasilkan output yang berbeda pula. Penelitian ini berusaha mengelompokkan *Smart City* yang ada di dunia menjadi 2 level yaitu level tinggi (*high*) dan rendah (*low*). Hasil pengelompokan yang didapatkan akan lebih baik menggunakan metode *k-means clustering* daripada *hierarchical clustering*. Pengelompokan ini dilakukan dengan harapan adanya perbaikan dari *Smart City* di level rendah (*low*) untuk memperbaiki tata kelola kotanya baik secara fisik maupun administratif. Saran peneliti terhadap hasil yang didapatkan kepada *Smart City* level rendah (*low*) adalah dengan meningkatkan indeks kesehatan dan mobilitas transportasi yang ada, seperti memperbaiki cara kerjanya agar lebih terintegrasi antara infrastruktur fasilitas dan pelayanannya. Peningkatan kedua indeks tersebut dapat meningkatkan indeks ekonomi yang cukup memiliki ketimpangan pada dua klaster yang terbentuk. Pengelompokan didasari pada indikator yang tertuang dalam *motion index*, namun tidak semua indikator dibahas dalam penelitian ini. Atas hal tersebut peneliti selanjutnya dapat memperbaiki kekurangan tersebut sehingga dapat melengkapi indikator yang belum terpenuhi.

## Referensi

- [1] M. Angelidou, "Smart city policies: A spatial approach," *Cities*, vol. 41, pp. S3–S11, Jul. 2014, doi: 10.1016/j.cities.2014.06.007.
- [2] E. Muller, J.-A. Héraud, F. Ardizzon, and J. Pellegrin, "Smart cities and clusters," 2019.
- [3] L. Leydesdorff and M. Deakin, "The triple-helix model of smart cities: A neo-evolutionary perspective," *Journal of Urban Technology*, vol. 18, no. 2, pp. 53–63, Apr. 2011, doi: 10.1080/10630732.2011.601111.
- [4] M. De Jong, S. Joss, D. Schraven, C. Zhan, and M. Weijnen, "Sustainable-smart-resilient-low carbon-eco-knowledge cities; Making sense of a multitude of concepts promoting sustainable urbanization," *J Clean Prod*, vol. 109, pp. 25–38, Dec. 2015, doi: 10.1016/j.jclepro.2015.02.004.
- [5] J. Braverman and H. Todosow, "The vision of a smart city The vision of a smart city View project," 2000. [Online]. Available: <https://www.researchgate.net/publication/241977644>
- [6] T. Hertz, "Knowledge-Based Urban Development," 2016.
- [7] C. Cantuarias-Villessuzanne, R. Weigel, and J. Blain, "Clustering of european smart cities to understand the cities' sustainability strategies," *Sustainability (Switzerland)*, vol. 13, no. 2, pp. 1–20, Jan. 2021, doi: 10.3390/su13020513.
- [8] I. Digel, D. Mussabalina, M. Urdabayev, N. Nurmukhametov, and A. Akparova, "Evaluating development prospects of smart cities: Cluster analysis of Kazakhstan's regions," *Problems and Perspectives in Management*, vol. 20, no. 4, pp. 76–87, 2022, doi: 10.21511/ppm.20(4).2022.07.
- [9] N. Tri, S. Saptadi, P. Chyan, and V. P. Taga, "Using K-Means Algorithm to Investigate Community Behavior in Treating Waste toward Smart City," *International Journal on Advanced Science Engineering Information Technology*, vol. 11, no. 4, pp. 1455–1462, 2021.

- [10] A. G. G and B. U. D, “Cluster Analysis of The Regions of Kazakhstan by The Level of Innovative Development,” *International Research Journal*, vol. 9, no. 111, pp. 133–137, 2021.
- [11] M. Kubina, D. Šulyová, and J. Vodák, “Comparison of smart city standards, implementation and cluster models of cities in North America and Europe,” *Sustainability (Switzerland)*, vol. 13, no. 6, Mar. 2021, doi: 10.3390/su13063120.
- [12] P. Berrone and J. E. Ricart, “IESE Cities in Motion Index,” Barcelona, 2022.
- [13] P. Chévez, D. Barbero, I. Martini, and C. Discoli, “Application of the k-means clustering method for the detection and analysis of areas of homogeneous residential electricity consumption at the Great La Plata region, Buenos Aires, Argentina,” *Sustain Cities Soc*, vol. 32, pp. 115–129, Jul. 2017, doi: 10.1016/j.scs.2017.03.019.
- [14] N. M. Garcia, “Smart Cities Indexes 2022,” Nov. 09, 2022. <https://www.kaggle.com/datasets/nicolasmiguezarcia/smart-cities-kpis-indexes> (accessed Nov. 09, 2022).
- [15] Numbeo, “Quality of Life,” 2022. [https://www.numbeo.com/quality-of-life/indices\\_explained.jsp](https://www.numbeo.com/quality-of-life/indices_explained.jsp) (accessed Nov. 09, 2022).
- [16] I. Ghozali, *Aplikasi Analisis Multivariate dengan Program IBM SPSS 19*. Semarang: Badan Penerbit Universitas Diponegoro, 2011.
- [17] Neil. J. Salkind, Ed., *Encyclopedia of Research Design*. United Kingdom: SAGE Publications, 2010.
- [18] A. R. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. New Jersey: Pearson Education Inc, 2007.
- [19] S. Pramana, B. Yuniarto, S. Mariyah, I. Santoso, and R. Nooraeni, *Data Mining dengan R: Konsep Serta Implementasi*. Bogor: IN MEDIA, 2018.
- [20] Sugiyanto, B. Surarso, and A. Sugiharto, “Analisa Peforma Metode Coseine dan Jaccard pada Pengujian Kesamaan Dokumen,” *Jurnal Masyarakat Informatika*, vol. 5, no. 10, pp. 1–8, 2014.
- [21] A. M. Sikana and A. W. Wijayanto, “Analisis Perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical Clustering,” *Jurnal Ilmu Komputer*, vol. 14, no. 2, pp. 66–78, 2021.
- [22] P. J. Rousseeuw, “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,” 1987.
- [23] T. Caliński and J. Harabasz, “A Dendrite Method Foe Cluster Analysis,” *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, 1974, doi: 10.1080/03610927408827101.
- [24] D. L. Davies and D. W. Bouldin, “A Cluster Separation Measure,” *IEEE Trans Pattern Anal Mach Intell*, vol. PAMI-1, no. 2, pp. 224–227, 1979, doi: 10.1109/TPAMI.1979.4766909.
- [25] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, “On clustering validation techniques,” *J Intell Inf Syst*, vol. 17, no. 2–3, pp. 107–145, Dec. 2001, doi: 10.1023/A:1012801612483.