

## ***Cluster Analysis of Hospital Inpatient Service Efficiency Based on BOR, BTO, TOI, AvLOS Indicators using Agglomerative Hierarchical Clustering***

Analisis Kluster Efisiensi Pelayanan Rawat Inap Rumah Sakit Berdasarkan Indikator BOR, BTO, TOI, AvLOS Menggunakan Agglomerative Hierarchical Clustering

**Tresna Maulana Fahrudin<sup>1</sup>, Prismahardi Aji Riyantoko<sup>2</sup>, Kartika Maulida Hindrayani<sup>3</sup>,  
Made Hanindia Prami Swari<sup>4</sup>**

<sup>1,2,3</sup> Sains Data, Universitas Pembangunan Nasional Veteran Jawa Timur, Indonesia

<sup>4</sup> Informatika, Universitas Pembangunan Nasional Veteran Jawa Timur, Indonesia

<sup>1\*</sup> tresna.maulana.ds@upnjatim.ac.id, <sup>2</sup>prismahardi.aji.ds@upnjatim.ac.id,

<sup>3</sup>kartika.maulida.ds@upnjatim.ac.id, <sup>4</sup>madehanindia.fik@upnjatim.ac.id

### ***Informasi Artikel***

*Received: May 2021*

*Revised: June 2021*

*Accepted: July 2021*

*Published: August 2021*

### ***Abstract***

*Purpose: The research proposed an approach for grouping hospital inpatient service efficiency that have the same characteristics into certain clusters based on BOR, BTO, TOI, and AvLOS indicators using Agglomerative Hierarchical Clustering.*

*Design/methodology/approach: Applying Agglomerative Hierarchical Clustering with dissimilarity measures such as single linkage, complete linkage, average linkage, and ward linkage.*

*Findings/result: The experiment result has shown that ward linkage was given a quite good score of silhouette coefficient reached 0.4454 for the evaluation of cluster quality. The cluster formed using ward linkage was more proportional than the other dissimilarity measures. Ward linkage has generated cluster 0 consists of 23 members, cluster 1 consists of 34 members, while both of cluster 2 and 3 consists of only 1 member respectively. The experiment reported that each cluster had problems with inpatient indicators that were not ideal and even exceeded the ideal limit, but cluster 0 generated the ideal BOR and TOI parameters, both reached 52.17% (12 of 23 hospital inpatient) and 78.36% (18 of 23 hospital inpatient) respectively.*

*Originality/value/state of the art: Based on previous research, this study provides an alternative to produce more proportional, representative and quality clusters in mapping hospital inpatient service efficiency that have the same*

---

*characteristics into certain clusters using Agglomerative Hierarchical Clustering Method compared to the K-means Clustering Method which is often trapped in local optima.*

---

### Abstrak

*Keywords:* cluster analysis, hospital, inpatient, agglomerative hierarchical clustering, silhouette coefficient  
Kata kunci: analisis klaster, rumah sakit, rawat inap, agglomerative hierarchical clustering, silhouette coefficient

**Tujuan:** Penelitian ini mengajukan suatu pendekatan untuk memetakan efisiensi pelayanan rawat inap rumah sakit yang memiliki karakteristik yang sama ke dalam klaster tertentu berdasarkan indikator BOR, BTO, TOI dan AvLOS menggunakan Metode *Agglomerative Hierarchical Clustering*.

**Perancangan/metode/pendekatan:** Menerapkan Metode *Agglomerative Hierarchical Clustering* dengan ukuran ketidakmiripan *single linkage, complete linkage, average linkage*, dan *ward linkage*.

**Hasil:** Penelitian ini menunjukkan bahwa *ward linkage* memberikan nilai *silhouette coefficient* yang cukup baik sebesar 0.4454 terhadap evaluasi kualitas klaster. Klaster yang terbentuk menggunakan ukuran ketidakmiripan *ward linkage* cukup proporsional dibandingkan ukuran ketidakmiripan lainnya. *Ward linkage* menghasilkan klaster 0 terdiri dari 23 anggota, klaster 1 terdiri dari 34 anggota, klaster 2 dan 3 masing-masing hanya terdiri dari 1 anggota. Hampir tiap klaster memiliki permasalahan dengan indikator rawat inap yang tidak ideal bahkan juga melebihi batas ideal, namun klaster 0 menunjukkan parameter BOR dan TOI yang dikategorikan ideal dimana masing-masing mencapai 52.17% (12 dari 23 rawat inap) dan 78.36% (18 dari 23 rawat inap).

**Keaslian/ state of the art:** Berdasarkan penelitian terdahulu, penelitian ini memberikan alternatif untuk menghasilkan klaster yang lebih proporsional, representatif dan berkualitas dalam memetakan efisiensi layanan rawat inap rumah sakit yang memiliki karakteristik yang sama ke dalam klaster tertentu menggunakan Metode *Agglomerative Hierarchical Clustering* dibandingkan Metode *K-means Clustering* yang sering terjebak pada *local optima*.

---

## 1. Pendahuluan

Rumah sakit sebagai tempat penyelenggaraan upaya kesehatan dengan tujuan mewujudkan derajat kesehatan bagi masyarakat serta memelihara dan meningkatkan pelayanan kesehatan. Beberapa upaya kesehatan direalisasikan oleh pihak rumah sakit dalam bentuk promotif,

preventif, kuratif dan rehabilitatif. Secara promotif, artinya dilakukan upaya pemeliharaan dan peningkatan kesehatan, sedangkan secara preventif, artinya dilakukan upaya pencegahan penyakit. Upaya kesehatan secara kuratif artinya penyembuhan penyakit bagi pasien, dan secara rehabilitatif artinya upaya pemulihan bagi pasien [1]. Salah satu bentuk kegiatan untuk merealisasikan upaya kesehatan tersebut adalah pelayanan rawat inap. Unit rawat inap berperan penting terhadap penyembuhan dan pemulihan pasien sehingga harus dikelola dengan baik serta adanya kontrol melalui penilaian internal dan eksternal.

Terdapat dua sisi untuk meninjau efisiensi pengelolaan rumah sakit, yakni sisi medis untuk meninjau mutu pelayanan medis, dan sisi ekonomi untuk meninjau pendayagunaan sarana yang tersedia. Sarana utama yang tersedia untuk unit rawat inap adalah ketersediaan tempat tidur pasien, penggunaan tempat tidur ini yang akan dinilai tingkat keefisiensannya yang berhubungan dengan lamanya pasien dirawat, kosongnya tempat tidur serta terisinya tempat tidur oleh pasien baru. Dalam bidang ilmu kesehatan dikenal grafik Barber Johnson untuk membandingkan tingkat efisiensi penggunaan tempat tidur dari suatu unit maupun antar unit. Rawat inap menjadi objek perhatian khusus dari suatu rumah sakit, karena tidak hanya memiliki kontribusi signifikan terhadap pendapatan rumah sakit, tetapi juga menunjukkan efisiensi pelayanan masuk dan keluarnya pasien. Rawat inap memiliki indikator untuk kinerja meliputi empat parameter yakni *Bed Occupancy Ratio* (BOR), *Bed Turn Over* (BTO), *Turn Over Interval* (TOI), dan *Average Length of Stay* (AvLOS) [2].

Parameter BOR menunjukkan persentase (%) pemakaian tempat tidur pada waktu tertentu, sedangkan parameter BTO menunjukkan frekuensi (kali) pemakaian tempat tidur pada waktu tertentu. Parameter TOI menunjukkan rata-rata (hari) dari tempat tidur telah terisi lalu kosong hingga ditempati oleh pasien baru atau pasien berikutnya, sedangkan parameter AvLOS menunjukkan rata-rata (hari) lama rawatan pasien [3]. Berdasarkan grafik Barber Johnson terkait standar ideal dan efisiensi pengelolaan tempat tidur rumah sakit telah ditentukan yakni nilai BOR berkisar 75-85%, nilai BTO berkisar minimal 30 kali, nilai TOI berkisar 1-3 hari, dan nilai AvLOS berkisar 3-12 hari [4]. Sedangkan standar ideal pengelolaan tempat tidur rumah sakit berdasarkan Departemen Kesehatan telah ditentukan yakni nilai BOR berkisar 60-85%, nilai BTO berkisar minimal 40-50 kali, nilai TOI berkisar 1-3 hari, dan nilai AvLOS berkisar 6-9 hari [1].

Berdasarkan data profil kesehatan Kota Surabaya tahun 2017 [5] terdapat 59 rumah sakit yang tercatat sesuai indikator kinerja pelayanan rumah sakit menggunakan parameter BOR, BTO, TOI dan AvLOS. Namun nilai parameter tersebut berupa representasi angka yang belum memberikan informasi tentang bagaimana analisis hasil evaluasi indikator kinerja pelayanan 59 rumah sakit yang ada di Kota Surabaya dan guna menggali pengetahuan baru yang dapat memberikan gambaran seberapa baik kinerja tiap rumah sakit. Salah satu cara untuk mendapatkan informasi dan pengetahuan baru dari data tersebut yang belum dikategorikan yakni dengan cara memetakan data indikator kinerja tiap rumah sakit tersebut ke dalam kluster atau kelompok tertentu. Teknik pengelompokan yang dapat digunakan untuk mendapatkan sejumlah kluster data yakni berdasarkan kemiripannya [6].

Penelitian di bidang kesehatan dan rumah sakit telah diprediksi akan meningkat seiring dengan bertumbuhnya data dari waktu ke waktu. Data medis bersumber dari data profil pasien, riwayat kontrol, tindak lanjut kontrol kesehatan dan data klinis [7]. Tseng, dkk meneliti tentang analisis kluster sumber data medis di rumah sakit distrik di Taiwan menggunakan data tahun 2007-2011.

Penelitiannya melaporkan bahwa terjadi persaingan yang cukup kompetitif pasar rumah sakit distrik yang harus diseimbangkan dengan layanan kesehatannya. Penelitiannya menyimpulkan bahwa terjadi peningkatan jumlah rumah sakit swasta dan dibandingkan rumah sakit umum yang justru terjadi penurunan di Taiwan. Selain itu, klasterisasi diterapkan untuk mengetahui jumlah tempat tidur yang digunakan, jumlah dokter, jumlah staf medis, jumlah pasien yang dipulangkan, jumlah pasien rawat jalan dan unit gawat darurat. Metode klasterisasi digunakan adalah metode K-means of Ward's Minimum Variance Iterations [8].

Zhang, dkk mengimplementasikan *hierarchical cluster analysis* dalam penelitian klinis dengan populasi heterogen, mereka fokus untuk mengkaji hasil visualisasinya melalui bahasa pemrograman R. Penelitiannya menunjukkan eksperimen berupa data *unsupervised* yang terdiri dari variabel prediktor bertipe numerik seperti tekanan darah, detak jantung, suhu dan pemeriksaan laboratorium, dimana hasil klaster dihubungkan dengan suatu variabel biner seperti kategori sembuh atau meninggal. Penelitiannya juga menerapkan *statistical quantity, heat map, scatter plot on heat map*, dan dikaitkan dengan jumlah kejadian [9]. Beberapa penelitian lain yang berkaitan dengan penerapan K-means untuk klasterisasi data kesehatan pada 3 area utama yakni pengendalian infeksi rumah sakit, pemeringkatan rumah sakit, dan identifikasi pasien beresiko tinggi [10]. Peneliti lain juga meneliti tentang sistem rekomendasi rumah sakit [11], dan implementasi metode Hierarchical Clustering pada pengelompokan dokumen [12].

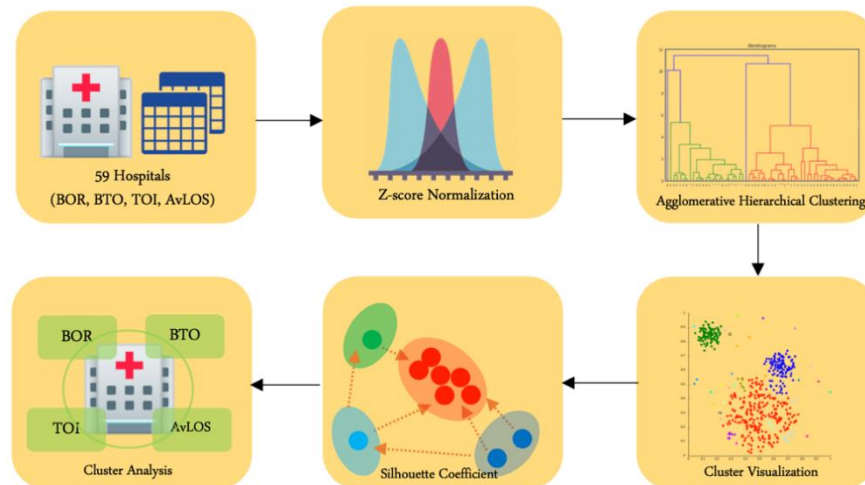
Oleh karena itu, penelitian ini bertujuan untuk menganalisis dan mengelompokkan rumah sakit di Surabaya berdasarkan indikator pelayanan rawat inap menggunakan metode Agglomerative Hierarchical Clustering. Pada tahap awal penelitian ini, sejumlah indikator pelayanan rawat inap seperti parameter BOR, BTO, TOI dan AvLOS diperoleh dari data profil kesehatan Kota Surabaya. Tahap kedua, nilai parameter yang bertipe numerik tersebut dinormalisasi menggunakan metode Z-score untuk mendapatkan nilai skala yang seimbang. Tahap ketiga, metode Agglomerative Hierarchical Clustering diterapkan untuk mengelompokkan rumah sakit berdasarkan indikator pelayanan rawat inapnya. Tahap keempat, mengevaluasi kualitas klaster yang terbentuk menggunakan silhouette coefficient. Tahap kelima, Hasil klaster yang terbentuk dianalisis untuk mendapatkan *insight* atau pengetahuan baru sebagai bahan evaluasi serta strategi rawat inap rumah sakit kedepannya.

## 2. Metode/Perancangan

Profil kesehatan kota Surabaya berisi data atau informasi yang menampilkan situasi dan kondisi kesehatan masyarakat di kota Surabaya, salah satunya adalah evaluasi kualitas rawat inap rumah sakit. Dinas Kesehatan Kota Surabaya juga melakukan monitoring dan evaluasi terhadap kinerja dari penyelenggaraan pelayanan minimal bidang kesehatan dan penyelenggaraan Rencana Pembangunan Jangka Menengah (RPJMD) tiap tahunnya dan dituangkan menjadi laporan profil kesehatan kota Surabaya sebagai bahan acuan untuk perencanaan program di tahun berikutnya. Untuk menunjang penelitian ini, profil kesehatan kota Surabaya dimanfaatkan untuk diperoleh data efisiensi layanan rawat inap sebanyak 59 rumah sakit yang terdiri dari rumah sakit umum dan swasta.

Desain sistem penelitian ini ditunjukkan pada Gambar 1, dimana terdapat data efisiensi layanan rawat inap dari 59 rumah sakit yang terdiri dari 4 parameter yakni BOR, BTO, TOI dan AvLOS. Representasi angka tiap parameter pada rumah sakit cukup bervariasi, hal ini menunjukkan terdapat kondisi beberapa parameter yang belum memenuhi syarat minimum atau

standar rumah sakit. Tantangannya adalah bagaimana sistem mampu memetakan rumah sakit yang memiliki kondisi atau kualitas yang sama ke dalam kluster tertentu. Selanjutnya 4 parameter tersebut nilainya cukup beragam, perlu dilakukan normalisasi agar mendapatkan skala yang seimbang.



Gambar 1. Desain Sistem Penelitian yang Diusulkan

## 2.1 Z-score Normalization

Salah satu metode normalisasi yang cukup baik untuk menyeimbangkan skala data yakni menggunakan Z-score.

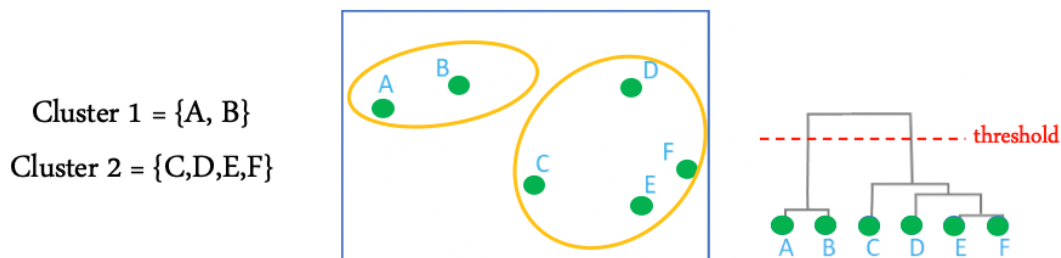
Z-score menggunakan rata-rata dan standar deviasi untuk mengubah nilai skala tiap fitur pada data asli dan disiapkan sebagai data latih dalam fase pembelajaran. Rata-rata dan standar deviasi dihitung tiap individu fitur dengan tujuan memenuhi asumsi yang mendasari analisis ragam. Formula Z-score ditunjukkan pada Pers. (1).  $x'$  adalah data ternormalisasi,  $x_i$  adalah suatu nilai dari sebuah variabel,  $\mu_i$  adalah rata-rata dari suatu variabel, dan  $\sigma_i$  adalah standar deviasi dari suatu variabel [13].

$$x' = \frac{x_i - \mu_i}{\sigma_i} \quad (1)$$

Mengubah nilai variasi data asli ke Z-score tidak mengubah bentuk distribusi, tetapi hanya mengubah rentang nilai tertentu. Z-score menunjukkan seberapa jauh data observasi terhadap rata-ratanya. Ini akan menguntungkan bagi tiap fitur agar tidak saling mendominasi nilainya misal saat perhitungan jarak. Nilai Z-score berkisar dari -3 standar deviasi yang berada pada sisi kiri dari kurva distribusi normal hingga +3 standar deviasi yang berada pada sisi kanan dari kurva distribusi normal.

## 2.2 Agglomerative Hierarchical Clustering

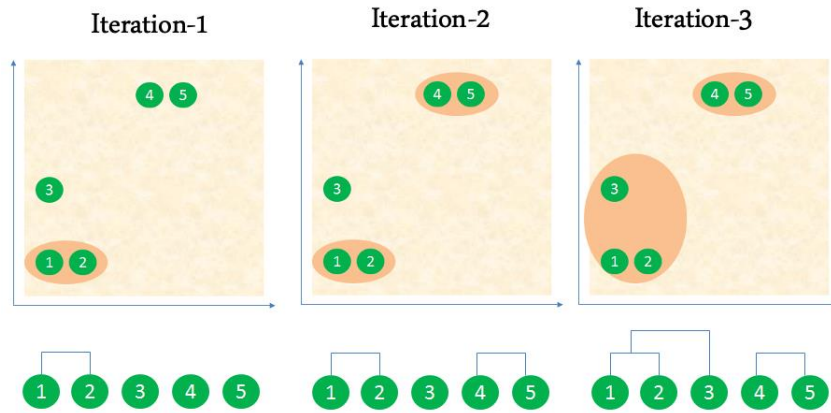
Setelah dilakukan normalisasi, dataset akan masuk ke tahap pembentuk model kluster menggunakan sebuah algoritma *machine learning*. Pada penelitian ini dipilih kelompok metode Hierarchical Clustering sebagai metode pilihan yang baik untuk memetakan 59 rumah sakit tersebut ke dalam kluster tertentu. Hierarchical Clustering adalah metode analisis kelompok yang berusaha untuk membangun sebuah hirarki kelompok data. Metode hirarki mengelompokkan titik-titik data ke pohon kluster yang disebut dendrogram, dimana menampilkan hubungan antar kluster. Dendrogram terdiri dari akar pohon dan daun pohon, dimana akar pohon mewakili 1 kluster yang berisikan semua titik data, sedangkan daun pohon mewakili satu titik data.



**Gambar 2.** Pemetaan Kluster pada Dendrogram

Pada Gambar 2 mengilustrasikan pemetaan kluster pada dendrogram, dimana terdapat 6 titik data yang terdiri dari {A, B, C, D, E, dan F} akan dipetakan menjadi beberapa kluster. Sebuah *threshold* berupa garis horizontal akan digunakan sebagai penentu jumlah kluster pada saat pohon hirarki terbentuk. Dengan memotong pohon pada level yang diinginkan menggunakan *threshold* tersebut, maka didapatkan pengklusteran titik-titik data ke dalam grup-grup yang terpisah. Misalnya, didapatkan 2 kluster, kluster 1 adalah titik data A dan B, sedangkan kluster 2 adalah titik data C, D, E dan F. Pengaturan *threshold* akan berpengaruh terhadap jumlah kluster yang terbentuk. Jika *threshold* diletakkan di bagian atas pohon hirarki, maka jumlah kluster yang didapatkan semakin sedikit. Sebaliknya, jika *threshold* diletakkan di bagian bawah pohon hirarki, maka jumlah kluster yang didapatkan semakin banyak.

Strategi pengelompokkan Hierarchical Clustering umumnya terdapat 2 jenis, yaitu Agglomerative Hierarchical Clustering (AHC) dan Divisive Hierarchical Clustering (DHC). Pada penelitian ini Agglomerative Hierarchical Clustering dipilih karena memiliki kelebihan untuk menghasikan kluster yang berkualitas [14] dan komputasi waktu yang hemat untuk melakukan klusterisasi data dengan jumlah kecil. Pada Gambar 3 diilustrasikan bagaimana cara kerja Agglomerative Hierarchical Clustering, dimana kluster-kluster kecil yang pada awalnya berisi titik data individual diaglomerasikan (dikelompokkan) menjadi yang lebih besar. Hal ini berbanding terbalik dengan cara kerja Divisive Hierarchical Clustering, dimana pengklusteran dimulai dengan 1 kluster yang berisi semua set data, kemudian kluster ini dibagi menjadi beberapa sub-kluster yang lebih kecil dan seterusnya sedemikian sampai 1 sub-kluster berisi hanya 1 titik data. Divisive Hierarchical Clustering tidak praktis karena teknik perhitungannya yang memakan waktu dan biaya komputasi (*cost*) yang besar [15].



**Gambar 3.** Cara Kerja Metode Agglomerative Hierarchical Clustering

Prosedur algoritma Agglomerative Hierarchical Clustering secara umum mengikuti Tabel 1 berikut:

**Tabel 1.** Prosedur Secara Umum Algoritma Agglomerative Hierarchical Clustering

Langkah	Keterangan
Langkah 1	Hitung matrik jarak antar data
Langkah 2	Gabungkan dua kelompok terdekat berdasarkan parameter kedekatan yang ditentukan
Langkah 3	Perbarui matrik jarak antar data untuk merepresentasikan kedekatan di antara kelompok baru dan kelompok yang masih tersisa
Langkah 4	Ulangi langkah 2 dan 3 hingga hanya satu kelompok yang tersisa
Langkah 5	Tentukan <i>threshold</i> untuk mendapat jumlah klaster yang diinginkan

Salah satu teknik untuk menghitung matriks jarak antar data yakni menggunakan Euclidean Distance. Euclidean Distance mengukur jarak spasial antara 2 objek yang memiliki sejumlah fitur, lalu jarak disimpan ke dalam sebuah matriks yang terdiri dari baris dan kolom yang menunjukkan besaran jarak di antara 2 objek. Misal terdapat 2 titik yakni  $u = (x_1, y_1)$  dan  $v = (x_2, y_2)$ , maka Euclidean Distance akan mengukur jarak antara  $u$  dan  $v$  mengikuti Pers. (2) berikut [16].

$$D(x, y) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2)$$

Selain itu, jika terdapat lebih dari 2 dimensi atau disebut sebagai  $n$  dimensi, misal  $a = (x_1, x_2 \dots x_n)$  dan  $b = (y_1, y_2 \dots y_n)$ , maka Pers. (2) dapat digeneralisasikan menjadi Pers. (3). Euclidean Distance akan menghitung jarak antara titik  $a$  dan titik  $b$ .

$$D(a, b) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (3)$$

Pers. (3) menunjukkan dimensi semakin bertambah sejumlah  $n$ , maka dapat disederhanakan dengan Pers. (4), dimana perhitungan jarak antar 2 titik data akan dijumlahkan sesuai dimensi ke- $i$  hingga  $n$  dimensi terpenuhi.

$$D(a, b) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

Setelah matriks jarak didapatkan menggunakan Euclidean Distance, Agglomerative Hierarchical Clustering menggabungkan individu-individu klaster. Dalam klaster, untuk menggabungkan dua atau lebih objek menjadi satu klaster, biasanya digunakan ukuran kemiripan dan ketidakmiripan. Pada kasus ini, Agglomerative Hierarchical Clustering menggunakan ukuran ketidakmiripan, dimana semakin tidak mirip dua objek yang diukur maka semakin rendah peluang untuk dikelompokkan dalam satu klaster. Beberapa metode dalam ukuran ketidakmiripan seperti yang diilustrasikan pada Gambar 4 antara lain (a) single linkage, (b) complete linkage, (c) average linkage dan (d) ward linkage. Masing-masing metode memiliki kelebihan dan kekurangannya masing-masing.

### 2.2.1 Ukuran Ketidakmiripan dengan Single Linkage

Single linkage menggunakan *minimum distance between cluster* [14] [15], metode ini bekerja dengan baik terhadap data yang terpisah dengan baik dan klaster yang konsentris. Kekurangannya adalah jika data memiliki pola *noise*, maka akan menghasilkan klaster yang tidak beraturan dan hirarki miring [12]. Formula single linkage ditunjukkan pada Pers. (5), dimana  $d(C_{ij})$  adalah sebuah jarak dari 2 klaster ke  $i$  dan  $j$ , lalu dihitung jaraknya terhadap  $C_k$  yaitu klaster berikutnya yang akan digabungkan berdasarkan jarak minimum. Terbentuklah calon klaster baru yang memiliki jarak minimum yakni  $d(C_i, C_k)$  atau  $d(C_j, C_k)$  yang akan dipilih.

$$d(C_{(ij)}, C_k) = \min\{d(C_i, C_k), d(C_j, C_k)\} \quad (5)$$

### 2.2.2 Ukuran Ketidakmiripan dengan Complete Linkage

Complete linkage menggunakan *maximum distance between cluster* [14] [15], metode ini bekerja dengan baik terhadap data yang memiliki *outlier* dan menghasilkan klaster yang sangat padat. Namun metode ini kurang cocok terhadap klaster yang konsentris [12]. Formula complete linkage ditunjukkan pada Pers. (6), dimana  $d(C_{ij})$  adalah sebuah jarak dari 2 klaster ke  $i$  dan  $j$ , lalu dihitung jaraknya terhadap  $C_k$  yaitu klaster berikutnya yang akan digabungkan berdasarkan jarak maksimum, kebalikan dari single linkage. Terbentuklah calon klaster baru yang memiliki jarak maksimum yakni  $d(C_i, C_k)$  atau  $d(C_j, C_k)$  yang akan dipilih.

$$d(C_{(ij)}, C_k) = \max\{d(C_i, C_k), d(C_j, C_k)\} \quad (6)$$



### 2.2.3 Ukuran Ketidakmiripan dengan Average Linkage

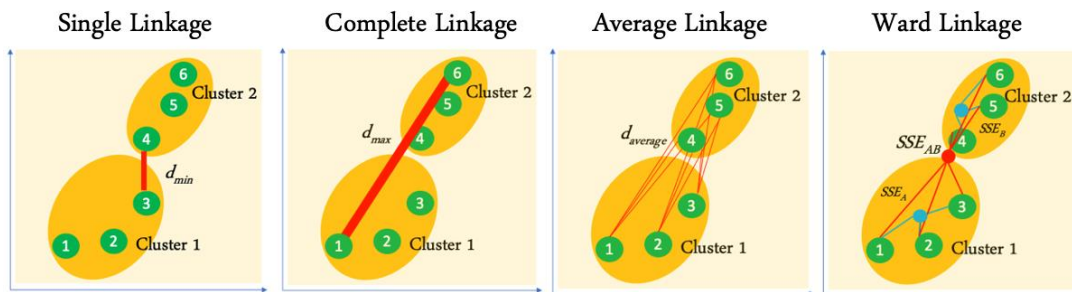
Average linkage menggunakan *mean distance between cluster* [14] [15], metode ini bekerja dengan baik terhadap data yang berbentuk gumpalan [12]. Metode ini relatif yang terbaik, namun harus dibayar dengan waktu komputasi yang paling tinggi dibandingkan dengan metode-metode *hierarchical* yang lain. Formula average linkage ditunjukkan pada Pers. (7), dimana  $d(C_{ij})$  adalah sebuah jarak dari 2 kluster ke  $i$  dan  $j$ , lalu dihitung jaraknya terhadap  $C_k$  yaitu kluster berikutnya yang akan digabungkan berdasarkan jarak rata-rata. Terbentuklah calon kluster baru yang memiliki jarak rata-rata dari penjumlahan  $d(C_i, C_k)$  dan  $d(C_j, C_k)$  lalu dibagi 2.

$$d(C_{(ij)}, C_k) = \frac{d(C_i, C_k) + d(C_j, C_k)}{2} \quad (7)$$

### 2.2.4 Ukuran Ketidakmiripan dengan Ward Linkage

Ward linkage menggunakan fungsi objektif yang optimal sebagai dasar untuk memilih pasangan kluster. Fungsi objektif yang digunakan bebas selama mencerminkan sebuah tujuan [17], misal berupa evaluasi pembentukan model. Pada kasus klusterisasi ini, fungsi objektif yang digunakan adalah SSE (*sum of squared error*), dimana akan dihitung selisih dari total SSE lokal yakni masing-masing kluster terhadap SSE globalnya. Formula ward linkage ditunjukkan pada Pers. (8), dimana  $d(C_{ij})$  adalah sebuah jarak dari 2 kluster ke  $i$  dan  $j$ , lalu dihitung jaraknya terhadap  $C_k$  yaitu kluster berikutnya yang akan digabungkan berdasarkan jarak minimum SSE. Terbentuklah calon kluster baru yang memiliki jarak minimum SSE yang didapatkan dari jumlah SSE lokal yakni  $d(C_i, C_k)$  dan  $d(C_j, C_k)$  dikurangi SSE global  $d(C_i, C_j)$ , dimana  $n_i, n_j, n_k$  adalah banyaknya data dalam set  $i, j$  dan  $k$ .

$$d(C_i \cup C_j, C_k) = \frac{n_i + n_k}{n_i + n_j + n_k} d(C_i, C_k) + \frac{n_j + n_k}{n_i + n_j + n_k} d(C_j, C_k) - \frac{n_k}{n_i + n_j + n_k} d(C_i, C_j) \quad (8)$$



Gambar 4. (a) Single Linkage, (b) Complete Linkage, (c) Average Linkage, (d) Ward Linkage

### 2.2.5 Evaluasi Kualitas Kluster Menggunakan Silhouette Coefficient

Semua model kluster yang terbentuk menggunakan 4 jenis linkage yang sudah dijelaskan di atas akan dievaluasi kualitas klusternya menggunakan silhouette coefficient. Metode evaluasi ini merupakan interpretasi dan validasi konsistensi dalam kelompok data. Silhouette coefficient

mengkombinasikan antara kohesi dan separasi [14]. Untuk menghitung silhouette coefficient, Langkah pertama, hitung jarak rata-rata objek ke- $i$  terhadap semua objek lain di dalam klasternya, ini biasa disebut  $a_i$  mengikuti Pers. (9), dengan  $j$  adalah objek lain dalam satu cluster A dan  $d(i, j)$  adalah jarak antara objek  $i$  dengan  $j$ .

$$a_i = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (9)$$

Langkah kedua, hitung jarak rata-rata objek ke- $i$  terhadap semua klaster lain yang tidak mengandung objek tersebut. Temukan nilai minimum yang terkait dengan semua klaster. ini disebut  $b_i$  mengikuti Pers. (10).

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \quad (10)$$

dengan  $d(i, C)$  adalah jarak rata-rata objek ke- $i$  dengan semua objek pada cluster lain  $C$  dimana  $A \neq C$ .

$$b(i) = \min_{C \neq A} d(i, j) \quad (11)$$

Langkah ketiga, hitung silhouette coefficient objek ke- $i$  menggunakan Pers. (12).

$$s_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (12)$$

Nilai silhouette coefficient bervariasi mulai dari -1 hingga 1, artinya jika nilai silhouette coefficient semakin mendekati nilai 1, maka semakin baik pengelompokan data dalam satu klaster. Sebaliknya, jika nilai silhouette coefficient semakin mendekati nilai -1, maka semakin buruk pengelompokan data di dalam satu klaster [18].

### 3. Hasil dan Pembahasan

Pada subbab ini akan dibahas lebih rinci hasil eksperimen yang telah dilakukan termasuk penjelasan bentuk dan struktur dataset, hasil penskalaan data menggunakan normalisasi Z-score, pemodelan klaster menggunakan Agglomerative Hierarchical Clustering, evaluasi kualitas klaster menggunakan silhouette coefficient dan analisis klaster. Hasil eksperimen ini diharapkan dapat menemukan *insight* yang dapat memberikan manfaat bagi evaluasi rumah sakit ke depannya.

#### 3.1. Dataset Indikator Layanan Rawat Inap Rumah Sakit

Dataset yang digunakan pada penelitian ini didapatkan dari profil kesehatan Kota Surabaya pada tahun 2017 yang terdiri dari 59 rumah sakit beserta parameter BOR, BTO, TOI dan AvLOS. Parameter ini cukup mewakili untuk melihat indikator pelayanan rawat inap dari berbagai rumah sakit, namun belum diketahui rawat inap rumah sakit mana saja yang memiliki kondisi yang sama dalam suatu klaster. Tabel 2 menunjukkan indikator pelayanan rawat inap rumah sakit di

Surabaya tahun 2017 berupa data terstruktur, dimana nama rumah sakit sebagai indeks, sedangkan 4 parameter lainnya sebagai fitur, dan tanpa label kelas (*ground truth*).

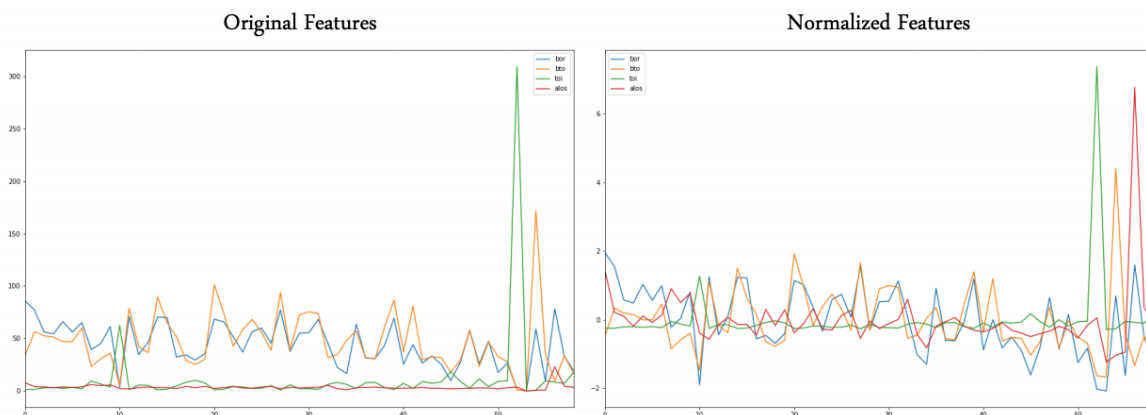
**Tabel 2.** Data Efisiensi Layanan Rawat Inap Rumah Sakit di Surabaya Tahun 2017

No.	Nama Rumah Sakit	BOR	BTO	TOI	AvLOS
1.	RSUD Dr Soetomo	85,67	33,94	1,54	7,65
2.	RSUD Dr Moh Soewandhi	77,09	56,6	1,48	4,18
3.	RSUD Haji	56,23	52,6	3,04	3,84
4.	RSUD Bhakti Darma Husada	54,47	51,43	3,23	3
5.	RS Primasatya Husada Citra (PHC)	66,06	47,06	2,63	3,86
6.	RS Tk III Brawijaya	56,16	47,21	3,39	3,33
...	...	...	...	...	...
59.	RS Bedah Surabaya	19,28	16,44	17,92	3,47

BOR dan BTO memiliki skala atau rentang nilai puluhan, namun berbeda halnya dengan TOI dan AvLOS, dimana rentang nilainya kurang dari 10. Hal ini mengakibatkan BOR dan BTO mendominasi selama proses perhitungan jarak, oleh karena itu perlu dilakukan penskalaan agar skala fitur menjadi seimbang.

### 3.2. Normalisasi Data Menggunakan Z-score

Z-score adalah salah satu di antara banyaknya pilihan metode normalisasi, Z-score mampu mengetahui seberapa jauh data observasi terhadap rata-ratanya. Pada Gambar 5(a) terlihat bahwa skala data belum seimbang, dimana terdapat nilai suatu fitur saling berjauhan dengan fitur lainnya, ini sifat alamiah dari variasi nilai fitur asli. Pada Gambar 5(b) terlihat telah dilakukan penskalaan fitur, sehingga nilai-nilai fitur saling berdekatan dalam suatu skala tertentu, dimana Z-score menghasilkan skala data dimulai dari rentang -3 hingga 3.



**Gambar 5.** (a) Skala Data Sebelum Dilakukan Normalisasi, (b) Skala Data Setelah Dilakukan Normalisasi

Hasil normalisasi Z-score tidak mengubah distribusi data, namun mampu mengubah skala data menjadi lebih baik dan seimbang. Hal ini akan menguntungkan agar fitur-fitur tidak saling mendominasi satu sama lain, fitur berada pada skala yang sama dan fitur tetap menjadi pembeda satu sama lain.

### 3.3. Pemodelan Kluster Menggunakan *Agglomerative Hierarchical Clustering*

Untuk memetakan rumah sakit yang memiliki kondisi atau kualitas yang sama berdasarkan indikator pelayanan rawat inap ke dalam kluster tertentu, maka diimplementasikan sebuah model *machine learning* berbasis *unsupervised learning* menggunakan metode *Agglomerative Hierarchical Clustering*. Metode ini cukup baik untuk mengelompokkan suatu individu data yang memiliki kemiripan dengan individu data lainnya berbasis pohon hirarki. Dengan menggunakan hirarki, dapat diketahui dengan mudah kedekatan suatu data beserta jaraknya. Metode *Agglomerative Hierarchical Clustering* memiliki sejumlah linkage untuk mengukur ketidakmiripan suatu data observasi, ini akan membantu untuk memetakan individu data yang tidak mirip dengan individu data lainnya ke dalam kluster lainnya. Beberapa metode ukuran ketidakmiripan yang digunakan pada penelitian ini antara lain *single linkage*, *complete linkage*, *average linkage* dan *ward linkage*.

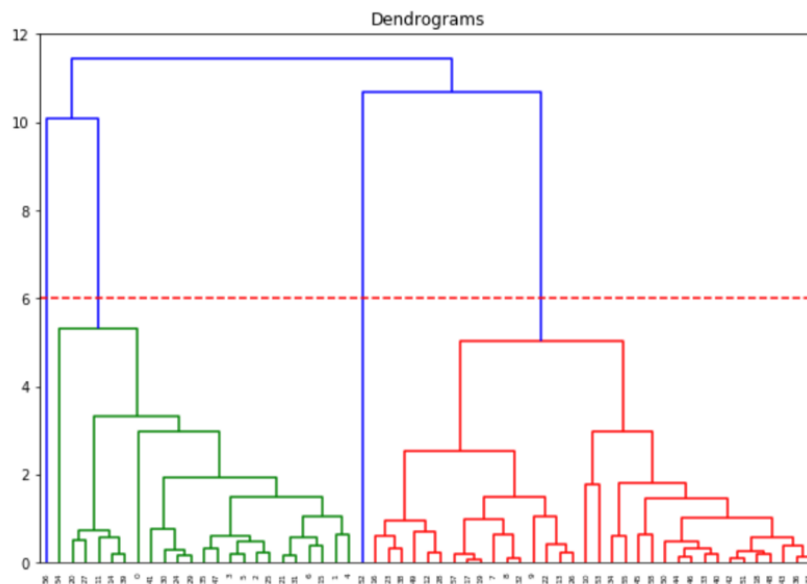
Pada Tabel 3 terlihat bahwa perbandingan proporsi data yang bergabung ke dalam kluster tertentu dari keempat metode linkage, dimana jumlah kluster ditentukan sebanyak 4 kluster. Eksperimen menunjukkan bahwa *single linkage* memetakan 56 rumah sakit atau sebanding dengan 94.9% bergabung ke dalam kluster 0 dari 59 rumah sakit yang ada, sedangkan 1.7% rumah sakit lainnya bergabung ke dalam kluster 1, 2 dan 3 masing-masing. Hal ini juga terjadi pada metode *complete linkage* dan *average linkage*, dimana hasil pengelompokkannya sama dengan *single linkage*. *Single*, *complete* dan *average linkage* menghasilkan kluster yang dominan terhadap kluster tertentu dengan mengambil porsi 94.9% dari data observasi, hal ini menunjukkan ketimpangan jumlah anggota kluster.

**Tabel 3.** Perbandingan Proporsi Anggota Kluster Berdasarkan Ukuran Ketidakmiripan pada Layanan Rawat Inap Rumah Sakit di Surabaya Tahun 2017

Cluster	Single Linkage	Complete Linkage	Average Linkage	Ward Linkage
Cluster 0	56 (94.9%)	56 (94.9%)	56 (94.9%)	23 (38.98%)
Cluster 1	1 (1.7%)	1 (1.7%)	1 (1.7%)	34 (57.62%)
Cluster 2	1 (1.7%)	1 (1.7%)	1 (1.7%)	1 (1.7%)
Cluster 3	1 (1.7%)	1 (1.7%)	1 (1.7%)	1 (1.7%)

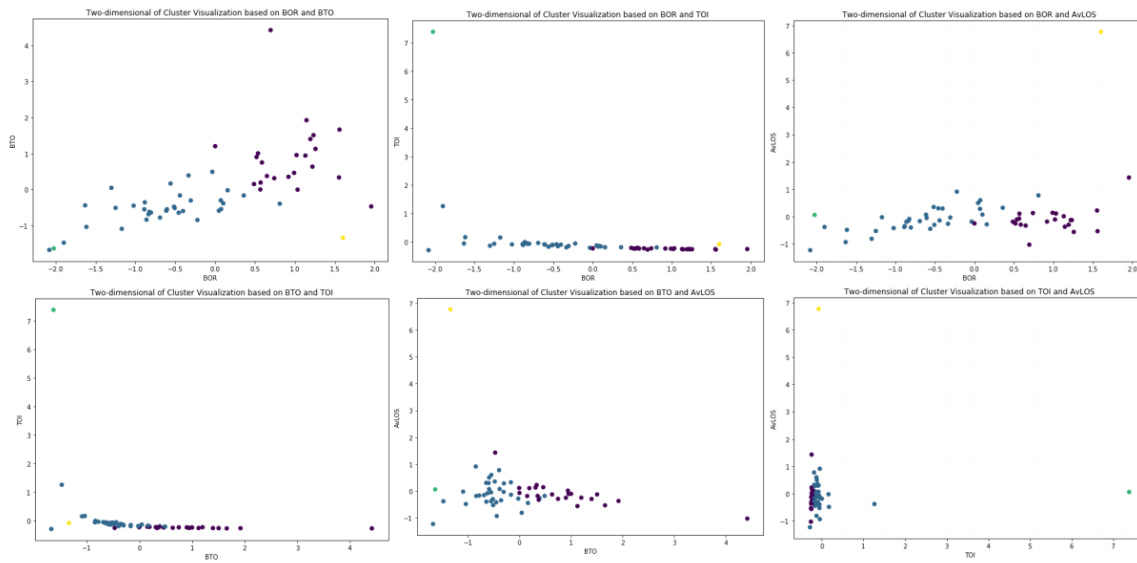
Hal yang berbeda ditunjukkan oleh *ward linkage*, dimana kluster yang terbentuk lebih proporsional. Ada 23 rumah sakit yang bergabung ke dalam kluster 0 atau sebanding dengan 38.98% dari 59 rumah sakit yang ada, 34 rumah sakit bergabung ke dalam kluster 1 atau sebanding 57.62%, dan 1.7% lainnya bergabung ke dalam kluster 2 dan 3 masing-masing. *Ward*

linkage memberikan kinerja yang lebih baik dalam mengelompokkan data ke dalam 4 kluster dibandingkan dengan single, complete dan average linkage. Seperti yang ditunjukkan pada Gambar 6, dendrogram yang terbentuk menggunakan ward linkage untuk memetakan pelayanan rumah sakit berdasarkan indikator rawat inap direpresentasikan dalam dua warna yakni hijau dan merah beserta ada 2 data observasi yang menjadi kluster tersendiri berwarna biru. Garis horizontal berupa *threshold* diletakkan pada axis y untuk memotong pohon dendrogram sesuai jumlah kluster yang diinginkan yakni 4 kluster. Kami menentukan jumlah kluster sebanyak 4 agar dapat mengetahui rumah sakit mana saja yang memiliki kondisi yang sama berdasarkan masing-masing parameter BTO, BOR, TOI dan AvLOS yang ada.



**Gambar 6.** Dendrogram Pemetaan Pelayanan Rumah Sakit Berdasarkan Indikator Rawat Inap Menggunakan Ward Linkage

Gambar 7 menunjukkan hubungan antara dua fitur yang direpresentasikan ke dalam diagram dua dimensi beserta warna klasternya. Visualisasi data ini berdasarkan hasil klusterisasi menggunakan ward linkage, dimana 4 kluster divisualisasikan ke dalam warna biru, ungu, hijau dan kuning. Terlihat pemetaan kluster yang cukup baik dari berbagai sudut pandang 2 fitur yang saling berhubungan pada axis  $x$  dan axis  $y$ . Misal Axis  $x$ : BOR dan Axis  $y$ : BTO, terlihat suatu kluster berwarna biru berkelompok di bagian sisi kiri (axis  $x$ : -2 hingga 1), sedangkan kluster lainnya berwarna ungu berkelompok di bagian sisi kanan (axis  $x$ : 0 hingga 2). Kluster lainnya yang berwarna hijau berada di antara gerombolan kluster berwarna biru di bagian bawah kiri (axis  $x$ : -2.1), sedangkan kluster berwarna kuning berkelompok sendiri di bagian bawah kanan (axis  $x$ : 1.6).



**Gambar 7.** Visualisasi 2-Dimensi Parameter Layanan Rawat Inap Beserta Hasil Klaster

### 3.4. Evaluasi Hasil Klaster Menggunakan Silhouette Coefficient

Kualitas klaster dapat diukur berdasarkan bagaimana sebuah individu data bergabung ke dalam klaster yang memiliki kemiripan tinggi (kohesi) dan menjauh terhadap klaster lain yang memiliki kemiripan yang rendah. Salah satu metode evaluasi untuk kasus klustering adalah silhouette coefficient, Tabel 4 menunjukkan perbandingan antara masing-masing metode linkage. Single, complete, dan average linkage mencapai nilai silhouette coefficient sebesar 0.5998 masing-masing, hanya ward linkage saja yang memberikan nilai silhouette coefficient sebesar 0.4454. Silhouette coefficient memberikan *threshold* antara -1 hingga 1, dimana semakin ke kiri maka semakin buruk klaster yang terbentuk, semakin ke kanan maka semakin baik klaster yang terbentuk. Hasil klasterisasi pelayanan rumah sakit berdasarkan indikator rawat inap pada penelitian ini menunjukkan kualitas klaster yang cukup bagus dengan nilai rata-rata silhouette coefficient sebesar sebesar 0.5612.

**Tabel 4.** Evaluasi Kualitas Klaster Menggunakan Silhouette Coefficient

#	Single Linkage	Complete Linkage	Average Linkage	Ward Linkage
Silhouette Coefficient	0.5998	0.5998	0.5998	0.4454
Average	0.5612			

### 3.5. Insight Hasil Analisis Klaster

Untuk mendapatkan insight dari model klaster telah terbentuk menggunakan metode Agglomerative Hierarchical Clustering, maka dipilih hasil klasterisasi pelayanan rawat inap rumah sakit berdasarkan ukuran ketidakmiripan milik ward linkage. Model ini cukup representatif dan proporsional untuk mengetahui rumah sakit mana saja yang masuk ke klaster

tertentu. Tabel 5 menunjukkan hubungan antara kategori keidealan pengelolaan tempat tidur rumah sakit (berdasarkan departemen kesehatan) dengan indikator kinerja pelayanan rumah sakit menggunakan parameter BOR, BTO, TOI dan AvLOS beserta data observasi yang tersebar pada tiap kluster. Kolom yang diwarnai merah menunjukkan suatu parameter tidak ideal, warna hijau menunjukkan parameter yang ideal dan berwarna oranye menunjukkan melebihi batas ideal.

**Tabel 5.** Analisis Kluster pada Parameter BOR, BTO, TOI, AvLOS

Cluster	Kategori	BOR	BTO	TOI	AvLOS
Cluster 0	Tidak Ideal	10 (43.48%)	1 (4.35%)	2 (8.69%)	22 (95.65%)
	Ideal	12 (52.17%)	2 (8.69%)	18 (78.36%)	1 (4.35%)
	Melebihi Batas Ideal	1 (4.35%)	20 (86.96%)	3 (13.04%)	0
Total		23 (38.98%)			
Cluster 1	Tidak Ideal	33 (97.06%)	27 (79.41%)	1 (2.94%)	33 (97.06%)
	Ideal	1 (2.94%)	4 (11.76%)	0	1 (2.94%)
	Melebihi Batas Ideal	0	3 (8.82%)	33 (97.06%)	0
Total		34 (57.62%)			
Cluster 2	Tidak Ideal	1 (100%)	1 (100%)	0	1 (100%)
	Ideal	0	0	0	0
	Melebihi Batas Ideal	0	0	1 (100%)	0
Total		1 (1.7%)			
Cluster 3	Tidak Ideal	0	1 (100%)	0	0
	Ideal	1 (100%)	0	0	0
	Melebihi Batas Ideal	0	0	1 (100%)	1 (100%)
Total		1 (1.7%)			

Cluster 0 terdiri dari 43.48% BOR dan 95.65% AvLOS tidak ideal, 86.96% BTO melebihi batas ideal, dan terdapat 52.17% BOR dan 78.36% TOI ideal. Cluster 1 terdiri dari 97.06% BOR dan 79.41% BTO tidak ideal, sedangkan 97.06% TOI melebihi batas ideal. Cluster 2 hanya terdiri 1 data observasi, dimana nilai BOR, BTO, AvLOS tidak ideal, sedangkan TOI melebihi batas ideal. Cluster 3 juga hanya terdiri dari 1 data observasi, dimana nilai BTO tidak ideal, BOR ideal, tetapi TOI dan AvLOS melebihi batas ideal. Ini menunjukkan bahwa pemakaian tempat tidur pada waktu tertentu, frekuensi (kali) pemakaian tempat tidur pada waktu tertentu, rata-rata (hari) dari tempat tidur telah terisi lalu kosong hingga ditempati oleh pasien baru atau pasien berikutnya, dan rata-rata (hari) lama rawatan pasien pada beberapa rumah sakit cukup bervariasi kondisinya, ada yang tidak ideal dan juga ada yang melebihi batas ideal. Hanya kluster 0 menunjukkan persentase yang cukup besar rumah sakit yang memiliki nilai BOR dan TOI yang ideal dibandingkan kluster lainnya.

#### 4. Kesimpulan dan Saran

Hasil pada penelitian menunjukkan bahwa Agglomerative Hierarchical Clustering memiliki beberapa jenis ukuran ketidaksamaan antara lain single linkage, complete linkage, average linkage dan ward linkage. Ward linkage bekerja dengan baik untuk melakukan pengelompokan pelayanan rawat inap rumah sakit secara proporsional yang memiliki karakteristik sama dan dapat dikelompokkan ke klaster tertentu berdasarkan indikator BOR, BTO, TOI, dan AvLOS. Walaupun nilai silhouette coefficient dari ward linkage yang mencapai 0.4454 lebih kecil dibandingkan dengan linkage lainnya yang mencapai 0.5998, namun ward linkage tetap digunakan sebagai model klasterisasi data yang dirujuk. Ward linkage menghasilkan klaster yang cukup proporsional dibandingkan linkage lainnya. Eksperimen menunjukkan bahwa klaster 0 terdiri dari 23 rawat inap rumah sakit, klaster 1 terdiri dari 34 rawat inap rumah sakit, klaster 2 dan 3 hanya terdiri dari masing-masing 1 rawat inap rumah sakit. Hampir tiap klaster memiliki permasalahan dengan indikator rawat inap yang tidak ideal bahkan juga melebihi batas ideal, namun yang menarik hanya klaster 0 yang menunjukkan BOR dan TOI yang ideal masing-masing sebesar 52.17% (12:23) dan 78.36% (18:23). Untuk penelitian ke depan, perlu melibatkan sampel data yang lebih banyak mencakup rumah sakit yang ada di Indonesia untuk menilai efisiensi rawat inap berdasarkan indikator yang ada tiap tahun dan menjaga kualitas pelayanan rumah sakit semakin baik.

#### Daftar Pustaka

- [1] S. P. S. Lubis dan C. Astuti, "Analisis Efisiensi Penggunaan Tempat Tidur di RSJ Prof. DR. M. Ildrem Medan Per Ruangan Berdasarkan Indikator Rawat Inap di Triwulan 1 Tahun 2018," *Jurnal Ilmiah Perekam dan Informasi Kesehatan Imelda*, vol. 3, no. 2, pp. 466-472, 2018.
- [2] N. F. Dewi dan S. K. Radityo, "The Performance Analysis of Inpatient Installation at Tria Dipa Hospital with Balanced Scorecard, 2013–2015," *KnE Social Sciences*, vol. 3, no. 11, p. 1566–1583, 2018.
- [3] R. Sidiq dan R. Afrina, "Kajian Efisiensi Pelayanan Rumah Sakit," *Idea Nursing Journal*, vol. 8, no. 1, pp. 29-34, 2017.
- [4] V. Cahyati, H. Rohman dan E. P. Nurcahyati, "Efektitas Kebijakan Rumah Sakit Bhayangkara Polda DIY Dilihat dari Grafik Barber Johnson," dalam *SMIKNAS*, Surakarta, 2019.
- [5] D. K. S. Profil Kesehatan Kota Surabaya Tahun 2017, Surabaya: Dinas Kesehatan Pemerintahan Kota Surabaya, 2017.
- [6] B. Santosa dan A. Umam, *Data Mining dan Big Data Analytics*, Yogyakarta: Penebar Media Pustaka, 2018.
- [7] L. Elezabeth, V. P. Mishra and J. Dsouza, "The Role of Big Data Mining in Healthcare Applications," in *7th International Conference on Reliability, Infocom Technologies and*



- Optimization (ICRITO)*, Amity Institute of Information Technology (AIIT), Noida, India, 2018.
- [8] S.-F. Tseng, T.-S. Lee dan C.-Y. Deng, "Cluster Analysis of Medical Service Resources at District Hospitals in Taiwan, 2007-2011," *Journal of the Chinese Medical Association*, vol. 78, no. 12, pp. 732-745, 2015.
- [9] Z. Zhang, F. Murtagh, S. V. Poucke, S. Lin dan P. Lan, "Hierarchical Cluster Analysis In Clinical Research With Heterogeneous Study Population: Highlighting Its Visualization With R," *Annals of Translational Medicine*, vol. 5, no. 4, p. 75, 2017.
- [10] J. K. Samriya, S. Kumar dan S. Singh, "Efficient K-means Clustering for Healthcare Data," *Advanced Journal of Computer Science and Engineering (AJCST)*, vol. 4, no. 2, pp. 1-7, 2016.
- [11] M. R. Khoie, T. S. Tabrizi, E. S. Khorasani, S. Rahimi dan N. Marhamati, "A Hospital Recommendation System Based on Patient Satisfaction Survey," *Applied Sciences*, vol. 7, no. 10, p. 966, 2017.
- [12] F. H. Saad, O. . I. E. Mohamed dan R. E. Al-Qutaish, "Comparison Of Hierarchical Agglomerative Algorithms For Clustering Medical Documents," *International Journal of Software Engineering & Applications (IJSEA)*, vol. 3, no. 3, pp. 1-15, 2012.
- [13] G. Aksu, C. O. Güzeller dan M. . T. Eser, "The Effect of the Normalization Method Used in Different Sample Sizes on the Success of Artificial Neural Network Model," *International Journal of Assessment Tools in Education*, vol. 6, no. 2, pp. 170-192, 2019.
- [14] P.-N. Tan, M. Steinbach, A. Karpatne dan V. Kumar, *Introduction to Data Mining 2nd Edition*, United States: Pearson, 2019.
- [15] M. Roux, "A Comparative Study of Divisive and Agglomerative Hierarchical Clustering Algorithms," *Journal of Classification*, vol. 35, no. 2, pp. 345-366, 2018.
- [16] M. . D. Malkauthekar, "Analysis of Euclidean Distance and Manhattan Distance Measure in Face Recognition," in *3rd International Conference on Computational Intelligence and Information Technology*, Mumbai, India, 2013.
- [17] J. H. W. Jr., "Hierarchical Grouping to Optimize an Objective Function," *Journal of American Statistical Association*, vol. 58, no. 301, pp. 236-244, 1963.
- [18] M. . Z. Rodriguez, C. . H. Comin, D. Casanova, O. . M. Bruno, D. R. Amancio, L. d. F. Costa dan F. A. Rodrigues, "Clustering Algorithms: A Comparative Approach," *PLoS ONE*, vol. 14, no. 1, pp. 1-34, 2019.