

Analisis Cluster Pada Data Kategorik dan Numerik dengan Pendekatan Cluster Ensemble (Studi Kasus: Puskesmas di Provinsi Kalimantan Timur Kondisi Desember 2017)

Cluster Analysis of Categorical and Numeric Data with Cluster Ensemble Approach (Case Study: Puskesmas in East Kalimantan on December 2017)

Nur Aini Ayu Lestari¹, Memi Nor Hayati², Fidia Deny Tisna Amijaya³

¹Laboratorium Statistika Komputasi FMIPA Universitas Mulawarman

²Laboratorium Statistika Terapan FMIPA Universitas Mulawarman

³Laboratorium Matematika Komputasi FMIPA Universitas Mulawarman

E-mail: nurainilstr@gmail.com

Abstract

Cluster analysis used to process categorical and numerical data at once is Cluster Ensemble algorithm Based on Mixed Data Clustering (algCEBMDC), which is a cluster algorithm with an ensemble cluster approach. The method used for numerical data is Agglomerative Nesting (AGNES) algorithm and for categorical data is the RObust Clustering using linK (ROCK) algorithm. The best clustering method and the optimum number of clusters in the AGNES algorithm is selected based on the maximum Pseudo-F value and the minimum icdrate value. The optimum number of clusters in the ROCK algorithm is selected using the minimum value of ratio S_W/S_B . The purpose of this study was to make a group of 179 Puskesmas in East Kalimantan on December 2017. Based on the results of the analysis, obtained 5 optimum cluster for numerical clustering with the AGNES algorithm and 2 optimum cluster for categorical clustering data with the ROCK algorithm. Final cluster for mixed data clustering obtained 2 optimum cluster at a threshold of 0.2 and 0.3 with value of ratio S_W/S_B is 7.283×10^{-17} . The first cluster consists of 83 Puskesmas and cluster two of 96 Puskesmas.

Keywords: AGNES, algCEBMDC, Cluster Ensemble, Puskesmas, ROCK

Pendahuluan

Pusat Kesehatan Masyarakat (Puskesmas) memegang peranan sebagai dasar pelayanan kesehatan yang menjamin pelayanan kesehatan minimal masyarakat. Puskesmas berperan memberikan pertolongan pertama kepada masyarakat sesuai standar pelayanan kesehatan (Jamkesindonesia, 2019).

Menurut Peraturan Menteri Kesehatan Republik Indonesia Nomor 75 Tahun 2014 untuk melaksanakan upaya kesehatan, Puskesmas harus menyelenggarakan manajemen Puskesmas, pelayanan kefarmasian, pelayanan keperawatan kesehatan masyarakat, dan pelayanan laboratorium. Untuk menyelenggarakan hal tersebut dibutuhkan ketenagaan Puskesmas maupun non medis yang memadai agar pelayanan masyarakat dapat optimal.

Pada tahun 2018 Pemerintah Provinsi Kalimantan Timur akan merevitalisasi pelayanan Puskesmas dengan mengembalikan fungsi awal keberadaannya sebagai Puskesmas yang selalu siap melayani 24 jam (Amirullah, 2017). Demi mengoptimalkan kebijakan tersebut tentu dibutuhkan ketenagaan Puskesmas yang lebih memadai agar pelayanan 24 jam dapat optimal. Agar memudahkan pemerintah Provinsi Kalimantan Timur untuk menetapkan kebijakan yang diambil, perlu dilakukan pengelompokan Puskesmas untuk mengetahui Puskesmas mana

saja yang perlu membenahan jumlah ketenagaan Puskesmas.

Salah satu metode yang umum digunakan untuk pengelompokan objek adalah *clustering*. Metode ini melakukan pengelompokan berdasarkan kemiripan atribut dari data obyek (Han dkk, 2001).

Clustering umumnya diterapkan hanya untuk memproses salah satu tipe data yaitu numerik atau kategorik. Data Puskesmas di Provinsi Kalimantan Timur memiliki tipe kategorik dan numerik, sehingga perlu metode *clustering* yang dapat diterapkan pada kedua tipe data tersebut secara bersamaan. He dkk (2005) mengembangkan sebuah metode pengelompokan untuk data campuran yakni metode *ensemble*. Pengelompokan *ensemble* (*cluster ensemble*) adalah teknik pengelompokan untuk menggabungkan hasil pengelompokan beberapa algoritma pengelompokan untuk mendapatkan kelompok yang lebih baik.

Salah satu jenis algoritma yang dapat digunakan untuk memproses data campuran adalah algoritma *Cluster Ensemble Based Mixed Data Clustering* (algCEBMDC) yaitu suatu algoritma pengelompokan dengan pendekatan *cluster ensemble* (He dkk, 2005).

Analisis Multivariat

Menurut Widarjono (2015) analisis multivariat berasal dari kata *multy* dan *variate* yang berarti alat analisis lebih dari dua variabel. Analisis multivariat berhubungan dengan semua teknik statistik yang secara simultan menganalisis sejumlah pengukuran pada suatu objek. Analisis multivariat sering digunakan untuk memecahkan permasalahan dalam penelitian yang sifatnya sangat kompleks. Analisis ini dikelompokkan menjadi dua kelompok besar yaitu:

1. Metode dependensi (*dependence methods*) adalah yang digunakan untuk menjelaskan atau meramalkan nilai variabel terikat berdasarkan lebih dari satu variabel bebas yang mempengaruhinya. Contoh dari metode ini adalah analisis varians, analisis kovarians, regresi berganda, analisis konjoin dan koreksi kanonik.
2. Metode interdependensi (*interdependence methods*) adalah yang digunakan untuk memberikan arti kepada suatu set variabel (kelompok variabel). Beberapa contoh analisis interdependensi adalah analisis faktor, analisis *cluster*, analisis klasifikasi dan penskalaan multidimensional.

Data Mining

Menurut Tan dkk (2006), *data mining* adalah suatu proses untuk menemukan informasi yang menarik dan tersembunyi dari suatu kumpulan data yang berukuran besar yang tersimpan dalam suatu basis data, *data warehouse* atau tempat penyimpanan data lainnya. Beberapa metode yang sering disebut-sebut dalam literatur *data mining* antara lain *clustering*, *classification*, *association rules mining*, *neural network*, *genetic algorithm* dan lain-lain (Santosa, 2007).

Transformasi Data

Transformasi data digunakan untuk mengubah data ke dalam bentuk yang sesuai untuk suatu analisis. Transformasi data dilakukan baik terhadap data numerik maupun data kategorik. Transformasi data numerik dilakukan dengan standarisasi data. Jika rentang nilai antar variabel memiliki perbedaan skala yang cukup besar yang dapat menyebabkan bias dalam analisis *cluster* maka data asli perlu standarisasi.

Rentang nilai yang besar akan menyebabkan perhitungan jarak menjadi tidak valid (Yulianto & Hidayatullah, 2014). Standarisasi data dapat dilakukan dengan cara semua dimensi atau sub-variabel penyusun atau item ditransformasi ke dalam data standar atau data *Z* (nilai rata-rata sama dengan nol, variansi sama dengan satu dan data tanpa satuan/relatif). Menurut Sartono, dkk (2003) cara menentukan nilai standarisasi adalah dengan menggunakan persamaan berikut

$$Z_{i,l} = \frac{x_{il} - \bar{x}_l}{s_l} \tag{1}$$

dimana

- n : banyaknya data
- x_{il} : data ke- i variabel ke- l
- \bar{x}_l : rata-rata variabel ke- l
- $Z_{i,l}$: standarisasi data untuk data ke- i variabel ke- l
- s_k : standar deviasi variabel ke- l

Analisis Cluster

Clustering atau analisis *cluster* adalah suatu teknik *data mining* untuk mengelompokkan himpunan objek (*dataset*) ke dalam beberapa grup/*cluster* hanya berdasarkan kemiripan karakteristik dari atribut yang dimiliki oleh data objek sedemikian sehingga data objek yang berada di dalam *cluster* yang sama memiliki kemiripan satu sama lain tetapi mereka tidak mirip dengan data objek yang berada dalam *cluster* yang berbeda (Han dkk, 2012).

Metode dalam analisis *cluster* dibagi menjadi dua yaitu, metode hierarki dan non-hierarki. Terkait dengan pengertian dan tujuan dilakukannya analisis *cluster*, dapat dinyatakan bahwa suatu kelompok yang baik adalah kelompok yang memiliki ciri-ciri sebagai berikut (Hair dkk, 2014):

- a. Homogenitas (kesamaan) yang tinggi antar anggota dalam satu kelompok (*within cluster*).
- b. Heterogenitas (perbedaan) yang tinggi antar kelompok yang satu tahap dengan kelompok yang lain (*between cluster*).

Salah satu faktor yang sangat berpengaruh terhadap hasil dari kelompok yang dibentuk antar objek pengamatan adalah jarak antar objek pengamatan (Sharma, 1996). Oleh karena itu, dibutuhkan suatu alat ukur untuk menentukan jarak antar objek lain pengamatan. Salah satu jenis metode pengukuran jarak yang umum digunakan adalah jarak *Euclidean*. Misalkan terdapat dua observasi dengan variabel-variabel berdimensi m , yaitu $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{im}]^T$ dan $\mathbf{x}_j = [x_{j1}, x_{j2}, \dots, x_{jm}]^T$. Konsep jarak *Euclidean* yang mengukur jarak antara observasi \mathbf{x}_i dan \mathbf{x}_j dinyatakan sebagai berikut (Johnson dan Wichern, 2002):

$$d_{ij} = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j)} \tag{2}$$

dengan:

- d_{ij} = jarak antara objek ke- i dan ke- j
- \mathbf{x}_i = vektor objek ke- i
- \mathbf{x}_j = vektor objek ke- j

Cluster Metode Hierarki

Cluster metode hierarki digunakan jika banyak gerombol yang akan dibentuk tidak diketahui sebelumnya. Terdapat dua prosedur pada metode berhierarki yaitu *Agglomerative Hierarchical Clustering* dan *Divisive Hierarchical Clustering* (Mattjik dan Sumertajaya, 2011).

Pada teknik pengelompokan hierarki *agglomerative* setiap objek merupakan satu kelompok tersendiri. Lalu dua kelompok terdekat digabungkan dan seterusnya sehingga diperoleh satu kelompok yang berunsurkan semua objek. Beberapa teknik pengelompokan antar kelompok dinyatakan sebagai berikut (Johnson & Wichern, 2002):

a. *Single Linkage*

Metode *Single Linkage* adalah proses *clustering* yang didasarkan pada jarak terdekat antar objeknya. Jika dua objek terpisah oleh jarak pendek, maka kedua objek tersebut akan bergabung menjadi satu *cluster* dan demikian seterusnya. Pengukuran jarak metode ini dapat ditulis dalam rumus berikut:

$$d_{w(i,j)} = \min\{d_{wi}, d_{wj}\} \tag{3}$$

Dimana $d_{w(i,j)}$ merupakan ukuran kemiripan antara kelompok ke- w dengan kelompok (i,j) yang merupakan penggabungan antara kelompok ke- i dan ke- j .

b. *Complete Linkage*

Pengukuran jarak pada metode *complete linkage* dilakukan berdasarkan jarak maksimum atau terjauh. Rumus yang digunakan dalam menentukan jarak pada metode ini adalah sebagai berikut:

$$d_{w(i,j)} = \max\{d_{wi}, d_{wj}\} \tag{4}$$

Dimana $d_{w(i,j)}$ merupakan ukuran kemiripan antara kelompok ke- w dengan kelompok (i,j) yang merupakan penggabungan antara kelompok ke- i dan ke- j .

c. *Average Linkage*

Kriteria yang digunakan dalam mengukur jarak pada metode ini adalah rata-rata jarak seluruh individu dalam satu kelompok dengan jarak seluruh individu dalam kelompok lain. Dapat dituliskan dalam rumus sebagai berikut:

$$d_{w(i,j)} = \frac{\sum_q \sum_r d_{qr}}{n_{(ij)} n_w} \tag{5}$$

Dimana $d_{w(i,j)}$ merupakan ukuran kemiripan antara kelompok ke- w dengan kelompok (i,j) yang merupakan penggabungan antara kelompok ke- i dan ke- j . $n_{(ij)}$ dan n_w merupakan jumlah pengamatan dalam kelompok ij dan w . Sedangkan, d_{qr} merupakan jarak antara pengamatan ke- q

dalam kelompok ij dengan pengamatan ke- r dalam kelompok w .

Algoritma AGNES

Algoritma AGNES pertama kali diperkenalkan oleh Kaufmann dan Rousseeuw pada tahun 1990. Algoritma AGNES merupakan algoritma *agglomerative hierarchical clustering* yang cukup populer bekerja pada numerik (Han dkk, 2012). Algoritma AGNES dimulai dengan menghitung matriks jarak antar objek, setiap objek berfungsi sebagai *cluster*, kemudian secara bertahap menggabungkan setiap pasangan *cluster* terdekat berdasarkan ukuran jarak dan metode penggabungan yang digunakan, sampai semua *cluster* tergabung dalam satu *cluster*.

Secara formal langkah-langkah algoritma AGNES ditulis oleh Tan dkk (2006) sebagai berikut:

1. Hitung matriks jarak
2. Anggap setiap objek sebagai sebuah *cluster*
3. Gabungkan dua *cluster* terdekat (berdasarkan ukuran jarak dan metode penggabungan yang digunakan)
4. Perbarui matriks jarak untuk mengetahui jarak antar *cluster* baru dengan *cluster* asli.
5. Ulangi langkah 3 dan 4 sampai hanya ada satu *cluster*.

Dalam penelitian ini digunakan ukuran jarak *euclidean* dan metode penggabungan *single linkage*, *complete linkage*, serta *average linkage*.

Algoritma ROCK

Algoritma ROCK pertama kali diperkenalkan oleh Guha dkk pada tahun 1999. Algoritma ROCK merupakan algoritma *agglomerative hierarchical clustering* untuk data kategori. Algoritma pengelompokan hierarki tradisional yang menggunakan jarak antar titik untuk pengelompokan dinilai kurang sesuai untuk data kategori. Oleh karena itu, dibentuk sebuah konsep baru, yaitu *link* untuk mengukur kesamaan/kedekatan antar sepasang titik data (Guha dkk, 1999).

Pengelompokan data kategori dengan algoritma ROCK dilakukan dengan tiga langkah sebagai berikut:

1. Menghitung *similarity* yaitu *Similarity* ukuran kemiripan antara pasangan objek ke- i dan objek ke- j dengan rumusan yang didefinisikan sebagai berikut:

$$sim(x_i, x_j) = \frac{|Y_i \cap Y_j|}{|Y_i \cup Y_j|}, i \neq j \tag{6}$$

dengan,

$i = 1, 2, 3, \dots, n$ dan $j = 1, 2, 3, \dots, n$

Y_i = Himpunan pengamatan ke- i , dengan

$Y_i = \{x_{1i}, x_{2i}, \dots, x_{m_{kategori}i}\}$,

Y_j = Himpunan pengamatan ke- j , dengan
 $Y_j = \{x_{1j}, x_{2j}, \dots, x_{m_{kategorij}}\}$,
 $|Y|$ = Bilangan kardinal atau jumlah anggota dari himpunan Y .

2. Menentukan tetangga
 Pengamatan dinyatakan sebagai tetangga jika nilai $sim(x_i, x_j) \geq \theta$.
3. Menghitung *link* antar objek pengamatan.
 Besarnya *link* dipengaruhi oleh nilai *threshold* (q) yang merupakan parameter yang ditentukan oleh peneliti yang dapat digunakan untuk mengontrol seberapa dekat hubungan antar objek. Besarnya nilai q yang digunakan adalah $0 < q < 1$. Algoritma ROCK berhenti ketika jumlah dari kelompok yang diharapkan sudah terpenuhi atau tidak ada lagi *link* antara kelompok-kelompok. Cara menghitung jumlah *link* untuk semua kemungkinan pasangan dari n objek adalah menggunakan matriks A . Matriks A merupakan matriks berukuran $n' \times n$ yang bernilai 1, jika x_i dan x_j dinyatakan mirip (tetangga) dan bernilai 0, jika x_i dan x_j dinyatakan tidak mirip (bukan tetangga). Jumlah *link* antar pasangan x_i dan x_j diperoleh dari hasil kali antara baris ke x_i dan kolom ke x_j dari matriks A . Jika *link* antara x_i dan x_j semakin besar, maka semakin besar kemungkinan x_i dan x_j berada dalam satu kelompok yang sama (Dutta dkk, 2005).
4. Penggabungan kelompok
 Penggabungan kelompok menggunakan algoritma ROCK didasarkan atas ukuran kebaikan (*goodness measure*) antar kelompok dengan rumus seperti pada Persamaan (7). *Goodness measure* adalah persamaan untuk menghitung jumlah *link* dibagi dengan kemungkinan *link* yang terbentuk berdasarkan ukuran kelompoknya.

$$g(C_i, C_j) = \frac{link[C_i, C_j]}{(n_i + n_j)^{1+2f(\theta)} - n_i^{1+2f(\theta)} - n_j^{1+2f(\theta)}} \quad (7)$$

dengan $link[C_i, C_j] = \sum_{x_i \in C_i, x_j \in C_j} link(x_i, x_j)$ yang menyatakan bahwa jumlah *link* dari semua kemungkinan pasangan objek yang ada dalam C_i dan C_j dimana (Tyagi & Sharma, 2012):

n_i = jumlah anggota dalam kelompok ke- i
 n_j = jumlah anggota dalam kelompok ke- j , dan

$$f(q) = \frac{1-q}{1+q}$$

Cluster Ensemble

Strehl dan Gosh (2002) memperkenalkan sebuah metode yang digunakan untuk mengombinasikan sekumpulan solusi *cluster* yang disebut *cluster ensemble*. Metode *cluster ensemble* memiliki keunggulan dibanding metode *cluster* yang lain. Penelitian yang dilakukan oleh Strehl dan Gosh (2002) menunjukkan bahwa metode *cluster ensemble* mampu meningkatkan kualitas dan kekekaran solusi *cluster*.

Clustering pada *cluster ensemble* dilakukan dengan mengombinasikan berbagai solusi dari berbagai metode *clustering* hingga diperoleh satu *cluster* akhir yang lebih baik. Input yang dibutuhkan adalah solusi *clustering* yang telah diperoleh dengan menggunakan berbagai hasil *clustering* tanpa melihat karakteristik data awal (Iam-on & Garret, 2010).

Algoritma CEBMDC

Algoritma *algCEBMDC* dikembangkan untuk menyelesaikan masalah yang berkaitan dengan *clustering* data dengan tipe campuran (kategorik dan numerik). Pertama, data asli yang bertipe campuran dipisah menjadi dua yaitu data dengan tipe kategorik dan data dengan tipe numerik. Selanjutnya, kedua data tersebut diproses secara terpisah dengan menggunakan algoritma *clustering* yang sesuai dengan tipe masing-masing data. Terakhir, *cluster-cluster* yang dihasilkan oleh kedua algoritma digabungkan dan dipandang sebagai data baru dengan tipe kategorik, kemudian diproses dengan menggunakan algoritma *clustering* data kategorik untuk mendapatkan hasil akhir (He, 2002).

Kinerja Pengelompokan Data Numerik

Penentuan jumlah kelompok optimum juga dapat dilakukan dengan melihat nilai maksimum dari *Pseudo-F*. Berikut merupakan rumus yang digunakan untuk menghitung nilai *Pseudo-F* *Sum of Square Total* (SST):

$$SST = \sum_{l=1}^{m_{numerik}} \sum_{i=1}^n (x_{il} - \bar{x}_l)^2 \quad (8)$$

Sum of Square Within Group (SSW):

$$SSW = \sum_{c=1}^C \sum_{l=1}^{m_{numerik}} \sum_{i=1}^{n_c} (x_{ilc} - \bar{x}_{lc})^2 \quad (9)$$

Sum of Square Between Group (SSB):

$$SSB = SST - SSW \quad (10)$$

R-Square:

$$R^2 = \frac{SSB}{SST} = \frac{[SST - SSW]}{SST} \quad (11)$$

Pseudo-F:

$$Pseudo - F = \frac{(R^2/C - 1)}{(1 - R^2/n - C)} \quad (12)$$

dengan keterangan sebagai berikut:

- $m_{numerik}$: jumlah variabel numerik dalam pengamatan
- C : jumlah kelompok yang dibentuk dalam pengamatan
- n : jumlah objek pengamatan
- n_c : jumlah anggota pada kelompok ke- c , untuk $c = 1, 2, \dots, C$
- \bar{x}_l : rata-rata keseluruhan objek pada variabel ke- l , untuk $l = 1, 2, \dots, m_{numerik}$
- \bar{x}_{lc} : rata-rata variabel ke- l pada kelompok ke- c , untuk $c = 1, 2, \dots, C$

Setelah mendapatkan jumlah kelompok yang optimum dalam setiap metode. Selanjutnya, proses analisis dilanjutkan dengan menentukan metode pengelompokan yang paling optimum berdasarkan nilai *internal cluster dispersion rate (icdrate)* minimum.

$$icdrate = 1 - R^2 \quad (13)$$

Kinerja Pengelompokan Data Kategorik

Sum of Square Total (SST) untuk sebuah variabel dengan data kategorik dapat dirumuskan seperti Persamaan (14). Untuk total *Sum of Square Within (SSW)* dirumuskan dalam Persamaan (15). Serta *Sum of Square Between (SSB)* dapat dirumuskan seperti Persamaan (16) (Dewi, 2012).

$$SST = \frac{n}{2} - \frac{1}{2n} \sum_{k=1}^K n_k^2 \quad (14)$$

$$SSW = \frac{n}{2} - \frac{1}{2n} \sum_{c=1}^C \frac{1}{n_c} \sum_{k=1}^K n_{kc}^2 \quad (15)$$

$$SSB = \frac{1}{2} \left(\sum_{c=1}^C \frac{1}{n_c} \sum_{k=1}^K n_{kc}^2 \right) - \frac{1}{2n} \sum_{k=1}^K n_k^2 \quad (16)$$

Sedangkan *Mean of Squares Total (MST)*, *Mean of Squares Within (MSW)*, dan *Mean of Squares Between (MSB)* dapat dirumuskan sebagai berikut:

$$MST = \frac{SST}{(n - 1)} \quad (17)$$

$$MSW = \frac{SSW}{(n - C)} \quad (18)$$

$$MSB = \frac{SSB}{(C - 1)} \quad (19)$$

S_w (simpangan baku dalam kelompok) dan S_B (simpangan baku antar kelompok) untuk data kategori dapat dirumuskan sebagai berikut

$$S_w = [MSW]^{\frac{1}{2}} \quad (20)$$

$$S_B = [MSB]^{\frac{1}{2}} \quad (21)$$

Kinerja suatu metode pengelompokan untuk data kategori semakin baik apa bila nilai rasio S_w dan S_B semakin kecil. Artinya terdapat homogenitas maksimum dalam kelompok dan heterogenitas maksimum antar kelompok (Bunkers & James, 1996).

Pusat Kesehatan Masyarakat (Puskesmas)

Menurut Peraturan Menteri Kesehatan Republik Indonesia Nomor 75 Tahun 2014 Tentang Pusat Kesehatan Masyarakat, Puskesmas adalah unit pelaksana teknis Dinas Kesehatan Kabupaten/kota yang bertanggung jawab menyelenggarakan pembangunan kesehatan di suatu wilayah kesehatan. Puskesmas memiliki tiga fungsi, yaitu sebagai pusat penggerak pembangunan yang berwawasan kesehatan, pusat pemberdayaan masyarakat dan keluarga dalam pembangunan kesehatan serta pusat pelayanan kesehatan masyarakat tingkat pertama.

Sumber daya manusia Puskesmas terdiri atas tenaga kesehatan dan tenaga non kesehatan. Jenis dan jumlah tenaga kesehatan dan non kesehatan dihitung berdasarkan analisis beban kerja, dengan mempertimbangkan jumlah pelayanan yang diselenggarakan, jumlah penduduk dan persebarannya, karakteristik wilayah kerja, luas wilayah kerja, ketersediaan fasilitas pelayanan kesehatan tingkat pertama lainnya di wilayah kerja, dan pembagian waktu kerja. Ketentuan mengenai jenis dan jumlah minimal tenaga kesehatan dan tenaga non kesehatan diatur oleh Peraturan Menteri Kesehatan Nomor 75 Tahun 2014 tentang Puskesmas.

Hasil Penelitian dan Pembahasan

1. Data Penelitian

Data dalam penelitian ini diambil dari publikasi Data Dasar Puskesmas Kondisi Desember 2017 di *website* resmi Kementerian Kesehatan Republik Indonesia yang terdiri dari data variabel numerik dan variabel kategorik. Data variabel numerik terdiri dari data kepadatan penduduk di wilayah kerja (X_1), jumlah dokter umum (X_2), jumlah tenaga dokter gigi (X_3), jumlah tenaga perawat (X_4), jumlah tenaga bidan (X_5), jumlah tenaga perawat (X_6), jumlah tenaga kesehatan masyarakat (X_7), jumlah tenaga kesehatan lingkungan (X_8), jumlah tenaga gizi (X_9), jumlah ahli laboratorium medik (X_{10}) dan jumlah tenaga penunjang kesehatan (X_{11}).

Data variabel kategorik terdiri dari data kemampuan penyelenggaraan (X_{12}) yang dikategorikan menjadi 2 kategori yaitu non rawat inap (kode: 0) dan rawat inap (kode: 1) serta data karakteristik wilayah kerja (X_{13}) yang

dikategorikan menjadi 3 kategori yaitu terpencil/sangat terpencil (kode: 0), perdesaan (kode: 1), dan perkotaan (kode: 2).

2. Analisis Cluster Data Numerik

Analisis *cluster* untuk data numerik dilakukan menggunakan algoritma AGNES. Algoritma AGNES bekerja dengan menghitung jarak antar objek, setiap objek berfungsi sebagai *cluster*, kemudian secara bertahap menggabungkan setiap pasangan *cluster* terdekat berdasarkan ukuran jarak dan metode penggabungan yang digunakan, hingga semua *cluster* tergabung menjadi satu *cluster*.

Metode yang digunakan dalam menghitung jarak antar objek adalah jarak *euclidean*, dan metode penggabungan yang digunakan adalah *single linkage*, *complete linkage*, dan *average linkage*. Pengelompokan dilakukan dengan membentuk 5, 4, 3, dan 2 *cluster*. Analisis *cluster* terhadap data numerik dilakukan dengan bantuan *software RStudio Version 1.1.383*.

Adapun tahapan dalam algoritma AGNES adalah sebagai berikut:

1. Standarisasi data untuk data variabel numerik menggunakan persamaan (1).
2. Inisialisasi data.
3. Menghitung matriks jarak *euclidean* menggunakan persamaan (2).
4. Pembentukan 5, 4, 3, dan 2 *cluster* menggunakan metode penggabungan *single linkage* pada persamaan (3).
5. Pembentukan 5, 4, 3, dan 2 *cluster* menggunakan metode penggabungan *complete linkage* pada persamaan (4).
6. Pembentukan 5, 4, 3, dan 2 *cluster* menggunakan metode penggabungan *average linkage* pada persamaan (5).
7. Pemilihan metode penggabungan terbaik dengan melihat nilai *Pseudo-F* tertinggi dan jumlah kelompok optimum dengan nilai *icdrate* terkecil.

Dari proses clustering menggunakan algoritma AGNES didapatkan hasil seperti pada Tabel 1.

Tabel 1. Hasil Kinerja Pengelompokan Data Numerik

Metode	Jumlah Cluster	Pseudo-F	Icdrate
Single linkage	2	0,1603	0,9991
	3	0,1497	-
	4	0,1526	-
	5	0,1547	-
Complete linkage	2	0,3001	-
	3	0,4585	0,9948
	4	0,3159	-
Average linkage	5	0,2357	-
	2	0,3001	-
	3	0,2299	-
Average linkage	4	0,1526	-
	5	0,3669	0,9916

Pada Tabel 1 diketahui pengelompokan Puskesmas di Provinsi Kalimantan Timur menggunakan metode *single linkage* dengan jumlah 2, 3, 4, dan 5 *cluster* diperoleh nilai *Pseudo-F* optimum berada pada jumlah 2 *cluster* yaitu sebesar 0,1603. Pengelompokan menggunakan metode *complete linkage* didapatkan nilai *Pseudo-F* optimum berada pada jumlah 3 *cluster* yaitu sebesar 0,4585. Pengelompokan menggunakan metode *average linkage* didapatkan nilai *Pseudo-F* optimum pada jumlah 5 *cluster* yaitu sebesar 0,3669.

Selanjutnya dari ketiga metode penggabungan tersebut dibandingkan nilai *icdrate* sehingga diperoleh metode penggabungan terbaik untuk pengelompokan Puskesmas di Provinsi Kalimantan Timur berdasarkan data numerik. Metode penggabungan terbaik dalam mengelompokkan Puskesmas di Provinsi Kalimantan Timur adalah dengan menggunakan metode *average linkage* dengan 5 *cluster*, karena nilai *icdrate* dari metode ini adalah nilai *icdrate* minimum dibandingkan dengan nilai *icdrate* kedua metode lainnya yaitu sebesar 0,9916.

Analisis Cluster Data Kategorik

Pada data kategorik yang terdiri dari variabel kemampuan penyelenggaraan (X_{12}) dan variabel karakteristik wilayah kerja (X_{13}), Puskesmas di Provinsi Kalimantan Timur kondisi Desember 2017 dikelompokkan menggunakan algoritma ROCK. Pada penelitian ini digunakan nilai θ sebesar 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; dan 0,9. Analisis *cluster* data kategorik menggunakan algoritma ROCK dilakukan dengan bantuan *software RStudio Version 1.1.383*. Adapun tahapan-tahapan dalam proses algoritma ROCK untuk 179 *cluster* dengan threshold 0,1 adalah sebagai berikut:

1. Inisialisasi data.
 2. Menghitung *similarity* antar objek.
- Contoh perhitungan *similarity* antara objek 1 dan objek 2 adalah sebagai berikut:

$$Y_1 = \{x_{121}, x_{131}\} = \{\text{rawat inap, perkotaan}\}$$

$$Y_2 = \{x_{122}, x_{132}\} = \{\text{rawat inap, perdesaan}\}$$

$$sim(x_1, x_2) = \frac{|Y_1 \cap Y_2|}{|Y_1 \cup Y_2|} = \frac{1}{3} = 0,333$$

Perhitungan *similarity* dari setiap pasangan objek membentuk matriks **sim** yang merupakan matriks simetris dengan ukuran 179 x 179 (banyaknya objek).

3. Menentukan tetangga.
- Suatu pasangan objek dinyatakan sebagai tetangga jika nilai dari $sim(x_i, x_j) \geq q$.

Informasi mengenai hubungan tetangga antar objek pengamatan dapat dinyatakan dengan matriks **A**. Matriks **A** merupakan matriks simetris berukuran 179 x 179 yang bernilai 1

jika objek tersebut memenuhi syarat bertetangga dan bernilai 0 jika objek tersebut tidak memenuhi jarak bertetangga.

4. Menghitung *Link*.
Link dari pasangan objek ke-i dan objek ke-j dapat dihitung dengan mengalikan baris ke-i dan kolom ke-j dari matriks A.
5. Menghitung *Goodness Measure*.
Nilai *goodness measure* dapat dihitung menggunakan Persamaan (7). Sebagai contoh dilakukan perhitungan nilai *goodness measure* untuk objek ke-1 dan objek ke-2 dengan $q = 0,1$ sebagai berikut:

$$\text{Dengan } f(0,1) = \frac{1 - 0,1}{1 + 0,1} = 0,818$$

$$g(C_1, C_2) = \frac{\text{link}[C_1, C_2]}{(n_1 + n_2)^{1+2f(\theta)} - n_1^{1+2f(\theta)} - n_2^{1+2f(\theta)}}$$

$$= \frac{95}{(1+1)^{1+2(0,818)} - 1^{1+2(0,818)} - 1^{1+2(0,818)}}$$

$$= 22,525$$

Pasangan objek atau *cluster* yang memiliki nilai *goodness measure* paling tinggi akan digabungkan menjadi satu *cluster*. Dari proses *clustering* menggunakan algoritma ROCK didapatkan hasil seperti pada Tabel 2.

Tabel 2. Hasil Kinerja Pengelompokan Data Kategorik

Threshold	Jumlah Cluster yang diinginkan	Jumlah Cluster yang terbentuk	Rasio S_w/S_B
0,1	2	2	0,0472
	3	2	0,0437
	4	2	$2,4541 \times 10^{-16}$
	5	3	0,0200
0,2	2	2	$6,5381 \times 10^{-16}$
	3	2	0,0592
	4	2	$2,4838 \times 10^{-16}$
	5	2	0,0331
0,3	2	2	$6,5381 \times 10^{-16}$
	3	2	0,0592
	4	3	0,0511
	5	3	0,0530
0,4	2, 3, 4, 5	6	0,0551
0,5	2, 3, 4, 5	6	0,0551
0,6	2, 3, 4, 5	6	0,0551
0,7	2, 3, 4, 5	6	0,0551
0,8	2, 3, 4, 5	6	0,0551
0,9	2, 3, 4, 5	6	0,0551

Berdasarkan Tabel 2 di atas terlihat bahwa kelompok optimum dihasilkan oleh *threshold* 0,1 dengan jumlah kelompok yang terbentuk adalah

dua kelompok dan nilai rasio S_w/S_B sebesar $2,4541 \times 10^{-16}$.

Analisis Cluster Data Campuran

Setelah mendapatkan hasil pengelompokan dari masing-masing pengelompokan berdasarkan tipe data. Kemudian hasil pengelompokan tersebut digabungkan dan dikelompokkan kembali menggunakan algoritma ROCK.

Hasil pengelompokan dari masing-masing variabel dengan tipe data numerik dan kategorik kemudian dinyatakan sebagai variabel baru dengan tipe data kategorik. Variabel X_{14} merupakan variabel yang berisikan hasil pengelompokan dari data numerik. Sedangkan X_{15} merupakan variabel yang berisikan hasil pengelompokan data kategorik. X_{14} dan X_{15} dianggap sebagai variabel kategorik baru yang akan dipakai untuk mengelompokkan Puskesmas di Provinsi Kalimantan Timur dengan metode ROCK sesuai dengan algCEMBDC.

Berikut merupakan struktur data baru yang digunakan dalam pengelompokan data campuran.

Tabel 3. Struktur Data Baru

No	Puskesmas	X_{14} (Output Cluster Numerik)	X_{15} (Output Cluster Kategorik)
1	Batu Kajang	1	1
2	Muser	1	1
3	Kerang	1	1
⋮	⋮	⋮	⋮
179	Bontang Barat	1	2

Dengan struktur data yang baru sesuai dengan Tabel 3 di atas kemudian dilakukan pengelompokan untuk mendapatkan *final cluster*. Nilai *threshold* yang digunakan dalam pengelompokan adalah 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; dan 0,9. Jumlah kelompok yang diharapkan adalah sama 2, 3, 4 atau 5 *cluster*. Hasil pengelompokan data campuran dengan algoritma ROCK dapat dilihat pada Tabel 4.

Berdasarkan Tabel 4 diketahui nilai minimum rasio antara simpangan baku dalam kelompok dengan simpangan baku antar kelompok adalah $7,283 \times 10^{-17}$. Anggota kelompok dari dua *cluster* yang terbentuk baik dengan *threshold* 0,2 dan 0,3 adalah sama. Sehingga pemilihan *threshold* 0,2 dan 0,3 tidak memberikan efek yang berbeda.

Final cluster yang didapatkan dari pengelompokan data campuran adalah dua kelompok dengan masing-masing kelompok beranggotakan 83 Puskesmas dan 96 Puskesmas.

Karakteristik Final Cluster

Hasil pengelompokan pada data campuran yang dilakukan menggunakan algoritma ROCK

dengan nilai *threshold* sebesar 0,2 dan 0,3 menghasilkan dua *cluster* optimum. Karakteristik Puskesmas berdasarkan variabel data numerik yang dimiliki dapat ditinjau dari segi kepadatan penduduk di wilayah kerja dan ketenagaan di tiap Puskesmas. Pada Tabel 6 dapat dilihat karakteristik variabel numerik antara Puskesmas pada *cluster* 1 pertama dan *cluster* 2.

Tabel 4. Hasil Kinerja Pengelompokan Data Campuran

Threshold	Jumlah Cluster yang diinginkan	Jumlah Cluster yang terbentuk	Rasio S_w/S_B
0,1	2	2	0,0159
	3	2	0,0159
	4	2	0,0159
	5	2	0,0159
0,2	2	2	0,0159
	3	2	$7,283 \times 10^{-17}$
	4	2	$1,339 \times 10^{-15}$
0,3	5	2	$2,728 \times 10^{-15}$
	2	2	0,0159
	3	2	$7,283 \times 10^{-17}$
0,4	4	2	0,0159
	5	3	0,0369
	2, 3, 4, 5	4	0,0839
0,5	2, 3, 4, 5	4	0,0839
0,6	2, 3, 4, 5	4	0,0839
0,7	2, 3, 4, 5	4	0,0839
0,8	2, 3, 4, 5	4	0,0839
0,9	2, 3, 4, 5	4	0,0839

Tabel 5. Anggota Cluster Hasil Pengelompokan

Cluster	Puskesmas Anggota
1	Pasir Belengkong, Suliliran Baru, Padang Pengrapat, Tanah Grogot, Lolo, Krayan, Kayungo, Mendik, Penyinggahan, Muara Pahu, Linggang Bigung, Melak, Tering Seberang, Sungai Merdeka, Handil Baru, Sanga-Sanga, Loa Janan, Batuah, Loa Duri, Loa Kulu, Jonggon Jaya, Muara Wis, Mangkurawang, Rapak Mahang, Loa Ipuh, Separi III, Teluk Dalam, Bunga Jadi, Teluk Lingga, Pulau Derawan, Maratua Bohe Bukut, Sambaliung, Suaran, Tanjung Redeb, Kampung Bugis, Merancang Ulu, Lebanan Makmur, Biatan Lempake, Gunung Intan, Sebakung Jaya, Long Hubung, Prapatan, Gunung Bahagia, Telaga Sari, Lamaru, Teritip, Manggar Baru, Gunung Samarinda, Muara Rapak, Karang Jati, Gunung Sari Ilir, Karang Rejo, Sumber Rejo, Gunung Sari Ulu, Baru Ilir, Sidomulyo, Margo Mulyo, Baru Tengah, Margasari, Bukuan, Sambutan, Sidomulyo, Sei Kapih, Samarinda Kota, Kampung Baka, Mangkupalas, Harapan Baru, Lok Bahu, Pasundan, Air Putih, Juanda, Segiri, Wonorejo, Karang Asam, Temindung, Sempaja, Remaja, Bengkuring, Bontang Selatan I, Bontang Selatan II, Bontang Utara I, Bontang Utara II, Bontang

Barat.

2 Batu Kajang, Muser, Kerang, Tanjung Aru, Suatang Baru, Kuaro, Long Ikis, Muara Komam, Long Kali, Muara Kedang, Resak,

Tabel 5. Anggota Cluster Hasil Pengelompokan (lanjutan)

Cluster	Puskesmas Anggota
	Tanjung Isuy, Belusuh, Lambing, Dilang Puti, Damai, Besiq, Dempar, Barong Tongkok, Sekolaq Darat, Gunung Rampah, Long Iram, Samboja, Muara Jawa, Muara Muntai, Kota Bangun, Rimba Ayu, Sebulu I, Sebulu II, Sungai Mariam, Muara Badak, Badak Baru, Marang Kayu, Perangat, Muara Kaman, Kahala, Kembang Janggut, Tabang, Ritan Baru, Muara Ancalong, Busang, Muara Wahau I, Muara Wahau II, Telen, Kongbeng, Muara Bengkal, Sepaso, Teluk Pandan, Sangatta Selatan, Rantau Pulung, Sangkulirang, Kaliorang, Sandaran, Kaubun, Karang Dalam, Muara Lesan/Kelay, Talisayan, Biduk-biduk, Tanjung Batu, Gunung Tabur, Tepian Buah/Segah, Teluk Bayur, Babulu, Waru, Penajam, Petung, Sotek, Sepaku I, Sepaku III, Maridan, Semoi II, Laham, Ujoh Bilang, Long Pahangai, Tiong Ohang, Sepinggan, Damai, Klandasan Ilir, Manggar, Batu Ampar, Karang Joang, Mekar Sari, Baru Ulu, Kariangau, Palaran, Trauma Centre, Bantuas, Maroman, Loa Bakung, Lempake, Sei Siring, Bontang Lestari

Tabel 6. Karakteristik Cluster Berdasarkan Data Numerik

Variabel	Cluster 1			Cluster 2		
	Mean	Min	Max	Mean	Min	Max
X_1	2.154,18	0,03	15.70	229,7	0,64	5.66
X_2	2,783	1	8	2,708	0	9
X_3	1,241	0	3	1,031	0	3
X_4	11,55	2	38	16,35	2	45
X_5	9,964	2	63	14,96	1	50
X_6	2,036	0	8	2,167	0	7
X_7	1,482	0	6	1,729	0	6
X_8	1,265	0	4	1,156	0	6
X_9	1,096	0	6	1,198	0	7
X_{10}	1,349	0	6	1,302	0	5
X_{11}	9,446	3	22	10,68	2	36

Karakteristik Puskesmas di Provinsi Kalimantan Timur kondisi Desember 2017 berdasarkan variabel-variabel berskala kategorik dapat ditunjukkan melalui modus atau kategori yang paling sering muncul pada setiap *cluster*. Perbandingan karakteristik Puskesmas berdasarkan variabel kategorik antara Puskesmas *cluster* satu dengan *cluster* dua disajikan pada Tabel 7.

Tabel 7. Karakteristik Cluster Berdasarkan Data Kategorik

Variabel	Cluster 1			Cluster 2		
	Mo	Su	Persent	Mo	Su	Persent

	de	m	ase	de	m	ase
X_{12}	0	83	100%	1	96	100%
X_{13}	0	50	60,24%	0	33	34,38%

Dari dua variabel kategorik dalam penelitian ini, satu variabel menunjukkan perbedaan mayoritas pada Puskesmas *cluster* satu dan *cluster* dua. Untuk variabel kemampuan penyelenggaraan (X_{12}), seluruh Puskesmas di *cluster* satu memiliki kemampuan penyelenggaraan non rawat inap, sedangkan seluruh Puskesmas di *cluster* dua memiliki kemampuan penyelenggaraan rawat inap. Untuk variabel karakteristik wilayah (X_{13}), mayoritas Puskesmas di *cluster* satu memiliki karakteristik wilayah kerja berupa daerah perkotaan yaitu sebesar 60,24%. Di sisi lain mayoritas Puskesmas di *cluster* dua juga memiliki karakteristik wilayah kerja berupa daerah perkotaan dengan persentase yang lebih kecil dari *cluster* dua yaitu sebesar 34,38%.

Berdasarkan uraian di atas dapat disimpulkan bahwa karakteristik dari masing-masing *cluster* berdasarkan data numerik dan data ketegorik adalah sebagai berikut:

1. Cluster Satu

Cluster satu berisikan 83 Puskesmas non rawat inap yang mayoritas memiliki wilayah kerja daerah perkotaan sebesar 60,24% dan kepadatan penduduk di wilayah kerja terbilang padat yaitu sebesar 2154,18 jiwa/km². Menurut standar ketenagaan Puskesmas dalam Peraturan Menteri Kesehatan Republik Indonesia Nomor 75 Tahun 2014, jumlah ketenagaan Puskesmas pada *cluster* satu yang telah memenuhi standar ketenagaan Puskesmas non rawat inap hanya tenaga dokter umum dan tenaga penunjang kesehatan. Sedangkan untuk tenaga kesehatan lainnya masih terdapat Puskesmas pada *cluster* satu yang belum memenuhi standar ketenagaan Puskesmas non rawat inap.

Tabel 8. Ketenagaan Puskesmas *Cluster* Satu yang Belum Memenuhi Standar

Jenis Tenaga	Jumlah Puskesmas yang Belum Memenuhi Standar Ketenagaan	
	yang Belum Memenuhi Standar	Persentase
X_2	0	0,00%
X_3	9	10,84%
X_4	18	21,69%
X_5	15	18,07%
X_6	1	1,20%
X_7	40	48,19%
X_8	7	8,43%
X_9	12	14,46%
X_{10}	3	3,61 %
X_{11}	0	0,00%

Puskesmas pada *cluster* satu mayoritas kekurangan tenaga kesehatan masyarakat dengan persentase 48,19% atau sebanyak 40 Puskesmas dari total 83 Puskesmas diikuti tenaga perawat

sebesar 21,69%, tenaga bidan sebesar 18,07%, dan tenaga ahli laboratorium medik sebesar 16,87%.

2. Cluster Dua

Cluster dua beranggotakan 96 Puskesmas dengan kemampuan penyelenggaraan rawat inap yang mayoritas wilayah kerjanya berupa daerah perkotaan walaupun tidak sebesar Puskesmas pada *cluster* satu yaitu sebesar 34,38%. Rata-rata kepadatan penduduk di wilayah kerja Puskesmas *cluster* dua terbilang jauh lebih kecil dari *cluster* satu yaitu sebesar 229,79 jiwa/km². Menurut standar ketenagaan Puskesmas dalam Peraturan Menteri Kesehatan Republik Indonesia Nomor 75 Tahun 2014, jumlah ketenagaan Puskesmas pada *cluster* dua yang telah memenuhi standar ketenagaan Puskesmas rawat inap hanya tenaga tenaga penunjang kesehatan. Sedangkan untuk tenaga kesehatan lainnya masih terdapat Puskesmas pada *cluster* satu yang belum memenuhi standar ketenagaan Puskesmas rawat inap.

Puskesmas pada *cluster* dua mayoritas kekurangan tenaga gizi dengan persentase 72,92% atau sebanyak 70 Puskesmas dari total 96 Puskesmas diikuti tenaga kesehatan masyarakat 46,88%, tenaga dokter umum sebesar 25%, dan tenaga dokter gigi sebesar 19,79%.

Tabel 9. Ketenagaan Puskesmas *Cluster* Satu yang Belum Memenuhi Standar

Jenis Tenaga	Jumlah Puskesmas yang Belum Memenuhi Standar Ketenagaan	
	yang Belum Memenuhi Standar	Persentase
X_2	24	25,00%
X_3	19	19,79%
X_4	7	7,29%
X_5	8	8,33%
X_6	11	11,46%
X_7	45	46,88%
X_8	13	13,54%
X_9	70	72,92%
X_{10}	12	12,50%
X_{11}	0	0,00%

Kesimpulan

Berdasarkan hasil analisis dan pembahasan, maka didapatkan kesimpulan sebagai berikut:

1. Hasil pengelompokan Puskesmas di Kalimantan Timur berdasarkan algCEBMDC adalah *Final cluster* yang didapatkan dari metode *ensemble* ROCK, yaitu menggabungkan hasil pengelompokan data numerik dengan hasil pengelompokan data kategorik menghasilkan *cluster* optimum pada nilai threshold 0,2 dan 0,3 dan rasio S_W/S_B sebesar $7,283 \cdot 10^{-17}$. Jumlah *cluster* optimum yang terbentuk adalah sebanyak dua *cluster*.

2. Karakteristik dari masing-masing *final cluster* adalah sebagai berikut:
- Cluster* satu berisikan 83 Puskesmas non rawat inap yang mayoritas memiliki wilayah kerja daerah perkotaan sebesar 60,24% dan kepadatan penduduk di wilayah kerja terbilang padat yaitu sebesar 2154,18 jiwa/km². Menurut standar ketenagaan Puskesmas dalam Peraturan Menteri Kesehatan Republik Indonesia Nomor 75 Tahun 2014, jumlah tenaga medis Puskesmas pada *cluster* satu yang telah memenuhi standar ketenagaan Puskesmas non rawat inap hanya tenaga dokter umum dan tenaga penunjang kesehatan.
 - Cluster* dua beranggotakan 96 Puskesmas dengan kemampuan penyelenggaraan rawat inap yang mayoritas wilayah kerjanya berupa daerah perkotaan walaupun tidak sebesar Puskesmas pada *cluster* satu yaitu sebesar 34,38%. Rata-rata kepadatan penduduk di *cluster* dua terbilang jauh lebih kecil dari *cluster* satu yaitu sebesar 229,79 jiwa/km². Menurut standar ketenagaan Puskesmas dalam Peraturan Menteri Kesehatan Republik Indonesia Nomor 75 Tahun 2014, jumlah tenaga medis Puskesmas pada *cluster* dua yang telah memenuhi standar ketenagaan Puskesmas rawat inap hanya tenaga penunjang kesehatan.

Daftar Pustaka

- Amirullah. (2017, April 11) Kaltim Revitalisasi Pelayanan Seluruh Puskesmas pada 2018. Diambil dari <https://kaltim.antaranews.com/berita/37798/kaltim-revitalisasi-pelayanan-seluruh-puskesmas-pada-2018>.
- Bunker, M. J., & James, R. M. (1996). Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique. *Journal of Climate*, Vol. 9, 130-146.
- Dewi, A. (2012). *Metode Cluster Ensemble untuk Pengelompokan Desa Perdesaan di Provinsi Riau*. Surabaya: Institut Teknologi Sepuluh November.
- Dutta, M., Mahanta, A. K. & Pujari, A. K. (2005). QROCK: A Quick Version of the ROCK Algorithm for Clustering of Categorical Data. *Pattern Recognition Letters*, Vol. 26, 2364-2373.
- Guha, S., Rastogi, R. & Shim, K. (1999). ROCK: A Robust Clustering algorithm for Categorical Attributes. *Proceedings of The 15th IEEE International Conference on Data Engineering*.
- Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2014). *Multivariate Data Analysis 7th Edition*. USA: Pearson Education, Inc.
- Han, J., Kamber, M. & Pei, J. (2012). *Data Mining Concepts and Techniques Third Edition*. USA: Elsevier.
- He, Z., Xu, X. & Deng, S. (2005). Clustering Mixed Numeric and Categorical Data: A Cluster Ensemble Approach. *CoRR, abs/cs/0509011*.
- Iam-on, N., & Garret, S. (2010). LinkCluE: A MATLAB Package for Link-Based Cluster Ensemble. *Journal of Statistical Software*. 36(9):1-36
- Jamkesindonesia. (2019). Tingkatkan Pelayanan Kesehatan Masyarakat. Diambil dari <http://jkn.jamsosindonesia.com/home/cetak/375/Tingkatkan%20Pelayanan%20Kesehatan%20Dasar>.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis 6th Edition*. USA: Pearson Education, Inc.
- Mattjik, A. A., & Sumetrajaya, I. M. (2011). *Sidik Peubah Ganda dengan Menggunakan SAS*. Bogor: IPB Press.
- Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Strehl, A., & Gosh, J. (2002). A Knowledge Reuse Framework for Combining Cluster Partitions. *Proceeding of the 2007 Sawtooth Software Conference*.
- Tan, P., Steinbach, M. & Kumar, V. (2006). *Introduction to Data Mining*. USA: Person Education, Inc.
- Tyagi, A., & Sharma, S. (2012). Implementation of ROCK Clustering Algorithm for the Optimization of Query Searching Time. *International Journal on Computer Science and Engineering*, Vol. 4 No. 05, 809-815.
- Widarjono, A. (2015). *Analisis Multivariat Terapan*. Yogyakarta: Penerbit UPP STIM YKPN
- Yulianto, S & Hidayatullah, K. H. (2014). Analisis Kluster untuk Pengelompokan Kabupaten/Kota di Provinsi Jateng berdasarkan Indikator Kesejahteraan Rakyat. *Jurnal Akd. Statistika Muhammadiyah Semarang*, 2(1), 58-59.