|  |  |  |
| --- | --- | --- |
|  | Available online to [www.journal.unipdu.ac.id](http://www.journal.unipdu.ac.id)    **S2-Accredited** – [SK No. 34/E/KPT/2018](http://arjuna.ristekdikti.go.id/index.php/news/view/138)  Journal Page is available to [www.journal.unipdu.ac.id:8080/index.php/register](http://www.journal.unipdu.ac.id:8080/index.php/register) |  |

Fuzzy C-Means Clustering dengan Reduksi Dimensi Core and Reduct untuk Analisis Klaster COVID-19 Indonesia berdasarkan Provinsi

Joko Eliyanto a, Sugiyarto b

a Magister Pendidikan Matematika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

b Matematika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

email: a joko1907050003@webmail.uad.ac.id, bsugiyarto@math.uad.ac.id

|  |  |  |
| --- | --- | --- |
| A R T I C L E I N F O |  | A B S T R A C T |
| Article history:  Received 24 January 2020 Revised 30 April 2020 Accepted 2 December 2020 Available online xxx | Memasuki awal Tahun 2021, Indonesia kembali meningkatkan upaya peangangan Pandemi COVID-19. Pemerintah Republik Indonesia melalui Instruksi Kementrian Dalam Negeri No 1 Tahun 2021 menetapkan kebijakan untuk melakukan Pembatasan Sosial Berskala Besar(PSBB) untuk pulau Jawa dan Bali. Pemilihan kawasan tersebut didasarkan pada parameter-parameter yang telah ditentukan oleh Kementrian Kesehatan. Melalui gugus tugas COVID-19, pemerintah juga membuat peta resiko untuk setiap kabupaten/kota d Indonesia berdasarkan parameter yang sama. Tindakan ini bisa dipandang sebagai tindakan pengelompokan wilayah didasarkan pada kriteria-kriteria tertentu. Makalah ini menawarkan pengelompokan wilayah menggunakan teknik clustering. Clustering adalah proses pengelompokan data sedemikian sehingga data yang berada dalam klaster yang sama memiliki kemiripan yang tinggi dan data pada klaster yang berbeda memiliki kemiripan yang rendah. Fuzzy c-means clustering dengan reduksi dimensi core and reduct adalah metode clustering yang mempertimbangkan keambiguan dalam data serta memiliki beban komputasi yang rendah dengan kualitas hasil yang dipertahankan. Makalah ini bertujuan untuk menghasilkan pengelompokan provinsi yang didasarkan pada dataset yang digunakan. Variabel-variabel penting untuk masing-masing provinsi seperti data konfirmasi positif, data pasien sembuh, pasien meninggal, jumlah penduduk, luas wilayah, jumlah pulau,jumlah rumah sakit rujukan dan koordinat provinsi. Hasilnya diperoleh dua klaster provinsi. Salah satu klister berisi 13 provinsi yaitu Lampung, Banten, Jawa Barat, DKI Jakarta, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, NTB, NTT, Kalimantan Selatan, Sulawesi Selatan dan Maluku. Hasil ini tidak jauh berbeda dengan hasil pengelompokan oleh pemerintah. Provinsi-provinsi ini disarankan diprioritaskan untuk mendapat perhatian lebih dalam penanganan Pandemi COVID-19. |
| Keywords:  COVID-19,  Clustering,  Core and Reduct,  Fuzzy C-Means,  Roughset |
| **IEEE style in citing this article:**  Joko Eliyanto and Sugiyarto, " Fuzzy C-Means Clustering dengan Reduksi Dimensi Core and Reduct untuk Analisis Klaster COVID-19 Indonesia berdasarkan Provinsi," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 7, no. 1, pp. 1-10, 2021. |
| 2021 Register: *Jurnal Ilmiah Teknologi Sistem Informasi* (Scientific Journal of Information System Technology) with CC BY NC SA license. | | |

1. Pendahuluan

Pada awal 2020, dunia dikejutkan dengan mewabahnya pneumonia baru yang bermula dari Wuhan, Provinsi Hubei yang kemudian menyebar dengan cepat ke lebih dari 190 negara dan teritori [1]. COVID-19 pertama dilaporkan di Indonesia pada tanggal 2 Maret 2020 sejumlah dua kasus (Wulandari et al. 2020). Pada tanggal 31 Maret 2020, Presiden Jokowi mengadakan Konferensi Pers, dengan tujuan untuk mengumumkan kepada publik mengenai kebijakan yang dipilihnya yaitu Pembatasan Sosial Berskala Besar (PSBB) pada provinsi-provinsi yang tingkat penularannya tinggi [2]. Pada 9 April 2020, pandemi sudah menyebar ke 34 provinsi dengan DKI Jakarta, Jawa Timur dan Jawa Barat sebagai provinsi terpapar virus corona di Indonesia. Dampak dari COVID-19 ini juga menyentuh sektor ekonomi, tercatat perumbuhan ekonomi Indonesia merosot hingga –5.32 % pada kuartal II 2020. Memasuki awal Tahun 2021, Indonesia kembali meningkatkan upaya peangangan Pandemi COVID-19. Pemerintah Republik Indonesia melalui Instruksi Kementrian Dalam Negeri No 1 Tahun 2021 menetapkan kebijakan untuk melakukan Pembatasan Sosial Berskala Besar(PSBB) untuk pulau Jawa dan Bali. Pemilihan kawasan tersebut didasarkan pada parameter-parameter yang telah ditentukan oleh Kementrian Kesehatan. Melalui gugus tugas COVID-19, pemerintah juga membuat peta resiko untuk setiap kabupaten/kota d Indonesia berdasarkan parameter yang sama. Tindakan ini bisa dipandang sebagai tindakan pengelompokan wilayah didasarkan pada kriteria-kriteria tertentu. Teknik lain untuk melakukan pengelompokan adalah clustering.

Clustering adalah proses pengelompokan data sedemikian sehingga data yang berada dalam klaster yang sama memiliki kemiripan yang tinggi dan data pada klaster yang berbeda memiliki kemiripan yang rendah [3]. Sebuah anggota dataset dapat masuk menjadi anggota suatu kelompok data melalui dua acara [4]. Pertama adalah memasukkan sebuah data ke sebuah kelompok dengan kriteria tertentu (misal jarak terpendek) dan kedua dengan menghitung derajat keanggotaan sebuah data terhadap klaster-klaster yang ada [5]. Cara yang kedua ini disebut sebagai fuzzy clustering [6]. Pada metode fuzzy, kemungkinan keanggotaan sebuah data terhadap suatu kelompok dapat dipertimbangkan. Metode yang populer digunakan untuk fuzzy clustering adalah fuzzy c-means [7]. Pada metode ini, data yang diubah ke dalam bentuk fuzzy adalah jarak antara objek dengan pusat klaster yang diberikan, fungsi objektif dari fuzzy c-means adalah hasil kali dari derajat keanggotaan data pada suatu klaster dengan kuadrat dari jarak titik data ke pusat klaster [8]. Jarak merupakan alat ukur yang penting pada algoritma clustering [9]. Fungsi objektif akan diminmumkan sehingga diperoleh klaster yang anggota-anggota di dalamnya sangat mirip dan perbedaan antar klaster tinggi. Bagaimanapun sebuah metode clustering akan menjadi sangat kompleks ketika data memiliki variabel dan record yang banyak [10]. Pada data berdimensi besar, sebelum proses clustering dapat dilakukan sebuah langkah reduksi dimensi.

Reduksi dimensi adalah cara untuk menurukan dimensi pada sebuah data set dengan tetap mempertahankan informasi yang penting pada dataset tersebut [11]. Teori rough set adalah alat matematika baru untuk menangani ketidakpastian dan ketidakjelasan sistem keputusan dan telah berhasil diterapkan di semua bidang. Metode ini digunakan untuk mengidentifikasi set reduksi set semua atribut dari sistem keputusan. Set reduksi digunakan sebagai teknik preprocessing untuk klasifikasi sistem keputusan untuk mengeluarkan pola potensial atau aturan asosiasi atau pengetahuan melalui teknik data mining [12]. Core and reduct adalah teknik dalam rough set yang bertujuan untuk mengurangi dan hanya menyisakan inti dari suatu dataset [13].

Fuzzy C-Means Clustering adalah metode Clustering yang biasanya diterapkan pada data bervolume besar. Data yang tersedia dalam berbagai jenis dapat direpresentasikan sebagai variabel [14]. Dalam sudut pandang matematika, variabel juga dipandang sebagai dimensi [15]. Semakin banyak record data dan variabel data, maka semakin kompleks dan lama pula proses komputasi yang harus dilakukan [16]. Untuk mengatasi hal tersebut, maka reduksi dimensi diterapkan untuk meningkatkan performa dari Fuzzy C-Means Clustering [17].

Penelitian ini berfokus pada penerapan fuzzy c-means clustering dengan dimensi core and reduct untuk klasterisasi COVID-19 di Indonesia berdasarkan Provinsi. Tujuan utama dari penelitian ini adalah menghasilkan klsster provinsi didasarkan pada kemiripan data yang ada. Sehingga provinsi-provinsi yang memiliki karakteristik yang sama dapat juga mendapat perlakuan yang sama. Hal ini diharapkan dapat mempercepat penanganan pandemic COVID-19 di Indonesia. Data yang digunakan pada penelitian ini adalah data konfirmasi positif, data pasien sembuh, pasien meninggal, jumlah penduduk, luas wilayah, jumlah pulau,jumlah rumah sakit rujukan dan koordinat provinsi. Metode Elbow digunakan untuk menentukan jumlah klaster yang direkomendasikan. Hasil klasterisasi diperbandingkan dengan hasil pengelompokan yang dilakukan oleh Pemerintah kemudian analisis terhadap keduanya disajikan.

1. State of the Art

Metode Fuzzy C-Means Clustering memiliki keunggulan yaitu dapat mempertimbangkan nilai kefuzzyan sebuah data untuk masuk ke dalam salah satu klaster [18]. Meskipun demikian, menerapkan metode ini dalam data yang berdimensi besar berpotensi untuk memperbesar error dan beban komputasi yang cukup besar. Keduanya saling berkaitan. Penerapan reduksi dimensi pada FCM dapat mengatasi hal itu. Metode reduksi dimensi dapat menurunkan dimensi dataset. Dataset yang tereduksi tersebut kemudian digunakan. Hasil dari penerapan Fuzzy C Means Clustering pada dataset yang tereduksi memiliki hasil yang relatif tidak jauh berbeda dengan dataset aslinya[19]. Beberapa metode reduksi dimensi dapat diterapkan, di antaranya adalah PCA dan Core and Reduct[20]. PCA dapat diterapkan dengan baik pada metode Fuzzy C-Means Clustering yang dikombinasikan dengan jarak Minkowski-Chebisev [21]. Penerapan reduksi dimensi Core and Reduct terbukti dapat menurunkan beban komputasi, menurunkan nilai fungsi objektif, dan kualitas klaster tetap dapat dipertahankan [17]. Selain pengembangannya, Fuzzy C-Means Clustering juga telah diterapkan di berbagai bidang kehidupan.

Fuzzy C-Means Clustering digunakan untuk pengelompokan kejahatan dengan tingkatan-tingkatan kriminalitas tertentu di Amerika Serikat [22]. Dalam bidang lain dilakukan penelitian untuk mengembangkan metode Fuzzy C-Means Clustering berdasarkan deferential evolutionary untuk melakukan studi klaster massa batuan. Hasilnya adalah bahwa metode baru ini memberikan nilai terbaik untuk semua kriteria serta memiliki stabilitas yang baik dalam berbagai kriteria yang dipertimbangkan dalam penelitian [23]. Metode Fuzzy C-Means ini bersama metode K-Means telah diterapkan untuk klasterisasi dan pemetaan lahan pertanian

di Minahasa Tenggara [24]. Beberapa penelitian klasterisasi kawasan di Indonesia terkait data COVID-19 telah dilakukan.

Metode K-Means Clustering telah diterapkan berdasarkan data jumlah kasus COVID-19, sembuh, meninggal dan CFR [25], [26]. Metode yang sama juga telah dibahas untuk diterapkan khusus untuk provinsi DKI Jakarta [27]. Metode klasterisasi lain yang diteliti dan diterapkan di Indonesia adalah algoritma K-Medoids Clustering berdasarkan data yang sama [28]. Untuk tujuan yang sama beberapa peneliti lain memilih untuk menggunakan pendekatan data map [29], [30]. Hierarchical clustering juga telah diterapkan untuk klasterisasi negara di dunia menggunakan data COVID-19 [31]. Meskipun demikian, untuk mengelompokan provinsi-provinsi yang memiliki resiko yang serupa akibat penyebaran COVID-19 tidak hanya bisa menggunakan data perkembangan kasus. Data lain seperti luas wilayah provinsi, jumlah penduduk, jumlah rumah sakit dan fasilitas kesehatan juga perlu dipertimbangkan. Pada penelitian ini, klasterisasi dilakukan menggunakan metode Fuzzy C-Means Clustering dengan Reduksi Core and Reduct. Data yang digunakan adalah data konfirmasi positif, data pasien sembuh, pasien meninggal, jumlah penduduk, luas wilayah, jumlah pulau,jumlah rumah sakit rujukan dan koordinat provinsi.

1. Method
2. **Fuzzy C-Means Clustering**

Fuzzy c-means (FCM) adalah metode clustering yang memungkinkan satu bagian data menjadi milik dua atau lebih cluster. Metode ini dicetuskan oleh Dunn pada tahun 1973 dan dikembangkan oleh Bezdek pada tahun 1980, metode ini sering digunakan dalam pengenalan pola[32]. Metode ini didasarkan pada minimalisasi fungsi objektif berikut:

 (2)

Keterangan:

: indeks *fuzzines* sebarang bilangan real yang lebih dari 1

: matriks derajat keanggotaan dari  pada klister 

: data ke-

: pusat klaster ke-

N : jumlah data

Partisi *fuzzy* dilakukan melalui optimasi berulang dari fungsi objektif yang ditunjukkan di atas, dengan pembaruan matriks keanggotaan  dan pusat klaster  oleh:

 (3)

dan

 (4)

Iterasi ini akan berhenti ketika  di mana  adalah adalah kriteria penghentian antara 0 dan 1, sedangkan  adalah langkah iterasi. Secara intuitif, matriks keanggotaan akan merepresentasikan kemungkinan keanggotanan sebuah data berdasarkan jaraknya. Semakin dekat sebuah data ke salah satu pusat kalster, maka semakin tinggi pula nilai kemungkinannya. Maka nilai dari fungsi objektifpun akan terus menurun. Artinya prosedur ini konvergen menuju ke minimum lokal atau titik pelana .

Untuk melakukan klasterisasi dengan teknik ini, dilakukan langkah-langkah sebagai berikut [17]:

1. Mengambil data yang mengandung variabel random dan  yang menyatakan objek dan atribut. Data berupa matriks yang berukuran  , yang mana k banyak data dan n adalah banyaknya atribut data.
2. Menentukan jumlah klaster = c, iterasi maksimal = MaxIter, error terkecil yang diharapkan =  dengan iterasi awal t=1 dan 
3. Menentukan matriks partisi awal  sebarang berukuran  yang terdiri dari bilangan randomsedemikian sehingga jumlah keseluruhan matriks partisi baru dalam kelas adalah 1 sesuai dengan kendala (1).



yang mana 

1. Menghitung pusat klaster 



1. Menghitung nilai fungsi objektif  menggunakan (1).
2. Menghitung perubahan matriks partisi.
3. Memeriksa kondisi untuk berhenti, jika  atau  maka proses penghitungan diberhentikan. Jika belum maka penghitungan terus dilanjutkan hingga salah satu di antara kriteria tersebut dipenuhi.
4. **Metode Elbow**

Metode Elbow merupakan metode yang bertujuan untuk menghasilkan informasi dalam menentukan jumlah cluster rekomendadi dengan cara melihat persentase hasil perbandingan antara jumlah cluster yang akan membentuk siku pada suatu titik [33].

Berikut ini tahapan algoritma metode Elbow dalam menentukan nilai k pada metode klasterisasi [33]:

1. Menginisialisasi awal nilai k ;
2. Menaikan nilai k ;
3. Menghitung hasil *sum of square error* dari tiap nilai k ;
4. Analisis hasil sum of square error dari nilai k yang mengalami penurunan secara drastis ;
5. Cari dan tetapkan nilai k yang berbentuk siku.

Pada metode Elbow nilai cluster terbaik yang akan diambil dari nilai *Sum of Square Error* (SSE) yang mengalami penurunan yang signifikan dan berbentuk siku.Untuk menghitung SSE menggunakan rumus (2). *Sum of Square Error* (SSE digunakan untuk mengukur perbedaan antara data yang diperoleh dengan model perkiraan yang telah dilakukan sebelumnya.

1. **Reduksi Dimensi Core and Reduct**

Teori *rough set* diajukan oleh seorang saintis asal Polandia Zdzislaw Pawlak [34]. Makalah pertama yang berisi membahas gagasan ini dipublikasi pada tahun 1982 *di International Journal of Computer and Information Sciences* [35]. Ide ini dikembangan untuk peningkatan *automatic generation systems* dan masalah *soft computing*. Teori ini menyediakan *framework* matematika baru untuk memproses informasi yang tidak presisi dan tidak lengkap, serta menemukan aturan tersembunyi dari sejumlah data yang banyak. Dalam banyak aplikasi analisis data, informasi dan pengetahuan disimpan dan diwakili dalam tabel informasi. Tabel informasi menyediakan cara yang mudah untuk menggambarkan sekumpulan objek hingga di dalam alam semesta oleh sekumpulan atribut hingga [36].

Tabel informasi atau juga disebut sistem informasi keputusan T dapat dinyatakan sebagai *quadruple* T=(U,A,C,D), yang mana U adalah himpunan dari semua objek [34]. A adalah himpunan variabel. Jika himpunan A kemudian dibagi menjadi conditional attribute himpunan C dan *decisional attribute* himpunan D, yaitu dan 

Tabel informasi adalah tabel dengan setiap barisnya adalah menyatakan entitas dan setiap kolomnya menyatakan variabel atribut tertentu. Entitas pada himpunan U dapat diklasifikasikan menjadi beberapa *decision class* yang berbeda dan beberapa *conditional attribute* yang berbeda. Tabel 1 merupakan tampilan dari tabel keputusan. Himpunan semesta adalah himpunan objek-objek atau entitas,  adalah himpunan *conditional attribute*, dan  adalah himpunan *decisional attribute*. Suku-suku  menyatakan *conditional attribute* ke-j dari entitas ke-i, dan  menyatakan *conditional attribute* ke-j dari entitas ke-i.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **U** | **C** | | | | **D** | | |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

Gambar 1. Susunan Tabel Infrormasi

Tabel informasi di atas mewakili semua informasi dan pengetahuan yang tersedia [37]. Pada tabel informasi, untuk setiap subset  hubungan *indiscernibility* didefinisikan dengan

 (11)

Pada tabel informasi T, himpunan atribut  disebut *reduct*, jika R memenuhi dua kondisi berikut:

1. 

2. Untuk setiap 

Kondisi satu menyatakan bahwa untuk setiap pasangan objek yang tidak dapat dibedakan oleh subset R maka juga tidak dapat dibedakan oleh A, dan sebaliknya. Kondisi kedua menyatakan bahwa ada pasangan objek yang tidak dapat dibedakan oleh R - {a} tetapi dapat dibedakan dengan A. Ini berarti bahwa R adalah set atribut minimum yang dapat menjaga hubungan *indiscernibility* IND(A). Biasanya, ada lebih dari satu reduksi dalam suatu tabel informasi. Himpunan semua reduksi dari tabel informasi T dilambangkan sebagai *RED (T).* Kemudian, *core* dari set atribut  adalah sebagai berikut:

 (12)

1. **Dataset**

Data yang digunakan adalah data jumlah pasien covid perprovinsi, jumlah kepadatan penduduk per provinsi, karakter wilayah per provinsi, dan jumlah fasilitas kesehatan per provinsi yang diperoleh dari halaman resmi covid19.go.id, kemenkes.go.id, dan bps.go.id. Tabel 1 menyajikan variabel-variabel yang digunakan dalam penelitian ini.

Table 1. Variabel Dataset

|  |  |  |
| --- | --- | --- |
| **Variable** | **Type Data** |  |
| Nama Provinsi | String |  |
| Jumlah Penduduk | Numeric |  |
| Luas Wilayah | Numeric |  |
| Persentase Luas Wilayah | Numeric |  |
| Jumlah Pulau | Numeric |  |
| Jumlah RS Rujukan | Numeric |  |
| Longitude | Numeric |  |
| Latitude | Numeric |  |
| Konfirmasi Positif | Numeric |  |
| Meninggal | Numeric |  |
| Sembuh | Numeric |  |
| Jumlah data | 34 |  |

1. **Algoritma Fuzzy C-Means Clustering dengan Reduksi Dimensi Core and Reduct**

Dalam proses klasterisasi COVID-19 Indonesia berdasarkan Provinsi menggunakan metode Fuzzy C-Means Clustering dengan reduksi dimensi Core and Reduct mengikuti Algoritma 1 berikut:

Algoritma 1. Fuzzy C-Means Core & Reduct Clustering

***Input:***

Input data berupa variabel  dan  yang menyatakan objek dan atribut. Data berupa matriks yang berukuran , yang mana n banyak data dan m adalah banyaknya atribut data.

*Proses Mulai:*

1. Normalisasi data menggunakan rumus:

 (12)

1. Menerapkan metode core and reduct sehingga jumlah variabel  akan menjadi sejumlah variabel baru , dengan .
2. Menerapkan metode fuzzy c-means clustering sehingga diperoleh klaster-klaster data.

Proses Selesai:

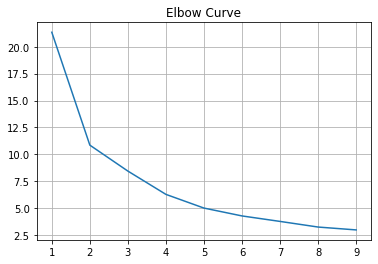
***Output:***

Hasil Label Klasterisasi

1. Results and Discussion

Data yang digunakan memiliki 11 variabel dan 34 baris data. Setelah melalui proses reduksi diperoleh tetap 11 variabel data. Langkah pertama adalah melakukan data preprocessing. Data numerik yang digunakan pada setiap variabel memiliki skala yang jauh berbeda. Terdapat data yang nilainya puluhan dan ratusan(data RS rujukan). Namun juga terdapat data yang nlainya jutaan(Jumlah Penduduk). Maka dilakukan normalisasi data dengan membagi data dengan range data di masing-masing variabel sebagaimana disajikan pada persamaan (12).

Kemudian selanjutnya dilakukan proses reduksi dimensi. Hasilnya adalah dataset baru dengan tetap 11 variabel. Ini bermakna bahwa setiap variabel dalam dataset bersifat penting dan sama sekali tidak identic. Oleh sebab itu metode klasterisasi selanjutnya akan dijalankan pada dataset utuh ini. Setelah data melalui proses reduksi dimensi, langkah selanjutnya adalah menentukan rekomendasi jumlah klister menggunakan metode Elbow. Hasil dari metode Elbow disajikan pada Gambar 3.



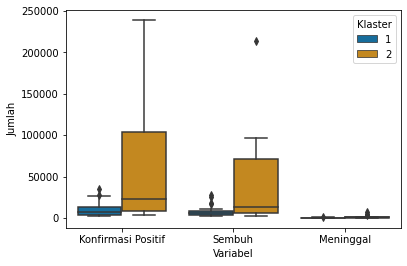
Gambar 3. Kurva Elbow

Berdasarkan Gambar 3, maka posisi “Siku lengan” dalam kurva Elbow ini paling nampak ter;ihat pada n=2. Sehingga jumlah klister rekomendasi yang dihasilkan adalah 2. Nilai ini sama dengan klasterisasi yang dilakukan pemerintah saat ini. Yaitu membagi provinsi-provinsi menjadi provinsi yang menerapkan kebijakan PSBB dan tidak menerapkan kebijakan PSBB. Kemudian metode Fuzzy C-Means Clustering dijalankan pada dataset ini. Hasil klasterisasi provinsi ditunjukkan pada Tabel 2.

Table 2. Hasil Klasterisasi FCM-CR

|  |  |
| --- | --- |
| **Klaster** | **Anggota** |
| Klaster 1 | Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Kepulauan Bangka Belitung, Kepulauan Riau, Kalimantan Barat,  Kalimantan Tengah, Kalimantan Utara, Kalimantan Timur, Sulawesi Utara,  Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara, Papua Barat, Papua |
| Klaster 2 | Lampung, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur ,Kalimantan Selatan, Sulawesi Selatan, Maluku |

Provinsi-provinsi dibagi menjadi 2 klaster dengan klister 1 berisikan 21 provinsi dan klister 22 berisikan 13 provinsi. Hasil ini berbeda jika dibandingkan dengan [25]. Pada makalah tersebut jumlah klister langsung ditentukan sejumlah 3 yaitu Tinggi, Sedang, dan Rendah. Pada penelitian ini jumlah klister ditentukan sesuai rekomendasi dari Metode Elbow. Sebagai catatan data pada penelitian tersebut pada bulan April 2020. Meskipun menggunakan metode Elbow sebagai rekomendasi jumlah klister, hasil pada penelitian [26] juga menunjukkan hasil yang berbeda. Pada penelitian tersebut justru menghasilkan klister yang cukup banyak. Metode K-Means yang digunakan menghasilkan 6 klaster. Sebagai catatan data yang digunakan pada penelitian tersebut adalah sampai bulan November 2020. Gambar 4 menyajikan grafik boxplot antara variabel data konfirmasi positif, meninggal, dan sembuh pada klaster 1 dan 2.



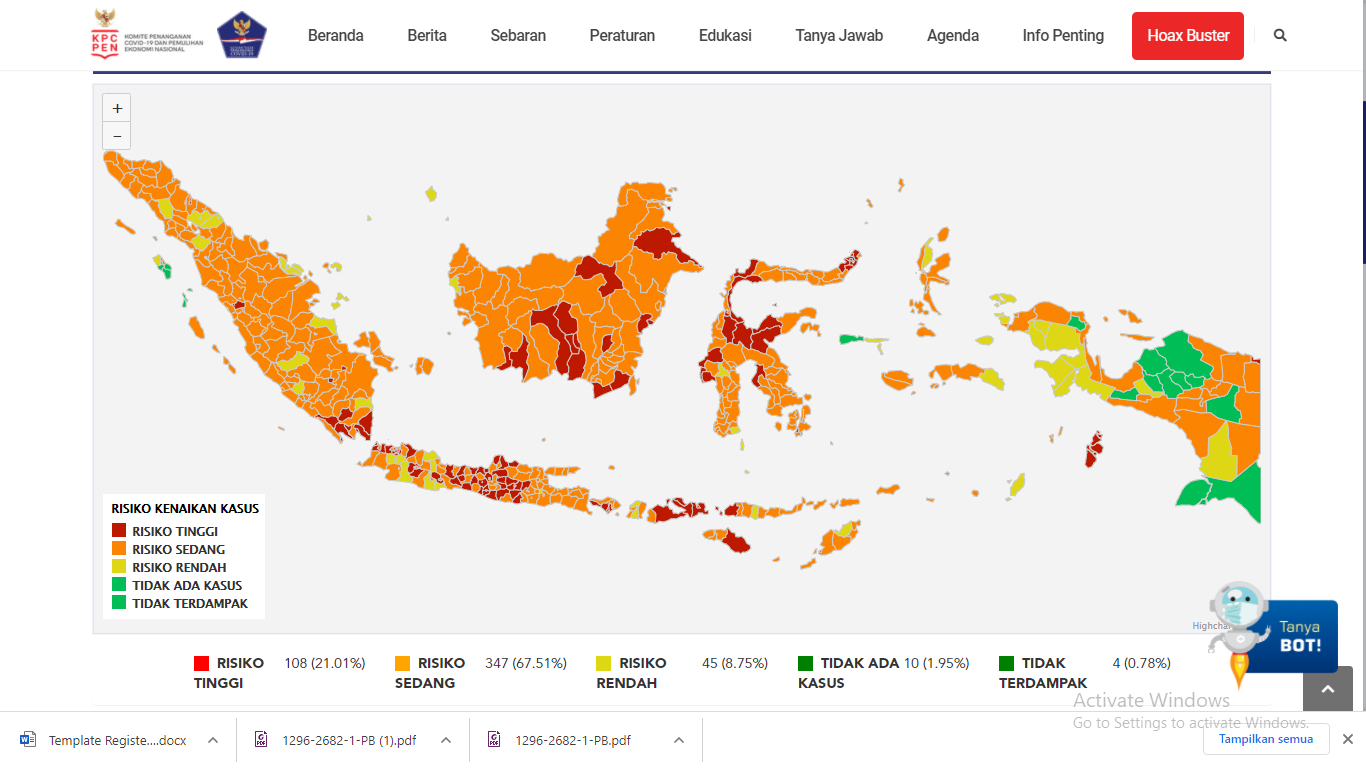
Gambar 4. Box Plot Klaster 1 VS Klaster 2

Berdasarkan Gambar 3 dapat diketahui bahwa Klaster 2 berisi provinsi dengan nilai-nilai parameter COVID-19 yang jauh lebih tinggi. Klaster ini bisa dimungkinkan untuk dijadikan klister yang menerima kebijakan PSBB. Tabel 3 menyajikan perbandingan nama-nama Provinsi yang menerima kebijakan PSBB mulai dari tanggal 11 Januari 2021 hingga 25 Januari 2021 dan nama-nama Provinsi pada klister 2 hasil penelitian ini.

Table 3 Perbandingan Klaster 2 dengan Provinsi PSBB

|  |  |
| --- | --- |
| **Klaster** | **Anggota** |
| Provinsi PSBB | DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, |
| Klaster 2  FCM-CR | Lampung, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur ,Kalimantan Selatan, Sulawesi Selatan, Maluku |

Pada Tabel 3 dapat dilihat bahwa semua Provinsi yang telah menerapkan kebijakan PSBB masuk ke dalam Klaster 2. Ini berarti hasil klasterisasi Fuzzy C-Means dengan Reduksi Dimensi Core and Reduct mendekati kebijakan yang telah diterapkan. Gambar 5 menyajikan peta resiko COVID-19 yang dapat dilihat di situs covid19.go.id. Gambar ini diambil pada tanggal 23 Januari 2021.



Gambar 5. Peta Resiko COVID-19 Indonesia

Berdasarkan gambar 5 dapat dilihat bawah ada beberapa daerah dengan jumlah kabupaten dengan resiko tinggi yang cukup banyak namun tidak diterapkan PSBB. Tentu, kebijakan ini sangat bergantung pada parameter-parameter yang digunakan. Perlu dipahami indikator-indikator yang digunakan sebagai berikut:

Indikator Epidemiologi:

1. Penurunan jumlah kasus positif & probable pada minggu terakhir sebesar ≥50% dari puncak
2. Penurunan jumlah kasus suspek pada minggu terakhir sebesar ≥50% dari puncak
3. Penurunan jumlah meninggal kasus positif & probable pada minggu terakhir sebesar ≥50% dari puncak
4. Penurunan jumlah meninggal kasus suspek pada minggu terakhir sebesar ≥50% dari puncak
5. Penurunan jumlah kasus positif & probable yang dirawat di RS pada minggu terakhir sebesar ≥50% dari puncak
6. Penurunan jumlah kasus suspek yang dirawat di RS pada minggu terakhir sebesar ≥50% dari puncak
7. Persentase kumulatif kasus sembuh dari seluruh kasus positif & probable
8. Laju insidensi kasus positif per 100,000 penduduk
9. Mortality rate kasus positif per 100,000 penduduk
10. Kecepatan Laju Insidensi per 100,000 penduduk

Indikator Surveilans Kesehatan Masyarakat:

1. Jumlah pemeriksaan sampel diagnosis meningkat selama 2 minggu terakhir
2. Positivity rate rendah (target ≤5% sampel positif dari seluruh orang yang diperiksa)

Indikator Pelayanan Kesehatan:

1. Jumlah tempat tidur di ruang isolasi RS Rujukan mampu menampung s.d >20% jumlah pasien positif COVID-19 yang dirawat di RS
2. Jumlah tempat tidur di RS Rujukan mampu menampung s.d >20% jumlah ODP, PDP, dan pasien positif COVID-19 yang dirawat di RS

Data yang digunakan untuk menghitung indikator-indikator di atas terbatas pada :

1. Data kasus positif dan pemeriksaan laboratorium berdasarkan data surveilans Kementerian Kesehatan.
2. Data pasien ODP, PDP, dan kapasitas pelayanan RS didapatkan berdasarkan data RS Online di bawah koordinasi Dirjen Pelayanan Kesehatan Kementerian Kesehatan.

Gambar 6 menujukkan pemetaan provinsi berdasarkan klasterisasi yang diperoleh menggunakan metode Fuzzy C-Means Clustering.



Gambar 6. Hasil Klasterisasi dengan Fuzzy C-Means dengan Reduksi Dimensi Core and Reduct

Klasterisasi ini didasarkan pada data konfirmasi positif, data pasien sembuh, pasien meninggal, jumlah penduduk, luas wilayah, jumlah pulau,jumlah rumah sakit rujukan dan koordinat provinsi. Terdapat beberapa provinsi yang masuk ke dalam Klaster 2 namun tidak mendapatkan perlakuan yang sama seperti 7 provinsi yang telah di PSBB. Kelima provinsi tersebut adalah Lampung, Nusa Tenggara Timur, Nusa Tenggara Barat, Kalimantan Selatan, Sulawesi Selatan, dan Maluku.

1. Conclusions

Fuzzy C-Means Clustering dengan Reduksi Dimensi Core and Reduct dapat diterpkan dengan baik untuk klasterisasi COVID-19 Indonesia berdasarkan Provinsi. Ada dua klaster yang dihasilkan Klaster 1 terdiri dari 21 Provinsi dan Klaster 2 terdiri dari 13 Provinsi. Klaster 2 memiliki nilai parameter COVID-19 seperti jumlah konfirmasi positif, sembuh dan meninggal yang jauh lebih tinggi dari klister 1. Tujuh provinsi yang telah mendapatkan kebijakan PSBB masuk ke dalam klaster 2. Provinsi-provinsi di luar 7 provinsi yang PSBB namun merupakan anggota klaster 2 adalah Lampung, Nusa Tenggara Timur, Nusa Tenggara Barat, Kalimantan Selatan, Sulawesi Selatan, dan Maluku. Kelima provinsi tersebut direkomendasikan untuk memperoleh perlakuan yang sama.

# References

[1] A. Susilo *et al.*, “Coronavirus Disease 2019 : Tinjauan Literatur Terkini Coronavirus Disease 2019 : Review of Current Literatures,” *J. Penyakit Dalam Indones.*, vol. 7, no. 1, pp. 45–67, 2020.

[2] A. Ristyawati, “Efektifitas Kebijakan Pembatasan Sosial Berskala Besar Dalam Masa Pandemi Corona Virus 2019 oleh Pemerintah Sesuai Amanat UUD NRI Tahun 1945,” *Adm. Law Gov. J.*, vol. 3, no. 2, pp. 240–249, 2020.

[3] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.

[4] M. Kamber, *Data Mining : Concepts and Techniques University of Illinois at Urbana-Champaign*. 2016.

[5] A. Gosain and S. Dahiya, “Performance Analysis of Various Fuzzy Clustering Algorithms: A Review,” *Procedia Comput. Sci.*, vol. 79, pp. 100–111, 2016.

[6] D. J. Bora and D. A. K. Gupta, “A Comparative study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm,” *Int. J. Comput. Trends Technol.*, vol. 10, no. 2, pp. 108–113, 2014.

[7] N. Grover, “A study of various Fuzzy Clustering Algorithms,” *Int. J. Eng. Res.*, vol. 3, no. 3, pp. 177–181, 2014.

[8] J. Nayak, B. Naik, D. P. Kanungo, and H. S. Behera, “A hybrid elicit teaching learning based optimization with fuzzy c-means (ETLBO-FCM) algorithm for data clustering,” *Ain Shams Eng. J.*, vol. 9, no. 3, pp. 379–393, 2018.

[9] Shraddha Pandit and Suchita Gupta, “A Comparative Study on Distance Measuring Approaches for Clustering,” *Int. J. Res. Comput. Sci.*, vol. 2, no. 1, pp. 29–31, 2011.

[10] Z. Shi and L. S. C. Pun-Cheng, “Spatiotemporal data clustering: A survey of methods,” *ISPRS Int. J. Geo-Information*, vol. 8, no. 3, 2019.

[11] A. Sarveniazi, “An Actual Survey of Dimensionality Reduction,” no. March, pp. 55–72, 2014.

[12] B. R. A. Moreira *et al.*, “Classifying Hybrids of Energy Cane for Production of Bioethanol and Cogeneration of Biomass-Based Electricity by Principal Component Analysis-Linked Fuzzy C-Means Clustering Algorithm,” *J. Agric. Sci.*, vol. 11, no. 14, p. 246, 2019.

[13] R. Vashist and M. . Garg, “Rule Generation based on Reduct and Core: A Rough Set Approach,” *Int. J. Comput. Appl.*, vol. 29, no. 9, pp. 1–5, 2011.

[14] J. Gould, “Variables in Research Designs,” *Concise Handb. Exp. Methods Behav. Biol. Sci.*, vol. 3, no. 4, pp. 75–110, 2001.

[15] B. Marr, *Big Data In Practice*, 1st ed. West Sussex: Wiley, 2016.

[16] C. C. Drovandi, C. C. Holmes, J. M. McGree, K. Mengersen, S. Richardson, and E. G. Ryan, “Principles of experimental design for Big Data analysis,” *Stat. Sci.*, vol. 32, no. 3, pp. 385–404, 2017.

[17] J. Eliyanto, Sugiyarto, Suparman, I. Djakaria, and M. A. H. Ruhama, “Dimension reduction using core and reduct to improve fuzzy C-means clustering performance,” *Technol. Reports Kansai Univ.*, vol. 62, no. 6, pp. 2855–2867, 2020.

[18] M. C. Hung and D. L. Yang, “An efficient fuzzy C-means clustering algorithm,” *Proc. - IEEE Int. Conf. Data Mining, ICDM*, no. February 2001, pp. 225–232, 2001.

[19] J. Eliyanto, “Meningkatkan Performa Fuzzy Clustering dengan Principal Component Analysis,” vol. 2, no. 2721, pp. 1–8, 2020.

[20] J. Eliyanto, “Reduksi Dimensi untuk Meningkatkan Performa Metode Fuzzy Klastering pada Big Data,” vol. 1, no. 1, pp. 27–36, 2019.

[21] S. Surono and R. D. A. Putri, “Optimization of Fuzzy C-Means Clustering Algorithm with Combination of Minkowski and Chebyshev Distance Using Principal Component Analysis,” *Int. J. Fuzzy Syst.*, 2020.

[22] M. Premasundari and C. Yamini, “a Violent Crime Analysis Using Fuzzy C-Means Clustering Approach,” *Ictact J. Soft Comput.*, vol. 09, no. April, pp. 2229–6956, 2019.

[23] A. Esmaeilzadeh and K. Shahriar, “Optimized fuzzy cmeans – fuzzy covariance – fuzzy maximum likelihood estimation clustering method based on deferential evolutionary optimization algorithm for identification of rock mass discontinuities sets,” *Period. Polytech. Civ. Eng.*, vol. 63, no. 2, pp. 674–686, 2019.

[24] J. Tamaela, E. Sediyono, and A. Setiawan, “Cluster Analysis Menggunakan Algoritma Fuzzy C-means dan K-means Untuk Klasterisasi dan Pemetaan Lahan Pertanian di Minahasa Tenggara,” *J. Buana Inform.*, vol. 8, no. 3, pp. 151–160, 2017.

[25] R. P. R and Y. A. E, “dengan Metode K - Means Clustering,” no. May, 2020.

[26] F. Virgantari and Y. E. Faridhan, “K-Means Clustering of COVID-19 Cases i n Indonesia ’ s Provinces,” 2020.

[27] A. Solichin and K. Khairunnisa, “Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means,” *Fountain Informatics J.*, vol. 5, no. 2, p. 52, 2020.

[28] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. Ilmi R.H.Zer, and D. Hartama, “Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia,” *Jti (Jurnal Teknol. Informasi)*, vol. 4, no. 1, pp. 166–173, 2020.

[29] S. S. Vallery, Happy Novita, “Jurnal Mantik,” *J. Mantik*, vol. 3, no. January, pp. 31–38, 2019.

[30] T. Khotimah *et al.*, “Clustering Perkembangan Kasus Covid-19 Di Indonesia Clustering The Development Of The Covid-19 Case In Indonesia,” vol. 1, no. 1, pp. 23–26, 2020.

[31] V. Zarikas, S. G. Poulopoulos, Z. Gareiou, and E. Zervas, “Clustering analysis of countries using the COVID-19 cases dataset,” *Data Br.*, vol. 31, p. 105787, 2020.

[32] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. London: Plenum Press, 1981.

[33] A. T. Rahman, Wiranto, and A. Rini, “Coal Trade Data Clustering Using K-Means (Case Study Pt. Global Bangkit Utama),” *ITSMART J. Teknol. dan Inf.*, vol. 6, no. 1, pp. 24–31, 2017.

[34] D. Xue and Y. Chen, *Solving Applied Mathematical Problems With MATLAB*. London: CRC Press, 2009.

[35] Z. Pawlak, “Information systems theoretical foundations,” *Inf. Syst.*, vol. 6, no. 3, pp. 205–218, 1981.

[36] Z. Pawlak, *Series D : System Theory , Knowledge Engineering and Problem Solving*. Warsaw: Kluwer Academic, 1991.

[37] Y. Yao and Y. Zhao, “Discernibility matrix simplification for constructing attribute reducts,” *Inf. Sci. (Ny).*, vol. 179, no. 7, pp. 867–882, 2009.