

# Penerapan K-Means dan Fuzzy C-Means untuk Pengelompokan Data Kasus Covid-19 di Kabupaten Indragiri Hilir

Sania Fitri Octavia<sup>1</sup>, Mustakim<sup>2</sup>

<sup>1,2</sup>Puzzle Research Data Technology (Predatech), Faculty of Science and Technology

<sup>1,2</sup>Department of Information System, Faculty of Science and Technology

Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: <sup>1</sup>saniafitri2017@gmail.com, <sup>2</sup>mustakim@uin-suska.ac.id

Submitted: 05/09/2021; Accepted: 25/09/2021; Published: 29/09/2021

**Abstrak**—Diawal tahun 2020 dunia sedang diguncangkan dengan merebaknya virus baru yaitu Corona Virus Disease (Covid-19). Virus ini menyebar dengan cepat hampir ke setiap negara, termasuk Indonesia. Penyebaran virus Covid-19 mulai terjadi di berbagai wilayah di Indonesia yang semakin hari semakin meningkat. Pada penelitian ini dilakukan pengelompokan wilayah yang terkena kasus Covid-19 di Kabupaten Indragiri Hilir guna untuk memberikan informasi terhadap pemerintah pusat mengenai penanganan terhadap Covid-19. Untuk melakukan klusterisasi pada penelitian ini akan menggunakan algoritma K-Means dan Fuzzy C-Means. Setelah dilakukan beberapa percobaan didapatkan perbandingan validitas yang diuji dengan Silhouette Index dan Partition Coefficient Index, yaitu nilai validitas SI dari K-Means adalah 0,950 sedangkan nilai validitas PCI dari Fuzzy C-Means adalah 0,960. Hasil yang diperoleh menunjukkan bahwa metode Fuzzy C-Means merupakan metode terbaik untuk melakukan clustering data Covid-19 di Kabupaten Indragiri Hilir dikarenakan nilai validitasnya mendekati 1 yang terletak pada K=3

**Kata Kunci:** Covid-19; Klustering; Fuzzy C-Means; K-Means; Partition Coefficient Index; Silhouette Index

**Abstract**—In the beginning of 2020 world was shocked because new virus spreaded, that is Coronavirus Disease 2019 (Covid-19). This virus spread quickly in almost country, including Indonesia. Covid-19 virus deployment started in various regions in Indonesia stay increasing everyday. this research has been done the region clustering that infected Covid-19 case in Indragiri Hilir district to inform to central government about Covid-19 handling. To do Clustering in this research used K-Means and Fuzzy C-Means Algorithm. After done some of test, it's obtained the ratio which was tested with Silhouette Index and Partition Coefficient, SI validity value of K-Means is 0,950 while PCI validity value of Fuzzy C-Means is 0,960. The results have been obtained shown that Fuzzy C-Means Method is the best Method to do Clustering Covid-19 data in Indragiri Hilir district Because the validity value is closed to 1 which is located in K=3.

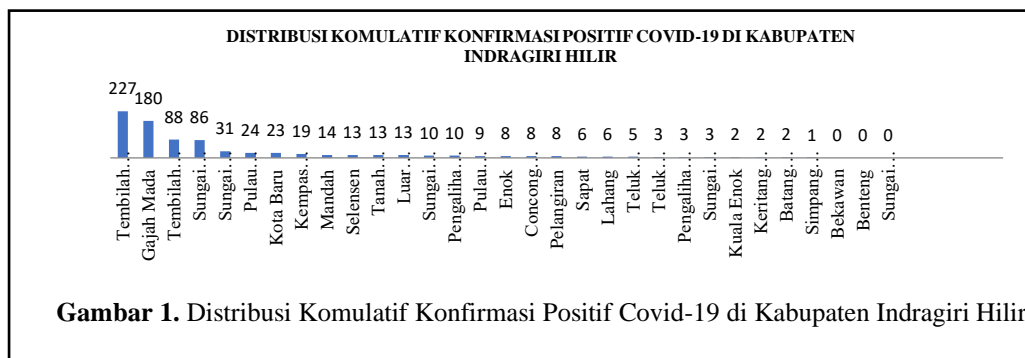
**Keywords:** Covid-19; Clustering; Fuzzy C-Means; K-Means; Partition Coefficient Index; Silhouette Index

## 1. PENDAHULUAN

Virus Covid-19 diidentifikasi pertama kali pada akhir Desember 2019 di Wuhan, atau lebih tepatnya ibu kota Provinsi Hubei, China. Corona Virus Disease atau Covid-19 adalah penyakit menular yang ditandai dengan gejala saluran pernapasan akut (Coronavirus 2 (Severe Acute Respiratory Syndrome Coronavirus 2 atau SARSCoV2) . Gejala yang biasa dialami oleh korban Covid-19 adalah mudah letih, demam dan batuk kering [2]. Gejala lain yang kurang dialami beberapa pasien antara lain nyeri, hidung tersumbat, sakit kepala, sakit tenggorokan, diare, kehilangan rasa dan penciuman, ruam bahkan jari tangan dan jari kaki berubah warna. World Health Organization (WHO) menyatakan bahwa virus Covid-19 dapat bertahan hidup berjam-jam, berhari-hari pada suhu 26-27 °C. Penularan virus dapat menular melalui droplet (percikan sangat kecil) yang mengandung virus penyakit dan dikeluarkan oleh penderita ketika berbicara, batuk ataupun bersin. Virus ini menyebar ke hampir setiap negara, termasuk Indonesia [3].

Indonesia mengonfirmasi kasus virus Covid-19 pertama kali pada Senin, 2 maret 2020 lalu dengan 2 kasus terkonfirmasi [4]. Berdasarkan data yang diperoleh dari situs web Covid19.go.id pada 04 Desember 2021, yang terkonfirmasi positif Covid-19 berjumlah 4.257.489 dengan jumlah kematian 143.863 dan sembuh 4.105.994. Keberadaan virus Covid-19 ini tentu dapat melumpuhkan semua kegiatan seperti kegiatan sosial dan kegiatan perekonomian. Pencegahan yang dapat dilakukan adalah menjaga jarak fisik, menggunakan masker dan tidak lupa untuk menerapkan pola hidup sehat. Pemerintah juga dapat menerapkan aktivitas keluar rumah dan anjuran physical distancing untuk mengantisipasi kenaikan jumlah kasus Covid-19. Adapun kebijakan yang paling dikedepankan adalah Pembatasan Sosial Berskala Besar (PSBB) di setiap wilayah guna memutuskan rantai penyebaran Covid-19 ini.

Penyebaran kasus Covid-19 yang tersebar diseluruh Provinsi dan Kabupaten di Indonesia relatif cepat dan berdampak negatif di seluruh wilayah [5]. Salah satunya yang tak lepas dari serangan covid-19 adalah Kabupaten Indragiri Hilir. Berdasarkan data yang diperoleh dari Dinas Kesehatan sejak 03 Maret hingga 04 Oktober 2021, di Kabupaten Indragiri Hilir terdapat jumlah kasus Covid-19 yang terkonfirmasi sebanyak 5613, sembuh hingga 5374, dalam perawatan sebanyak 51 dan meninggal sebanyak 188 orang, Berikut grafik konfirmasi positif Covid-19 di Kabupaten indragiri Hilir dapat dilihat pada Gambar 1.



Luasnya wilayah di Kabupaten Indragiri Hilir memungkinkan diperlukannya pengelompokan bagian bagian berdasarkan wilayah. Dalam upaya penanganan tersebut diperlukan faktor-faktor agar penanganannya tepat. Maka diperlukanlah suatu upaya untuk mengantisipasi penularan virus COVID-19 tersebut dengan mengelompokkan suatu kawasan agar dapat dilakukan prioritas penanganan, pengawasan dan peningkatan layanan kesehatan masyarakat. Teknik clustering pada data mining dapat digunakan untuk klusterisasi kasus Covid-19.

Data mining adalah metode menggali, mencari, dan menemukan informasi atau wawasan penting dalam database besar. Salah satu tujuan dari data mining adalah untuk menemukan pola-pola tertentu yang harus dicapai sehingga dapat digunakan untuk memprediksi dalam pengambilan keputusan yang akan datang[6]. Clustering adalah teknik data mining yang bertujuan untuk mengelompokkan data (objek) menjadi beberapa cluster atau kelompok. Kemudian objek serupa digabungkan ke dalam cluster yang sama dan objek yang berbeda menjadi bagian dari cluster yang berbeda[7]. Terdapat beberapa algoritma yang ada dalam clustering seperti K-Means Clustering dan Fuzzy C-Means.

Menurut Athifaturrofifah, Goejantoro & Yuniarti (2018) dalam penelitiannya menyatakan bahwa K-Means dapat memberikan hasil clustering yang lebih baik karena metode K-Means memberikan nilai SC yang lebih besar daripada K-Medoids. Menurut Syarif, Furqon & Adinugroho (2018), Algoritma K-Means ini mudah dipahami dan memiliki tingkat kompleksitas yang cukup efisien sehingga algoritma ini sering digunakan dalam berbagai kasus [8]. Menurut Yohanes (2016) berdasarkan hasil penelitiannya menunjukkan bahwa K-Means lebih cepat dalam hal waktu clustering. Sedangkan Fuzzy C-Means secara komputasi lebih baik dalam menentukan derajat keanggotaan cluster dalam pengelompokan data [9]. Menurut Fajar Agustini (2017) pada penelitiannya juga menyatakan bahwa algoritma Fuzzy C-Means memiliki kelebihan dalam penempatan pusat cluster yang lebih tepat jika dibandingkan dengan metode *cluster* lainnya.

Berdasarkan uraian diatas dari beberapa penelitian sebelumnya. Penelitian kedua metode tersebut memberikan hasil yang signifikan dan berbeda dalam pola cluster. Pengujian performasi algoritma adalah salah satu cara dalam clustering untuk menemukan metode mana yang baik untuk digunakan dalam pengelompokan. Oleh karena itu penelitian ini akan ditunjukkan perbandingan clustering data Covid-19 di Kabupaten Indragiri Hilir menggunakan Algoritma K-Means dan Fuzzy C-Means, yang bertujuan untuk mengetahui kemampuan masing-masing algoritma dalam melakukan clustering untuk mendapatkan hasil algoritma terbaik. Diharapkan dapat membantu pemerintah Kabupaten Indragiri Hilir dalam mengambil keputusan yang tepat serta mengoptimalkan penanganan Covid-19 di Kabupaten Indragiri Hilir.

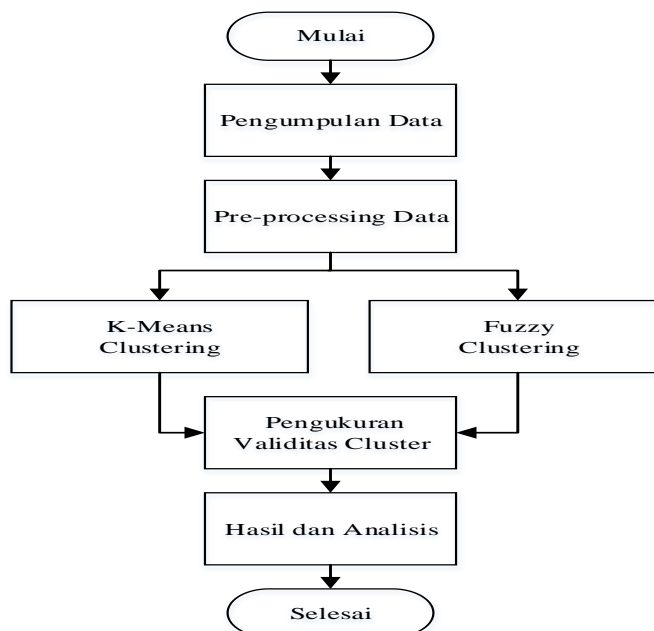
## 2. METODOLOGI PENELITIAN

### 2.1 Covid-19

Covid-19 atau dikenal dengan nama Virus Corona merupakan jenis baru dari coronavirus yang dapat menular ke manusia [9]. Covid-19 adalah suatu penyakit yang disebabkan oleh virus SARS-CoV-2 yang memiliki gejala mirip dengan flu biasa yang jika berlanjut dapat menyebabkan masalah pernapasan dan radang paru (Pneumonia) [10]. Gangguan saluran pernafasan pada pasien merupakan gejala utama yang ditimbulkan dari penularan Virus Corona . Gangguan pernapasan yang disebabkan oleh virus corona dimulai dari penyakit ringan seperti flu biasa hingga infeksi pernapasan seperti infeksi paru-paru (pneumonia). Pada gejala pertama, penderita biasanya mengalami batuk kering, flu, sakit kepala, demam, dan sakit tenggorokan [11].

### 2.2 Tahapan Penelitian

Pada penelitian ini terbagi menjadi lima tahapan utama yaitu pengumpulan data, pre-processing data, proses klusterisasi menggunakan algoritma, pengukuran validitas cluster serta analisis dan hasil algoritma terbaik. Metodologi penelitian ini dapat dilihat pada gambar 1.



**Gambar 1.** Metodologi Penelitian

Pada penelitian ini dataset yang digunakan adalah data kasus Covid-19 di Kabupaten Indragiri Hilir yang diperoleh langsung dari Dinas Kesehatan Kabupaten Indragiri Hilir pada tanggal 03 Maret 2020 hingga 4 Oktober 2021. Kemudian Langkah selanjutnya adalah melakukan proses preprocessing terhadap data yang diperoleh dengan melakukan cleaning dan normalisasi data. Normalisasi yang diterapkan pada data adalah normalisasi min max. Selanjutnya dilakukan proses clustering dengan menggunakan Algoritma K-Means dan Fuzzy C-Means. Tahap selanjutnya, proses analisis dan hasil dilakukan perhitungan validitas cluster dengan menggunakan Silhouette Index (SI) dan Partition Coefficient Index (PCI), kemudian didapatkan hasil algoritma terbaik.

**2.3 Clustering**

Clustering merupakan suatu proses pengelompokan data kedalam kelas atau cluster yang memiliki kesamaan objek. Dalam Clustering, data yang memiliki kesamaan akan digabungkan ke dalam cluster yang sama, begitupun sebaliknya data yang tidak memiliki kesamaan akan dimasukkan ke dalam cluster yang berbeda [12]. Potensi clustering dapat digunakan untuk mendefinisikan struktur data yang dapat digunakan lebih lanjut dalam berbagai aplikasi seperti klasifikasi, pengolahan gambar dan pengenalan pola [13].

**2.4 K-Means**

K-Means merupakan algoritma klasterisasi yang mengelompokkan data menjadi beberapa kelompok dengan beberapa cluster berdasarkan titik pusat klaster (centroid) terdekat dengan data [14]. Algoritma ini banyak digunakan karna cukup mudah untuk diimplementasikan dan relatif cepat [15]. Metode ini membagi data ke dalam klaster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu klaster yang sama dan sebaliknya data yang memiliki karakteristik berbeda dikelompokkan ke dalam klaster yang berbeda [16]. Untuk menghitung jarak semua data ke setiap centroid dari sebuah cluster, dapat menggunakan jarak Euclidean persamaan berikut:

$$d_{ik} = \sqrt{\sum_{j=1}^m (X_{ij} - C_{kj})^2} \tag{1}$$

**2.5 Fuzzy C-Means**

Fuzzy C-Means merupakan salah satu teknik Fuzzy clustering paling populer yang pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981. Dalam pendekatan ini, titik data memiliki nilai keanggotaannya dengan pusat cluster, yang akan diperbarui secara iterative [17]. Algoritma FCM merupakan algoritma pengelompokan data yang ditentukan oleh derajat keanggotaan. Untuk beberapa objek data yang menunjukkan keanggotaan parsial dan data yang tidak mempunyai batasan pada salah satu kelas saja dapat dikelompokkan berdasarkan derajat keanggotaan yaitu antara 0 sampai dengan 1 [18]. Meminimalkan fungsi objektif atau fungsi fitnes dapat dilakukan untuk mendapatkan struktur klaster dan hasil klaster yang optimal. Fungsi objektif yang dapat digunakan adalah sebagai berikut :

$$jm = \sum_{i=1}^c \sum_{j=1}^n (u_{ij})^m d^2(y_{1,z_1}) \tag{2}$$

$$u_{ij} = \left[ \sum_{k=1}^c \left( \frac{d_{ij}}{d_{kj}} \right)^{\frac{2}{m-1}} \right]^{-1} \quad 1 \leq i \leq c; 1 \leq j \leq n \tag{3}$$



$$Z_i = \frac{\sum_{j=1}^n (u_{ij})^m y_j}{\sum_{j=1}^n (u_{ij})^m} \tag{4}$$

**2.5 Silhouette Index**

SI digunakan untuk memvalidasi sebuah data cluster tunggal maupun cluster keseluruhan [19]. Metode validitas cluster ini dilakukan dengan membandingkan jarak rata-rata objek baik dalam cluster yang sama maupun cluster yang berbeda. Sebuah cluster dinyatakan terbaik jika nilainya mendekati satu [20]. Untuk menghitung nilai SI dapat menggunakan persamaan:

$$SI = \frac{1}{n} \sum_{i=1}^n \left( \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \right) \tag{5}$$

**2.5 Partition Coefficient Index**

PCI hanya menyesuaikan untuk nilai derajat keanggotaan yang berisi informasi geometris. Nilai dalam rentang [0,1] akan menunjukkan kualitas kluster yang semakin baik dengan nilai yang tinggi atau mendekati nilai 1[21]. Untuk menghitung nilai PCI dapat menggunakan persamaan:

$$PCI = \frac{1}{N} (\sum_{i=1}^N \sum_{j=1}^K U_{ij}) \tag{6}$$

**3. HASIL DAN PEMBAHASAN**

**3.1 Pengumpulan dan Preprocessing Data**

Data yang digunakan pada penelitian ini adalah data Covid-19 di Kabupaten Indragiri Hilir yang diperoleh dari Dinas Kesehatan pada 03 Maret 2020 hingga 04 Oktober 2021 dengan jumlah kasus terkonfirmasi sebanyak 5613, sembuh 5374, dalam perawatan sebanyak 51 dan meninggal sebanyak 188 orang. Atribut yang digunakan adalah selesai isolasi mandiri, selesai isolasi rs, total selesai isolasi, suspek meninggal, isolasi mandiri, sembuh, meninggal dan konfirmasi dirawat. Pada penelitian ini data yang digunakan adalah hasil cleaning data yang dapat dilihat pada Tabel 1.

**Tabel 1.** Hasil data cleaning

No.	Kecamatan	Selesai Isolasi Mandiri	Selesai Isolasi Rs	Total Selesai Isolasi	Suspek Meninggal	Isolasi Mandiri	Sembuh	Meninggal	Dirawat
1	Enok	535	17	552	1	2	120	9	2
2	Tanah Merah	178	10	188	2	3	85	1	4
3	Gaung Anak Serka	554	9	563	1	2	66	2	2
4	Keritang	387	47	434	2	4	141	21	4
5	Mandah	220	17	237	2	4	132	3	4
6	Pulau Burung	91	32	123	2	6	289	4	6
7	Sungai Batang	10	3	13	2	4	565	20	4
8	Tembilahan Hulu	1000	58	1058	3	8	121	2	8
...	...	...	...	...	...	...	...	...	...
19	Kempas	1166	17	1183	2	61	5246	181	64

**3.2 Normalisasi Data**

Normalisasi yang diterapkan pada data yang akan diteliti adalah min-max normalization. Hasil normalisasi data ditunjukkan pada tabel 2.

**Tabel 2.** Data Covid-19 hasil Normalisasi Data

No.	Kecamatan	Selesai Isolasi Mandiri	Selesai Isolasi Rs	Total Selesai Isolasi	Suspek Meninggal	Isolasi Mandiri	Sembuh	Meninggal	Dirawat
1	Enok	0,0714	0,0809	0,0715	0,0000	0,0167	0,0201	0,0444	0,0159
2	Tanah Merah	0,0229	0,0405	0,0232	0,0714	0,0333	0,0134	0,0000	0,0476
3	Gaung Anak Serka	0,0740	0,0347	0,0730	0,0000	0,0167	0,0097	0,0056	0,0159
4	Keritang	0,0513	0,2543	0,0559	0,0714	0,0500	0,0241	0,1111	0,0476
5	Mandah	0,0286	0,0809	0,0297	0,0714	0,0500	0,0224	0,0111	0,0476
6	Pulau Burung	0,0111	0,1676	0,0146	0,0714	0,0833	0,0524	0,0167	0,0794
7	Sungai Batang	0,0001	0,0000	0,0000	0,0714	0,0500	0,1051	0,1056	0,0476
8	Tembilahan Hulu	0,1345	0,3179	0,1386	0,1429	0,1167	0,0203	0,0056	0,1111
...	...	...	...	...	...	...	...	...	...
19	Kempas	0,1571	0,0809	0,1552	0,0714	1,0000	1,0000	1,0000	1,0000

### 3.3 Klasterisasi dengan Algoritma

Penelitian ini melakukan klasterisasi dengan algoritma K-Means dan Fuzzy C-Means menggunakan data yang sudah dinormalisasi pada tahap sebelumnya. Dilakukan beberapa percobaan clustering menggunakan algoritma Kmeans dan Fuzzy C-Means untuk menemukan jumlah cluster optimal.

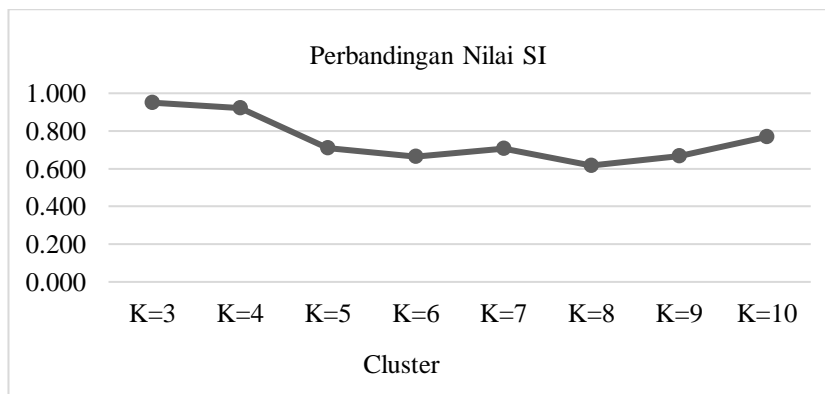
#### 3.3.1 K-Means

Proses ini dilakukan percobaan dengan jumlah 3 cluster sampai dengan 10 cluster bertujuan untuk mendapatkan cluster optimal. Hasil klasterisasi dapat dilihat pada Tabel 3.

**Tabel 3.** Hasil klasterisasi Algoritma K-Means

Percobaan	Klaster									
	1	2	3	4	5	6	7	8	9	10
K=3	17	1	1	-	-	-	-	-	-	-
K=4	16	1	1	1	-	-	-	-	-	-
K=5	13	1	1	1	3	-	-	-	-	-
K=6	3	1	1	1	11	2	-	-	-	-
K=7	1	1	1	1	11	2	2	-	-	-
K=8	1	5	1	1	6	1	2	2	-	-
K=9	8	1	1	1	1	2	1	2	2	-
K=10	2	1	1	1	5	1	1	5	1	1

Setelah dilakukan proses cluster kemudian dilakukan validitas cluster untuk menemukan klaster yang terbaik. Untuk menemukan nilai validitas cluster K-Means menggunakan *silhouette index*. Hasil validitas cluster dapat dilihat pada Gambar 2.



**Gambar 2.** Perbandingan Nilai SI

Dapat dilihat pada Gambar 2 bahwa semua nilai SI pada percobaan ini adalah positif. Artinya objek ke-i sudah berada di kelompok yang benar. Nilai SI tertinggi adalah 0,950 yang terletak pada nilai K=3. Maka didapatkan hasil klaster 1 berjumlah 17 kecamatan, klaster 2 berjumlah 1 1 kecamatan, dan klaster 3 berjumlah 1 kecamatan

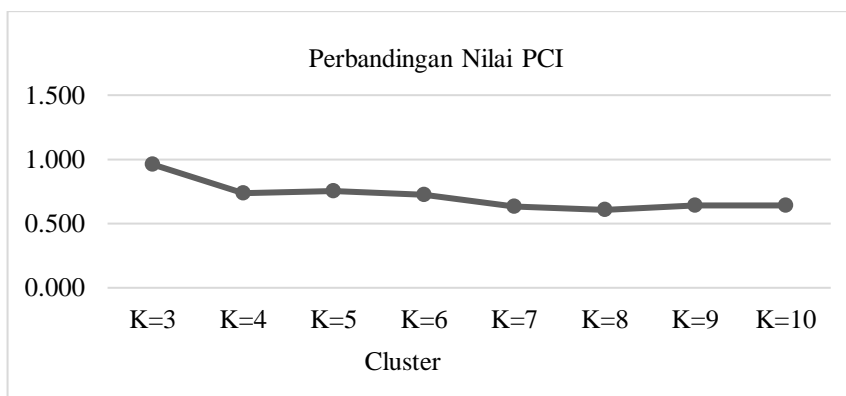
#### 3.3.2 Fuzzy C-Means

Pada proses ini juga dilakukan percobaan dengan jumlah 3 cluster sampai dengan 10 cluster dan melakukan *input* pangkat (w) yaitu 2 dan Maxiter berjumlah 100. Hasil klasterisasi dapat dilihat pada Tabel 4.

**Tabel 4.** Hasil klasterisasi Fuzzy C-Means

Percobaan	Klaster									
	1	2	3	4	5	6	7	8	9	10
K=3	17	1	1	-	-	-	-	-	-	-
K=4	1	11	6	-	-	-	-	-	-	-
K=5	1	1	11	2	4	-	-	-	-	-
K=6	1	3	2	11	1	1	-	-	-	-
K=7	1	5	2	8	1	1	1	-	-	-
K=8	3	1	1	1	5	2	5	1	-	-
K=9	3	2	1	1	5	4	1	1	1	-
K=10	2	1	1	2	1	5	1	3	1	2

Setelah melakukan clustering dengan menggunakan Fuzzy C-Means pada data, kemudian dilakukan validitas cluster untuk mencari cluster yang paling optimal. Untuk menemukan nilai validitas cluster Fuzzy C-Means menggunakan PCI. Hasil validitas cluster dapat dilihat pada Gambar 3.

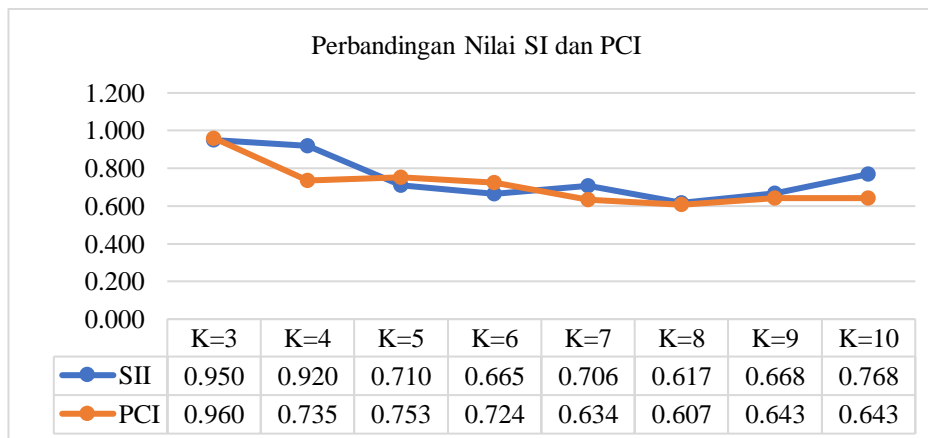


**Gambar 3.** Perbandingan Nilai PCI

Pada Gambar 3 Nilai PCI tertinggi terletak pada nilai K=3 dengan nilai yang diperoleh sebesar 0,960. Kemudian didapatkan hasil yaitu kluster 1 berjumlah 171 kecamatan, kluster 2 berjumlah 1 kecamatan, dan kluster 3 berjumlah 1 kecamatan.

### 3.4 Perbandingan Validitas Cluster

Setelah melakukan percobaan pengelompokkan dengan 2 Algoritma yaitu K-Means dan FCM, dengan percobaan dari 3 kluster sampai dengan 10 kluster maka langkah selanjutnya menentukan kluster terbaik dengan cara membandingkan nilai validitas algoritma K-means dengan SI dan algoritma Fuzzy C-Means dengan PCI pada setiap percobaan algoritma. Dari percobaan yang telah dilakukan dapat dilihat pada Gambar 4.



**Gambar 4.** Perbandingan Nilai SI dan PCI

Pada Gambar 4 hasil validitas K-Means dan FCM memiliki nilai tertinggi yaitu K-means bernilai 0,950 dan nilai Fuzzy C-Means bernilai 0,960. Berdasarkan ketentuan validasi, semakin mendekati 1 maka semakin baik kualitas kluster yang dihasilkan. Sehingga Fuzzy C-Means memiliki validitas cluster paling baik dibandingkan K-Means dengan perbandingan nilai SI dan PCI yaitu 0,960.

## 4. KESIMPULAN

Berdasarkan hasil dan analisis yang dilakukan, dapat disimpulkan bahwa pengelompokkan data Covid-19 menggunakan algoritma K-Means dan Fuzzy C-Means diperoleh cluster terbaik pada K=3. Namun jika dilihat dari segi validitasnya, Nilai SI pada K-Means adalah 0,950 sedangkan nilai PCI pada Fuzzy C-Means adalah 0,960. Sehingga dalam melakukan clustering data Covid-19, Fuzzy C-Means merupakan metode terbaik untuk melakukan pengelompokkan data dikarenakan nilai validitasnya mendekati 1 dengan percobaan nilai K=3. Maka diperoleh hasil kluster 1 terdiri dari 17 kecamatan, kluster 2 berjumlah 1 kecamatan dan kluster 3 berjumlah 1 kecamatan

## UCAPAN TERIMAKASIH

Ucapan terimakasih kepada Dinas Kesehatan dan Kantor Walikota Dumai yang telah memberikan kesempatan untuk melakukan penelitian. Terimakasih juga Keluarga besar Puzzle Research dan Data Technology yang selalu memberikan masukan, dorongan dan semangat dalam penyelesaian penelitian ini.

## REFERENCES

- [1] D. P. Sari, "Implementasi Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 Di Sumatera Barat," *CBIS J.*, vol. 01, pp. 50–56, 2021.
- [2] F. Adiba, "Penerapan Data Mining dalam Mengklasifikasikan Tingkat Kasus Covid-19 di Sulawesi Selatan Menggunakan Algoritma Naive Bayes," *Indones. J. Fundam. Sciencs.*, vol. 7, no. 1, pp. 18–28, 2021.
- [3] A. Prastyowati, "Mengenal Karakteristik Virus SARS-CoV-2 Penyebab Penyakit COVID-19 Sebagai Dasar Upaya Untuk Pengembangan Obat Antivirus Dan Vaksin," *BioTrends*, vol. 11, no. 1, pp. 1–10, 2020.
- [4] 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, doi: 10.21111/fij.v5i2.4905.
- [5] N. Dwitri, J. A. Tampubolon, S. Prayoga, F. Ilmi Zer, and D. Hartama, "Penerapan Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 Di Indonesia," *Jti (Jurnal Teknol. Informasi)*, vol. 4, no. 1, pp. 101–105, 2020.
- [6] B. Produktivitas, "Perbandingan algoritma k-means dengan k-medoids pada pengelompokan armada kendaraan truk berdasarkan produktivitas," vol. 06, pp. 229–240, 2021.
- [7] M. Nana Nurhaliza, "Clustering of Data Covid-19 Cases in the World Using DBSCAN Algorithms Pengelompokan Data Kasus Covid-19 di Dunia Menggunakan Algoritma," vol. 1, no. 1, pp. 1–8, 2021.
- [8] R. Syarif, M. T. Furqon, and S. Adinugroho, "Perbandingan Algoritme K-Means Dengan Algoritme Fuzzy C Means ( FCM ) Dalam Clustering Moda Transportasi Berbasis GPS," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 10, pp. 4107–4115, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2852>.
- [9] D. Caesaron, S. A. Salma, M. D. Prasetyo, and M. H. Rifai, "Edukasi dan Sosialisasi Pencegahan dan Pengendalian COVID-19 melalui Media Poster di Desa Bojongsoang, Kabupaten Bandung," *Abdimas J. Pengabd. Masy. Univ. Merdeka Malang*, vol. 6, no. 2, pp. 221–229, 2021, doi: 10.26905/abdimas.v6i2.5354.
- [10] Kemendagri, "Pendiri Covid-19 dan Implementasinya pada Perekonomian," *Kementerian Perdagang. Ri.*, vol. 14, no. 11, pp. 3497–3508, 2020.
- [11] N. Nurahman, D. A.-J. ILMIAH, and undefined 2021, "Algoritma K-Means Untuk Melihat Penularan Tertinggi Virus Covid-19 Diseluruh Provinsi Indonesia," *Ejournal.Lppmsttpagaralam.Ac.Id.*, vol. 12, no. 02, pp. 162–168, 2021, [Online]. Available: <http://ejournal.lppmsttpagaralam.ac.id/index.php/betrik/article/view/331>.
- [12] S. T. Siska, "VOL . 9 NO . 1 April 2016," *J. Teknol. Inf. Pendidik.*, vol. 9, no. 1, pp. 86–93, 2016.
- [13] N. Rofiqo, A. P. Windarto, and D. Hartama, "Penerapan Clustering Pada Penduduk Yang Mempunyai Keluhan Kesehatan Dengan Datamining K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 216–223, 2018, doi: 10.30865/komik.v2i1.929.
- [14] N. Purba, P. Poningsih, and H. S. Tambunan, "Penerapan Algoritma K-Means Clustering Pada Penyebaran Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Provinsi Riau," *J. Inf. Syst. Res.*, vol. 2, no. 3, pp. 220–226, 2021, [Online]. Available: <http://ejurnal.seminar-id.com/index.php/josh/article/view/736>.
- [15] D. Marlina, N. F. Putri, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," vol. 4, no. 2, pp. 64–71, 2018.
- [16] W. Yustianti, N. Rahmawati, and Y. Yamasari, "Jurnal Wiyli Naim Yuni.Pdf," *Journal Information Engineering and Educational Technology*, vol. 04, no. 01. p. 20, 2021.
- [17] J. Xu, J. Han, X. Kai, and F. Nie, "Robust and sparse fuzzy k-means clustering," *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2016-Janua, pp. 2224–2230, 2016.
- [18] A. E. Pramitasari and Y. Nataliani, "Perbandingan Clustering Karyawan Berdasarkan Nilai Kinerja Dengan Algoritma K-Means Dan Fuzzy C-Means," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 3, pp. 1119–1132, 2021, doi: 10.35957/jatisi.v8i3.957.
- [19] N. I. Selviana *et al.*, "Analisis Perbandingan K-Means dan Fuzzy C-Means untuk Pemetaan Motivasi Belajar Mahasiswa," no. November, pp. 95–105, 2016.
- [20] F. M. Izzadin, "OPTIMASI JUMLAH CLUSTER K-MEANS DENGAN METODE ELBOW DAN SILHOUETTE PADA PRODUKTIVITAS TANAMAN PANGAN DI PROVINSI JAWA TENGAH TAHUN 2018," 2018.
- [21] A. Ramadhan, Mustakim, and R. Handinata, "Implementasi Algoritma Fuzzy C Means Dan Moora Untuk Pengelompokan Dan Penentuan Wilayah Penanggulangan Bencana Banjir," no. November, p. Pekanbaru, 2019.