ANALISIS KLASTER MENGGUNAKAN METODE *FUZZY C-MEANS* PADA DATA COVID-19 DI PROVINSI LAMPUNG

(Skripsi)

Oleh

INDAH SUCIATI

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2021

ABSTRACT

CLUSTER ANALYSIS USING THE FUZZY C-MEANS METHOD ON COVID-19 DATA IN LAMPUNG PROVINCE

By

INDAH SUCIATI

Currently, COVID-19 has spread to various countries, one of which is Indonesia, where Lampung Province is one of the provinces in Indonesia affected by COVID-19. Cluster analysis is a tool for grouping a number of n objects based on p variables that have similar characteristics among these objects, so that the diversity within a cluster is smaller than the diversity between clusters. The purpose of this study is to cluster districts/cities based on positive cases, dead cases, and recovered cases on COVID-19 data in Lampung Province for a period of 6 months (March 2020-August 2020 and September 2020-February 2021) and a period of 12 months (March 2020-February 2021) using the fuzzy c-means method. In this study, the partition coefficient index is used to evaluate the optimal number of clusters, so that the best cluster results will be obtained for COVID-19 data in Lampung Province, then see if there is a shift in cluster results in the first 6 months and the last 6 months. The results obtained in this study are the number of clusters 2 is the optimal number of clusters for clustering each COVID-19 data in Lampung Province with the best cluster results in each data producing the same clustering, where Bandar Lampung City is a district/city in Lampung Province that is included in the high cluster, while for other districts/cities is included in the low cluster. In addition, it was found that there was no shift of cluster results in the COVID-19 data in Lampung Province for the first 6 months (March 2020-August 2020) and the last 6 months (September 2020-February 2021).

Key words: cluster analysis, fuzzy c-means method, partition coefficient index

ABSTRAK

ANALISIS KLASTER MENGGUNAKAN METODE *FUZZY C-MEANS* PADA DATA COVID-19 DI PROVINSI LAMPUNG

Oleh

INDAH SUCIATI

Saat ini, COVID-19 telah tersebar ke berbagai negara salah satunya yaitu Negara Indonesia, dimana Provinsi Lampung merupakan salah satu provinsi di Negara Indonesia yang terdampak COVID-19. Analisis klaster adalah suatu alat untuk mengelompokkan sejumlah n objek berdasarkan p variabel yang mempunyai kesamaan karakteristik diantara objek-objek tersebut, sehingga keragaman dalam suatu klaster tersebut lebih kecil dibandingkan dengan keragaman antar klaster. Tujuan penelitian ini yaitu melakukan pengklasteran kabupaten/kota berdasarkan kasus positif, kasus meninggal, dan kasus sembuh pada data COVID-19 di Provinsi Lampung kurun waktu 6 bulan (Maret 2020-Agustus 2020 dan September 2020-Februari 2021) serta kurun waktu 12 bulan (Maret 2020-Februari 2021) menggunakan metode *fuzzy c-means*. Dalam penelitian ini digunakan *partition coefficient index* untuk mengevaluasi jumlah klaster optimal, sehingga akan diperoleh hasil klaster terbaik untuk data COVID-19 di Provinsi Lampung, kemudian melihat apakah terjadi perpindahan hasil klaster dalam kurun waktu 6 bulan pertama dan 6 bulan terakhir. Hasil yang diperoleh pada penelitian ini yaitu jumlah klaster 2 merupakan jumlah klaster yang optimal untuk pengklasteran setiap data COVID-19 di Provinsi Lampung dengan hasil klaster terbaik pada setiap data menghasilkan klasterisasi yang sama, dimana Kota Bandar Lampung merupakan kabupaten/kota di Provinsi Lampung yang masuk ke dalam klaster tinggi, sedangkan untuk kabupaten/kota lainnya masuk ke dalam klaster rendah. Selain itu, diperoleh hasil bahwa tidak terdapat perpindahan hasil klaster pada data COVID-19 di Provinsi Lampung kurun waktu 6 bulan pertama (Maret 2020-Agustus 2020) dan 6 bulan terakhir (September 2020-Februari 2021).

Kata Kunci: analisis klaster, *fuzzy c-means, partition coefficient index*

ANALISIS KLASTER MENGGUNAKAN METODE *FUZZY C-MEANS* PADA DATA COVID-19 DI PROVINSI LAMPUNG

Oleh

INDAH SUCIATI

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2021

Judul Skripsi	: **ANALISIS KLASTER MENGGUNAKAN METODE FUZZY C-MEANS PADA DATA COVID-19 DI PROVINSI LAMPUNG**
Nama Mahasiswa	: Indah Suciati
Nomor Pokok Mahasiswa	: **1717031078**
Program Studi	: **Matematika**
Fakultas	: **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Ir. Netti Herawati, M.Sc., Ph.D.	**Subian Saidi, S.Si., M.Si.**
NIP. 19650125 199003 2 001	NIP. 19800821 200812 1 001

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

Ketua : **Ir. Netti Herawati, M.Sc., Ph.D.** .....

Sekretaris : **Subian Saidi, S.Si., M.Si.** .....

Penguji
Bukan Pembimbing : **Widiarti, S.Si., M.Si.** .....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T.
NIP. 19740705 200003 1 001

Tanggal Lulus Ujian Skripsi : **6 Agustus 2021**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama	: **Indah Suciati**
Nomor Pokok Mahasiswa	: **1717031078**
Jurusan	: **Matematika**
Judul Skripsi	: **Analisis Klaster Menggunakan Metode *Fuzzy C-Means* pada Data COVID-19 di Provinsi Lampung**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau telah dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai ketentuan akademik yang berlaku.

Bandar Lampung, 18 Agustus 2021
Penulis,

Indah Suciati
NPM. 1717031078

RIWAYAT HIDUP

Penulis bernama lengkap Indah Suciati, lahir di Pringsewu pada tanggal 20 Agustus 2000. Penulis merupakan anak kedelapan dari sembilan bersaudara yang lahir dari pasangan Bapak Kotiman dan Ibu Supiyah.

Penulis memulai pendidikan di SD Negeri 1 Pringsewu Timur pada tahun 2006-2012. Kemudian melanjutkan pendidikan ke sekolah menengah pertama di SMP Negeri 2 Pringsewu pada tahun 2012-2015 dan melanjutkan pendidikan ke sekolah menengah atas di SMA Negeri 1 Gadingrejo pada tahun 2015-2017. Pada tahun 2017, penulis diterima sebagai Mahasiswa S1 di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SBMPTN (Seleksi Bersama Masuk Perguruan Tinggi Negeri).

Selama menjadi mahasiswa, penulis aktif di organisasi dalam kampus antara lain menjadi generasi muda Himpunan Mahasiswa Jurusan Matematika (HIMATIKA) periode 2017 dan Anggota Bidang Minat dan Bakat HIMATIKA periode 2019.

Pada bulan Februari-Maret 2020, selama 40 hari penulis melaksanakan Kuliah Kerja Nyata (KKN) periode pertama di Desa Karang Mulya, Kecamatan Way Serdang, Kabupaten Mesuji, sebagai bentuk pengabdian mahasiswa kepada

masyarakat dan menjalankan Tri Dharma Perguruan Tinggi. Kemudian pada bulan Juli-Agustus tahun 2020, sebagai bentuk penerapan ilmu yang telah dipelajari, penulis melaksanakan Kerja Praktik (KP) di Kantor Badan Pusat Statistik (BPS) Kabupaten Pringsewu.

KATA INSPIRASI

"Sesungguhnya bersama kesulitan ada kemudahan. Maka apabila engkau telah selesai (dari suatu urusan), tetaplah bekerja keras (untuk urusan yang lain). Dan hanya kepada Tuhanmulah engkau berharap."
(Q.S. Al-Insyirah : 6-8)

"Allah tidak akan membebani seseorang melainkan sesuai dengan kesanggupannya."
(Q.S. Al-Baqarah : 286)

"Barangsiapa menempuh jalan untuk mendapatkan ilmu, Allah akan memudahkan baginya jalan menuju surga."
(H.R. Muslim)

"Janganlah pernah menyerah ketika Anda masih mampu berusaha lagi. Tidak ada kata berakhir sampai Anda berhenti mencoba"
(Brian Dyson)

"Jangan menyerah, hal-hal besar membutuhkan waktu, bersabarlah"
(Anonim)

PERSEMBAHAN

Dengan mengharap rahmat dan keridhaan Allah SWT, kupersembahkan karya sederhana ini kepada:

Bapak, Ibu, Kakak, Adik, dan Keponakan Tercinta

Yang telah mencurahkan semua pengorbanan, doa, kasih sayang, dan dukungan tiada henti. Terima kasih sudah menjadi rumah dan tempat berpulang dari segala keluh kesah atas hal-hal pelik yang datang.

Dosen Pembimbing dan Penguji

Yang senantiasa telah meluangkan waktu untuk memberikan bimbingan dan ilmu yang berharga kepada penulis.

Sahabat-Sahabatku

Yang selalu memberikan semangat, dukungan dan doa, serta selalu membantu penulis selama masa perkuliahan.

Almamater Tercinta Universitas Lampung

SANWACANA

Alhamdulillah puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah serta karunia-Nya kepada penulis sehingga skripsi dengan judul "**Analisis Klaster Menggunakan Metode *Fuzzy C-Means* pada Data COVID-19 di Provinsi Lampung**" dapat terselesaikan dengan baik.

Terselesaikan skripsi ini tidak lepas dari dukungan, bimbingan, saran, serta doa dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis mengucapkan terimakasih kepada:

1. Ibu Ir. Netti Herawati, M.Sc., Ph.D. selaku Dosen Pembimbing I yang telah membimbing penulis, menyumbangkan ilmu, memberikan motivasi dan arahan, serta kesediaan waktu yang diberikan.
2. Bapak Subian Saidi, S.Si., M.Si., selaku Dosen Pembimbing II dan Dosen Pembimbing Akademik yang telah membimbing penulis, memberikan motivasi dan arahan, serta kesediaan waktu yang diberikan baik dalam perkuliahan maupun penelitian.
3. Ibu Widiarti, S.Si., M.Si., selaku Dosen Penguji atas kesediaannya untuk menguji, memberikan saran dan kritik yang membangun dalam penyelesaian skripsi ini.
4. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
5. Bapak Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Seluruh dosen, staf, dan karyawan Jurusan Matematika yang telah memberikan ilmu dan bantuan yang berguna bagi penulis.

7. Bapak, Ibu, kakak-kakakku yaitu Mbak Sari, Mas Anton, Mbak Nourma, Mbak Lis, Mbak Wulan, dan Mbak Anggun, adikku yaitu Jaya, dan keponakan-keponakanku yaitu Devon, Cleo, Sachio, dan Elvano, serta seluruh keluarga yang selalu menjadi semangat tersendiri bagi penulis dan tiada henti memberi dukungan, kasih sayang, motivasi, serta doa kepada penulis.
8. Sahabat terbaik penulis, Dhea, Epmi, Yustika, Nita, Dewi, Inas, Atina, Ayu, Astri, Hana yang siap mendengarkan keluh kesah, memberi semangat dan keceriaan, serta selalu membantu penulis dalam hal apapun.
9. Anggota Bidang Minat dan Bakat HIMATIKA periode 2019, serta Keluarga Besar HIMATIKA periode 2019 yang telah memberikan banyak pembelajaran dan pengalaman kepada penulis.
10. Teman-teman "Lambetika C" atas kebersamaan dan canda tawanya selama masa perkuliahan.
11. Teman-teman seperbimbingan sebagai tempat saling bertukar pikiran, memberikan saran, dan saling menyemangati selama proses penyelesaian skripsi.
12. Teman-teman Jurusan Matematika angkatan 2017 atas kebersamaannya.
13. Semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu penulis dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih terdapat kekurangan baik isi maupun susunannya. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun agar dapat menjadi bahan perbaikan untuk kedepannya.

Bandar Lampung, 18 Agustus 2021
Penulis,

Indah Suciati

DAFTAR ISI

Halaman

DAFTAR TABEL vi

DAFTAR GAMBAR viii

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah 1
1.2 Tujuan 3
1.3 Manfaat 3

II. TINJAUAN PUSTAKA

2.1 Analisis Klaster 4
2.2 Logika *Fuzzy* 5
2.3 Jarak Euclidean 5
2.4 Metode *Fuzzy C-Means* (FCM) 6
2.5 Algoritma Metode *Fuzzy C-Means* (FCM) 7
2.6 Standarisasi Data 9
2.7 Indeks Validitas 10
2.8 *Partition Coefficient Index* (PCI) 10

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian 12
3.2 Data Penelitian 12
3.3 Metode Penelitian 13

IV. HASIL DAN PEMBAHASAN

4.1 Analisis Klaster Menggunakan Metode *Fuzzy C-Means* (FCM) 15
4.2 Pemilihan Jumlah Klaster Optimal Menggunakan *Partition Coefficient Index* (PCI) 31
4.3 Hasil Klaster Terbaik Menggunakan Metode *Fuzzy C-Means* (FCM) 32

V. KESIMPULAN

DAFTAR PUSTAKA.....36

LAMPIRAN

DAFTAR TABEL

Tabel Halaman

1. Standarisasi Data COVID-19 di Provinsi Lampung Kurun Waktu 6 Bulan (Maret 2020-Agustus 2020) 16

2. Standarisasi Data COVID-19 di Provinsi Lampung Kurun Waktu 6 Bulan (September 2020-Februari 2021) 16

3. Standarisasi Data COVID-19 di Provinsi Lampung Kurun Waktu 12 Bulan (Maret 2020-Februari 2021)....17

4. Membangun Matriks Partisi Awal untuk Jumlah Klaster 3....19

5. Pusat Klaster Akhir Data COVID-19 di Provinsi Lampung Kurun Waktu 6 Bulan (Maret 2020-Agustus 2020)....20

6. Pusat Klaster Akhir Data COVID-19 di Provinsi Lampung Kurun Waktu 6 Bulan (September 2020-Februari 2021)....22

7. Pusat Klaster Akhir Data COVID-19 di Provinsi Lampung Kurun Waktu 12 Bulan (Maret 2020-Februari 2021) 23

8. Nilai Akhir Perubahan Fungsi Objektif Data COVID-19 di Provinsi Lampung 25

9. Perubahan Akhir Nilai Keanggotaan Data COVID-19 di Provinsi Lampung Kurun Waktu 6 Bulan (Maret 2020-Agustus 2020) untuk Jumlah Klaster 2 dengan Pangkat Pembobot 2....26

10. Perubahan Akhir Nilai Keanggotaan Data COVID-19 di Provinsi Lampung Kurun Waktu 6 Bulan (September 2020-Februari 2021) untuk Jumlah Klaster 2 dengan Pangkat Pembobot 2....27

11. Perubahan Akhir Nilai Keanggotaan Data COVID-19 di Provinsi Lampung Kurun Waktu 12 Bulan (Maret 2020-Februari 2021) untuk Jumlah Klaster 2 dengan Pangkat Pembobot 2....27

12. Perubahan Fungsi Objektif Data COVID-19 di Provinsi Lampung Kurun Waktu 6 Bulan (Maret 2020-Agustus 2020) dengan Jumlah Klaster 2 dan Pangkat Pembobot 2.....28

13. Klasterisasi Data COVID-19 di Provinsi Lampung Kurun Waktu 6 Bulan (Maret 2020-Agustus 2020) untuk Jumlah Klaster 2 dengan Pangkat Pembobot 2 29

14. Klasterisasi Data COVID-19 di Provinsi Lampung Kurun Waktu 6 Bulan (September 2020-Februari 2021) untuk Jumlah Klaster 2 dengan Pangkat Pembobot 2 30

15. Klasterisasi Data COVID-19 di Provinsi Lampung Kurun Waktu 12 Bulan (Maret 2020-Februari 2021) untuk Jumlah Klaster 2 dengan Pangkat Pembobot 2 30

16. Perbandingan Nilai PCI Data COVID-19 di Provinsi Lampung 31

17. Hasil Klaster Terbaik Data COVID-19 di Provinsi Lampung Berdasarkan Kurun Waktu.....32

18. Hasil Klaster Terbaik Data COVID-19 di Provinsi Lampung.....33

DAFTAR GAMBAR

Gambar Halaman

1. Peta Sebaran Data COVID-19 di Provinsi Lampung Kurun Waktu 6 Bulan (Maret 2020-Agustus 2020 dan September 2020-Februari 2021) serta Kurun Waktu 12 Bulan (Maret 2020-Februari 2021).....34

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Coronavirus Disease (COVID-19) adalah penyakit menular yang ditandai dengan gejala pada bagian pernapasan akut *(SARS-CoV-2*). COVID-19 pertama kali teridentifikasi di Wuhan, Provinsi Hubei, Cina dengan kasus pertama yang dilaporkan oleh WHO *China Country Office* merupakan kasus yang saat itu belum diketahui secara pasti etiologinya. Pada tanggal 11 Maret 2020, WHO menetapkan COVID-19 sebagai pandemi (Dwitri, dkk., 2020). Saat ini, COVID-19 telah tersebar ke berbagai negara salah satunya yaitu Negara Indonesia, dimana Provinsi Lampung merupakan salah satu provinsi di Negara Indonesia yang terdampak COVID-19.

Penyebaran COVID-19 yang merata di kabupaten/kota Provinsi Lampung merupakan penyebaran yang cukup cepat dan berdampak negatif. Luasnya wilayah Provinsi Lampung memungkinkan diperlukannya pengklasteran kabupaten/kota di Provinsi Lampung. Pengklasteran ini akan menghasilkan titik-titik pusat penyebaran kasus COVID-19 yang nantinya dapat dievaluasi dan dijadikan suatu informasi. Menurut Muslimatin (2011), salah satu cabang ilmu statistika yang membahas tentang metode pengklasteran adalah analisis klaster.

Analisis klaster adalah suatu alat untuk mengelompokkan sejumlah n objek berdasarkan p variabel yang mempunyai kesamaan karakteristik diantara objek-objek tersebut, sehingga keragaman dalam suatu klaster tersebut lebih kecil dibandingkan dengan keragaman antar klaster (Sitepu, Irmeilyana, & Gultom, 2011). Analisis klaster dapat diklasifikasikan sebagai *soft clustering* dan *hard clustering* (Bora & Gupta, 2014). Metode *soft clustering* yang sangat populer adalah metode *Fuzzy C-Means* (FCM), dimana kelebihan metode ini terletak pada penempatan pusat klaster yang lebih tepat dibandingkan dengan metode klaster lainnya (Megawati, Mukid, & Rahmawati, 2013). Namun, metode ini terdapat kelemahan dalam penentuan jumlah klaster yang optimal. Untuk meminimalisir hal tersebut serta mengoptimalkan hasil pengklasteran, perlu dilakukan validasi dengan menggunakan indeks validitas (Sutoyo & Sumpala, 2015). Indeks validitas yang sering digunakan dalam metode FCM yaitu *Partition Coefficient Index* (PCI), dimana indeks validitas ini hanya mengevaluasi nilai derajat keanggotaan, tanpa memandang nilai vektor (data) yang biasanya mengandung informasi sebaran data (Wu & Yang, 2005).

Berdasarkan uraian di atas, penulis tertarik untuk melakukan penelitian mengenai analisis klaster menggunakan metode *Fuzzy C-Means* (FCM) pada data COVID-19 di Provinsi Lampung, penulis menggunakan *Partition Coefficient Index* (PCI) untuk mengevaluasi jumlah klaster optimal, sehingga akan diperoleh hasil klaster terbaik untuk data COVID-19 di Provinsi Lampung.

1.2 Tujuan

Tujuan dari penelitan ini adalah sebagai berikut:

1. Melakukan pengklasteran kabupaten/kota berdasarkan kasus positif, kasus meninggal, dan kasus sembuh COVID-19 di Provinsi Lampung menggunakan metode *Fuzzy C-Means* (FCM).
2. Menentukan jumlah klaster yang optimal untuk pengklasteran kabupaten/kota berdasarkan kasus positif, kasus meninggal, dan kasus sembuh COVID-19 di Provinsi Lampung dengan *Partition Coefficient Index* (PCI).
3. Memperoleh hasil klaster terbaik untuk pengklasteran kabupaten/kota berdasarkan kasus positif, kasus meninggal, dan kasus sembuh COVID-19 di Provinsi Lampung.
4. Melihat ada tidaknya perpindahan hasil klaster pada data COVID-19 di Provinsi Lampung untuk kurun waktu 6 bulan pertama (Maret 2020-Agustus 2020) dan 6 bulan terakhir (September 2020-Februari 2021).

1.3 Manfaat

Manfaat dari penelitian ini adalah sebagai berikut:

1. Dapat mengetahui hasil klaster terbaik untuk pengklasteran kabupaten/kota berdasarkan kasus positif, kasus meninggal, dan kasus sembuh COVID-19 di Provinsi Lampung.
2. Dapat menjadi referensi bagi peneliti apabila ingin melakukan penelitian mengenai analisis klaster menggunakan metode FCM.

II. TINJAUAN PUSTAKA

2.1 Analisis Klaster

Menurut Sitepu, Irmeilyana, & Gultom (2011), analisis klaster adalah suatu alat untuk mengelompokkan sejumlah n objek berdasarkan p variabel yang mempunyai kesamaan karakteristik diantara objek-objek tersebut, sehingga keragaman dalam suatu klaster tersebut lebih kecil dibandingkan dengan keragaman antar klaster. Menurut Talakua, Leleury, & Talluta (2017), suatu klaster memiliki ciri-ciri kesamaan yang tinggi antara anggota dalam satu klaster (*within cluster*) dan perbedaan yang tinggi antara klaster yang satu dengan klaster lain (*between cluster*).

Menurut Bora & Gupta (2014), analisis klaster dapat diklasifikasikan sebagai *soft clustering* dan *hard clustering*. Dalam *soft clustering*, himpunan *fuzzy* digunakan untuk mengelompokkan data sehingga setiap data dapat dimiliki oleh dua atau lebih klaster dengan derajat keanggotaan yang berbeda. Sedangkan pada teknik *hard clustering*, himpunan *crisp* digunakan untuk mengelompokkan data sehingga jika suatu data masuk ke dalam klaster tertentu maka tidak dapat masuk ke klaster lain. Dalam banyak situasi, *soft clustering* lebih alami daripada *hard clustering*. Teknik *soft clustering* yang sangat populer adalah *Fuzzy C-Means* (FCM).

2.2 Logika *Fuzzy*

Logika *fuzzy* pertama kali diperkenalkan oleh Lotfi A. Zadeh pada tahun 1965. Dasar logika *fuzzy* adalah teori himpunan *fuzzy*. Pada teori himpunan *fuzzy*, peranan nilai keanggotan sebagai penentu keberadaan elemen dalam suatu himpunan sangatlah penting. Pada himpunan tegas (*crisp*), nilai keanggotaan hanya terdapat dua kemungkinan yaitu 0 dan 1, sedangkan pada himpunan *fuzzy* nilai keanggotaan terletak pada rentang 0 sampai 1. Apabila *x* memiliki nilai keanggotaan *fuzzy* $\mu_A[x] = 0$, berarti x tidak menjadi anggota himpunan A, apabila x memiliki nilai keanggotaan *fuzzy* $\mu_A[x] = 1$, berarti x menjadi anggota penuh pada himpunan A, demikian pula apabila x memiliki nilai keanggotaan *fuzzy* $\mu_A[x]$ antara $0 - 1$, berarti x menjadi anggota pada himpunan A secara parsial (Kusumadewi & Purnomo, 2010).

2.3 Jarak Euclidean

Dalam pengklasteran digunakan suatu ukuran yang dapat menerangkan keserupaan atau kedekatan antar data, yaitu ukuran jarak atau similaritas. Ukuran jarak yang sering digunakan adalah ukuran jarak Euclidean, dimana jarak Euclidean digunakan untuk melihat kedekatan data terhadap pusat klaster (Johnson & Wichern, 1982). Menurut Fathia, Rahmawati, & Tarno (2016), rumus jarak Euclidean adalah sebagai berikut:

$$d_{ik} = \sqrt{\sum_{j=1}^{p} \left(x_{ij} - c_{kj}\right)^2} \qquad (2.1)$$

dengan:

d_{ik} = jarak objek ke-i dengan pusat klaster ke-k

p = jumlah variabel

x_{ij} = data dari objek ke-i pada variabel ke-j

c_{kj} = pusat klaster ke-k pada variabel ke-j

2.4 Metode *Fuzzy C-Means* (FCM)

Metode *Fuzzy C-Means* (FCM) adalah teknik pengelompokkan data yang keberadaan tiap-tiap titik data suatu klaster ditentukan oleh nilai keanggotaan. Nilai keanggotaan tersebut mencakup bilangan real pada interval [0-1]. Metode ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981 dengan kelebihan metode ini yaitu penempatan pusat klaster yang lebih tepat jika dibandingkan dengan metode klaster lainnya (Megawati, Mukid, & Rahmawati, 2013).

Konsep dasar metode FCM yang pertama adalah menentukan pusat klaster yang akan menandai lokasi rata-rata untuk tiap-tiap klaster. Pada kondisi awal, pusat klaster ini masih belum akurat. Tiap-tiap data memiliki nilai keanggotaan untuk tiap-tiap klaster. Dengan cara memperbaiki pusat klaster dan nilai keanggotaan tiap-tiap data secara berulang, maka akan terlihat bahwa pusat klaster akan bergerak menuju lokasi yang tepat. Perulangan ini didasarkan pada minimasi fungsi objektif (Muhardi & Nisar, 2015).

2.5 Algoritma Metode *Fuzzy C-Means* (FCM)

Menurut Kusumadewi (2010), algoritma dari *Fuzzy C-Means* (FCM) adalah sebagai berikut:

1. Memasukkan data yang akan diklaster, berupa matriks $\boldsymbol{X}$ berukuran $n \times p$ ($n =$ jumlah sampel data, $p =$ jumlah variabel setiap data) sebagai berikut:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \tag{2.2}$$

   dengan:

   $i = 1, 2, \ldots, n$

   $j = 1, 2, \ldots, p$

   $x_{ij} =$ data sampel ke-i variabel ke-j

2. Menentukan:
   a. Jumlah klaster ($1 < c < N$)
   b. Pangkat pembobot ($m > 1$)
   c. Maksimum iterasi ($MaxIter$)
   d. Galat terkecil yang diharapkan ($\varepsilon > 0$)
   e. Fungsi objektif awal ($P_0 = 0$)
   f. Iterasi awal ($t = 1$)
3. Membangkitkan bilangan acak μ_{ik} sebagai elemen-elemen matriks partisi awal $\boldsymbol{U}$. Matriks partisi awal $\boldsymbol{U}$ sebagai berikut:

$$\boldsymbol{U} = \begin{bmatrix} \mu_{11}(x_1) & \mu_{12}(x_1) & \cdots & \mu_{1c}(x_1) \\ \mu_{21}(x_2) & \mu_{22}(x_2) & \cdots & \mu_{2c}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{n1}(x_n) & \mu_{n2}(x_n) & \cdots & \mu_{nc}(x_n) \end{bmatrix} \tag{2.3}$$

dengan:

$i = 1, 2, \dots, n$

$k = 1, 2, \dots, c$

μ_{ik} = elemen matriks partisi awal $\boldsymbol{U}$ ke-i klaster ke-k

dengan aturan:

$$\sum_{k=1}^{c} \mu_{ik} = 1 \tag{2.4}$$

4. Menghitung pusat klaster ke-k pada variabel ke-j (v_{kj}) dengan rumus sebagai berikut:

$$v_{kj} = \frac{\sum_{i=1}^{n}((\mu_{ik})^m x_{ij})}{\sum_{i=1}^{n}(\mu_{ik})^m} \tag{2.5}$$

dengan:

$j = 1, 2, \dots, p$

$k = 1, 2, \dots, c$

m = pangkat pembobot

μ_{ik} = elemen matriks partisi awal $\boldsymbol{U}$ ke-i klaster ke-k

sehingga diperoleh matriks pusat klaster $\boldsymbol{V}$ sebagai berikut:

$$\boldsymbol{V} = \begin{bmatrix} v_{11} & \cdots & v_{1p} \\ \vdots & \ddots & \vdots \\ v_{c1} & \cdots & v_{cp} \end{bmatrix} \tag{2.6}$$

5. Menghitung nilai fungsi objektif pada iterasi ke-t (P_t) dengan rumus sebagai berikut:

$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} ((\mu_{ik})^m \, d_{ik}) \tag{2.7}$$

dengan:

$i = 1, 2, \dots, n$

$k = 1, 2, \dots, c$

d_{ik} = jarak Euclidean dari objek ke-i dengan pusat klaster ke-k

6. Menghitung perubahan matriks keanggotaan μ_{ik} dengan rumus sebagai berikut:

$$\mu_{ik} = \left[\frac{\left[\sum_{j=1}^{p} d_{ik}\right]^{\frac{1}{m-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{p} d_{ik}\right]^{\frac{1}{m-1}}} \right]^{-1} \quad (2.8)$$

dengan:

$i = 1, 2, \dots, n$

$j = 1, 2, \dots, p$

$k = 1, 2, \dots, c$

$m =$ pangkat pembobot

$\mu_{ik} =$ elemen matriks partisi awal $\boldsymbol{U}$ ke-i klaster ke-k

$d_{ik} =$ jarak Euclidean dari objek ke-i dengan pusat klaster ke-k

7. Memeriksa kondisi berhenti:
   a. Jika ($|P_t - P_{t-1}| < \varepsilon$) atau ($t > MaxIter$) maka iterasi berhenti.
   b. Jika tidak, maka $t = t + 1$ dan kembali ke langkah 4.

2.6 Standarisasi Data

Variabel pengklasteran harus distandarisasi apabila memungkinkan untuk menghindari masalah yang dihasilkan dari penggunaan nilai skala yang berbeda antar variabel pengklasteran. Standarisasi yang paling umum adalah konversi setiap variabel terhadap nilai standar (dikenal dengan *z-score*) dengan melakukan substraksi nilai tengah dan membaginya dengan standar deviasi tiap variabel.

Menurut Walpole & Myers (1995), rumus standarisasi data sebagai berikut:

$$z = \frac{x_i - \bar{x}}{s} \tag{2.9}$$

dengan:

$z =$ standarisasi variabel

$x_i =$ data ke- i

$\bar{x} =$ rata-rata keseluruhan data setiap variabel

$s =$ standar deviasi

2.7 Indeks Validitas

Menurut Sutoyo & Sumpala (2015), metode *Fuzzy C-Means* (FCM) sering digunakan karena memiliki tingkat keakuratan yang tinggi dan waktu komputasi yang cepat. Namun, metode ini terdapat kelemahan dalam penentuan jumlah klaster yang optimal. Untuk meminimalisir hal tersebut serta mengoptimalkan hasil pengklasteran, perlu dilakukan validasi jumlah klaster. Validasi jumlah klaster dilakukan menggunakan indeks validitas. Indeks validitas adalah ukuran validitas untuk memperoleh jumlah klaster optimal yang sepenuhnya dapat menjelaskan struktur data (Zhao & Franti, 2014).

2.8 *Partition Coefficient Index* (PCI)

Menurut Wu & Yang (2005), *Partition Coefficient Index* (PCI) adalah indeks validitas yang hanya mengevaluasi nilai derajat keanggotaan, tanpa memandang nilai vektor (data) yang biasanya mengandung informasi sebaran data.

Persamaan yang digunakan untuk menghitung *Partition Coefficient Index* (PCI) yaitu sebagai berikut:

$$PCI = \frac{1}{n}\sum_{k=1}^{c}\sum_{i=1}^{n}(\mu_{ik})^2 \quad (2.10)$$

dimana n adalah banyak objek penelitian, c adalah banyak klaster, dan μ_{ik} adalah nilai keanggotaan objek ke-i dengan pusat klaster ke-k. PCI memiliki rentang nilai antara 0 sampai 1 dengan jumlah klaster yang optimal ditunjukkan oleh nilai PCI yang paling besar.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada semester ganjil tahun ajaran 2021/2022 bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data COVID-19 di Provinsi Lampung kurun waktu 6 bulan yaitu Maret 2020-Agustus 2020 dan September 2020-Februari 2021, serta kurun waktu 12 bulan yaitu Maret 2020-Februari 2021 yang diperoleh dari situs www.bappeda.lampungprov.go.id dengan data bersumber dari Dinas Kesehatan Provinsi Lampung dan variabel yang digunakan dalam penelitian ini yaitu kasus positif, kasus meninggal, dan kasus sembuh COVID-19 di Provinsi Lampung.

3.3 Metode Penelitian

Penelitian ini dilakukan dengan langkah-langkah sebagai berikut:

1. Melakukan pengambilan data penelitian.
2. Melakukan analisis klaster untuk data penelitian dengan metode *Fuzzy C-Means* (FCM) menggunakan *software* RStudio 1.4.1103 dengan langkah-langkah sebagai berikut:
   a. Memasukkan data yang akan digunakan dalam pengklasteran.
   b. Membangun matriks yang merupakan nilai standarisasi dari data penelitian.
   c. Menentukan:
      1) Jumlah klaster
      2) Pangkat pembobot
      3) Maksimum iterasi sampai mencapai nilai konvergen
      4) Galat terkecil
      5) Fungsi objektif awal
      6) Iterasi awal
   d. Membentuk matriks partisi awal.
   e. Meghitung pusat klaster.
   f. Menghitung nilai fungsi objektif.
   g. Memperbaiki derajat keanggotan setiap klaster.
   h. Memeriksa kondisi berhenti.
   i. Melakukan pengklasteran untuk setiap objek berdasarkan nilai keanggotaan.
3. Menghitung serta membandingkan nilai *Partition Coefficient Index* (PCI) pada hasil klaster data COVID-19 di Provinsi Lampung.

4. Memperoleh jumlah klaster optimal dan hasil klaster terbaik untuk data COVID-19 di Provinsi Lampung.

V. KESIMPULAN

Berdasarkan hasil dan pembahasan pada bab sebelumnya, diperoleh kesimpulan bahwa jumlah klaster optimal untuk pengklasteran kabupaten/kota berdasarkan kasus positif, kasus meninggal, dan kasus sembuh COVID-19 di Provinsi Lampung menggunakan metode FCM adalah 2. Hal ini dikarenakan nilai PCI pada jumlah klaster 2 merupakan nilai PCI yang paling besar jika dibandingkan dengan jumlah klaster lainnya.

Berdasarkan jumlah klaster optimal sebelumnya, maka diperoleh hasil klaster terbaik untuk pengklasteran kabupaten/kota berdasarkan kasus positif, kasus meninggal, dan kasus sembuh COVID-19 di Provinsi Lampung untuk kurun waktu 6 bulan (Maret 2020-Agustus 2020 dan September 2020-Februari 2021) serta kurun waktu 12 bulan (Maret 2020-Februari 2021) dengan menggunakan metode FCM, yaitu untuk klasterisasi data penelitian menghasilkan hasil klaster yang sama, dimana Kota Bandar Lampung merupakan kabupaten/kota di Provinsi Lampung yang masuk ke dalam klaster tinggi, sedangkan untuk kabupaten/kota lainnya masuk ke dalam klaster rendah. Selain itu, diketahui bahwa tidak terdapat perpindahan hasil klaster pada data COVID-19 di Provinsi Lampung untuk kurun waktu 6 bulan pertama (Maret 2020-Agustus 2020) dan 6 bulan terakhir (September 2020-Februari 2021).

DAFTAR PUSTAKA

Bora, D.J. & Gupta, A.K. 2014. A Comparative Study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm. *International Journal of Computer Trends and Technology (IJCTT)*. **10**(2): 108-113.

Dwitri, N., dkk. 2020. Penerapan Algoritma K-Means dalam Menentukan Tingkat Penyebaran Pandemi COVID-19 di Indonesia. *Jurnal Teknologi Informasi*. **4**(1): 128-132.

Fathia, A.N., Rahmawati, R., & Tarno. 2016. Analisis Klaster Kecamatan di Kabupaten Semarang Berdasarkan Potensi Desa Menggunakan Metode Ward dan Single Linkage. *Jurnal Gaussian*. **5**(4): 801-810.

Johnson, R.A. & Wincern, D.W. 1982. *Applied Multivariate Statistical Analysis*. Prentice Hal, Inc., New Jersey.

Kusumadewi, S. 2010. *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Graha Ilmu, Yogyakarta.

Kusumadewi, S. & Purnomo, H. 2010. *Artificial Intelligence*: *Teknik dan Aplikasinya*. Graha Ilmu, Yogyakarta.

Megawati, N., Mukid, M.A., & Rahmawati, R. 2013. Segmentasi Pasar pada Pusat Perbelanjaan Menggunakan Fuzzy C-Means (Studi Kasus: Rita Pasaraya Cilacap). *Jurnal Gaussian*. **2**(4): 343-350.

Muhardi & Nisar. 2015. Penentuan Penerima Beasiswa dengan Algoritma Fuzzy C-Means di Universitas Megow Pak Tulang Bawang. *Jurnal Tim Darmajaya*. **1**(2): 158-174.

Muslimatin, B. 2011. Perbandingan Metode K-Means dan Metode Fuzzy C-Means (FCM) untuk Clustering Data (Studi Kasus pada Data Saham Harian PT. Astra, Tbk.). Skripsi. Jurusan Matematika Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim, Malang.

Sitepu, R., Irmeilyana, & Gultom, B. 2011. Analisis Cluster terhadap Tingkat Pencemaran Udara pada Sektor Industri di Sumatera Selatan. *Jurnal Penelitian Sains*. **14**(3): 11-17.

Sutoyo, M.S. & Sumpala, A.T. 2015. Penerapan Fuzzy C-Means untuk Deteksi Dini Kemampuan Penalaran Matematis. *Scientific Journal of Informatics*. **2**(15): 129-136.

Talakua, M.W., Leleury, Z.A., & Talluta, A.W. 2017. Analisis Cluster dengan Menggunakan Metode K-Means untuk Pengelompokkan Kabupaten/Kota di Provinsi Maluku berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014. *Barekeng.* **11**(2): 119-128.

Walpole, R.E. & Myers, R.H. 1995. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuawan*. Ed. ke-4. ITB, Bandung.

Wu, K.L. & Yang, M.S. 2005. A Cluster Validity Index for Crisp Clusters. *Pattern Analysis and Applications.* **26**(9): 1275-1291.

Zhao, Q. & Franti, P. 2014. WB-index: A Sum-of-Squares Based Index for Cluster Validity. *Elsevier*. **92**: 77–89.