Jurnal Sainsmat, September 2023, Halaman 141 - 152 Vol. XII, No. 2
ISSN 2579-5686 (Online) ISSN 2086-6755
http://ojs.unm.ac.id/index.php/sainsmat

Analisis *Cluster Fuzzy C-Means* dan Diskriminan untuk Pengelompokan Data Kesejahteraan Rakyat

Cluster Fuzzy C-Means Analysis and Discrimination for Collection of People's Welfare Data

Moh. Wahyu Warolemba[1)], Resmawan[2)], Dewi Rahmawaty Isa[1)]*

[1)] Program Studi Statistika, Universitas Negeri Gorontalo
[2)] Program Studi Matematika, Universitas Negeri Gorontalo

ABSTRAK

Analisis cluster adalah metode multivariat yang tujuannya membentuk pengelompokkan berdasarkan perbedaan karakteristik. Teknik statistik yang dapat dilakukan untuk pengelompokkan adalah Fuzzy c-means (FCM). Cara kerja FCM yaitu menggunakan derajat keanggotaan untuk menentukkan keberadaan tiap titik data dalam suatu kelompok (Cluster). Untuk memastikan pengelompokan akurat dan kriteria terpenuhi, validasi cluster perlu dilakukan validasi cluster sehingga menghasilkan pengelompokkan data yang baik. Penelitian ini menggunakan analisis cluster Fuzzy c-means dan validasi diskrimanan dengan tujuan untuk mengelompokkan 34 provinsi di Indonesia berdasarkan tingkat kesejahteraan rakyat. Data yang digunakan pada penelitian ini yaitu data indikator kesejahteraan rakyat di Indonesia tahun 2021. Variabel-variabel yang digunakan pada penelitian ini adalah umur/angka harapan hidup, rata-rata lama sekolah, angka partisipasi sekolah 16-18 tahun, dan rumah tangga dengan akses air yang layak. Berdasarkan hasil penelitian terbentuk dua cluster yaitu cluster 1 daerah dengan nilai indikator kesejahteraan rakyat rendah sebanyak 16 provinsi dan cluster 2 daerah dengan nilai indikator kesejahteraan rakyat tinggi sebanyak 18 provinsi.

Kata kunci: Fuzzy c means, Validasi diskriminan, Kesejahteraan rakyat.

ABSTRACT

Cluster analysis is a multivariate method that aims to classify objects into a cluster based on different characteristics. The statistical technique that can be used for clustering is Fuzzy C-means (FCM). The FCM uses membership degree to determine the existence of each data point in a cluster. To ensure accurate clustering and that the criteria are met, cluster validation must be done to produce good data clustering. Moreover, this study uses a discriminant analysis test to validate the results of cluster solutions. The study aims to classify 34 provinces in Indonesia based on the level of people's welfare. The data used in this study is data on indicators of people's welfare in Indonesia in 2021. The variables used in this study were age/life expectancy, the average length of schooling, the 16-18 years of enrollment, and households with proper access to water. Based on the study results, two

* *Korespondensi:*
email: mr.akhtar27@yahoo.co.id

clusters were formed: cluster 1, areas with low people's welfare indicator values in 16 provinces, and Cluster 2, regions with high people's welfare indicator values in 18 provinces.

Keywords: fuzzy c means, discriminant validation, people's welfare.

PENDAHULUAN

Kesejahteraan rakyat merupakan suatu kondisi yang bentuknya dinamis atau nilai kuantitatifnya tidak akan pernah berhenti karena terus berubah seiring dengan perkembangan kebutuhan hidup manusia (Alwi & Hasrul, 2018). Perkembangan kesejahteraan rakyat merupakan salah satu cita-cita luhur negara Indonesia sebagaimana disebutkan dalam Undang-Undang Dasar 1945, namun permasalahan kesejahteraan masih banyak dijumpai di berbagai provinsi di Indonesia. Faktor-faktor yang menyebabkan kesejahteraan rakyat yaitu kepadatan penduduk dan kemiskinan. Selain itu, Badan Pusat Statistik menuliskan beberapa indikator yang dapat mempengaruhi tingkat kesejahteraan rakyat yaitu kesehatan dan gizi, pendidikan, ketenagakerjaan, taraf dan pola konsumsi, kemiskinan, perumahan dan lingkungan.

Masalah kesejahteraan di Indonesia perlu ditangani khususnya di provinsi-provinsi yang memiliki tingkat kesejahteraan yang rendah. Oleh karena itu, untuk menangani masalah tersebut dilakukan analisis pengelompokan untuk memberikan gambaran mengenai kesejahteraan rakyat di setiap provinsi.

Berdasarkan permasalahan analisis yang dapat digunakan yaitu analisis *cluster*. Analisis *cluster* merupakan metode multivariat yang bertujuan untuk membentuk pengelompokkan berdasarkan karakteristik yang berbeda. Salah satu metode *cluster* yang dapat digunakan untuk pengelompokkan yaitu *cluster Fuzzy c-means* (FCM). FCM merupakan suatu teknik pengelompokkan data dimana keberadaan tiap titik data dalam suatu kelompok (*cluster*) ditentukan oleh derajat keanggotaan. *Cluster* yang baik dan tepat didapatkan dengan cara memperbaiki pusat *cluster* dan nilai keanggotaan tiap data secara berkala (Ningrat dkk, 2016). *Cluster* baik dapat dilihat dari nilai homogenitas tinggi yang dimiliki antar anggota dalam suatu *cluster* dan memiliki heterogenitas yang tinggi antara satu *cluster* dengan lainnya (Dani et al., 2021). Untuk mendapatkan metode pengelompokan yang tepat dan memenuhi kriteria pengelompokan yang baik dibutuhkan proses validasi *cluster* atau validasi pengelompokan, sehingga pada pengelompokan data akan didapatkan hasil yang lebih baik. Dalam penelitian ini digunakan uji analisis diskriminan untuk validasi hasil solusi cluster. Analisis diskriminan merupakan salah satu analisis multivariat yang termasuk *Dependence Method* yaitu terdapat variabel independen dan dependen, sehingga ada variabel yang nilainya tergantung dari data variabel independen. Pada analisis diskriminan juga ditentukan nilai *Hit Ratio*, *Cmax* dan *Press'Q* yang berfungsi sebagai indikator bahwa pengelompokan yang dilakukan sudah optimal atau belum.

Beberapa penelitian terkait analisis *cluster* dengan analisis diskriminan banyak dilakukan dari waktu kewaktu. Pada Penelitian yang dilakukan oleh Wijayanti, dkk (2021) menggunakan analisis *Fuzzy c-means* untuk mengelompokkan provinsi di Indonesia berdasarkan pada indikator kesehatan lingkungan. Penelitian Maghfiroh, dkk (2019)

menggunakan *Fuzzy c-means* dalam menggelompokkan kabupaten/kota berdasarkan fasilitas pelayanan kesehatan di Jawa timur. Penelitian terkait Analisis *Fuzzy c-means*, menunjukan keunggulan metode ini adalah mampu melakukan pengelompokan untuk data yang tersebar secara tidak teratur sehingga dapat dikatakan sebagai salah satu metode analisis *cluster* yang baik digunakan untuk pengelompokan.

Keberadaan setiap titik data dalam kelompok ditentukan oleh derajat keanggotaannya saat menggunakan Fuzzy c-means untuk pengelompokkan data. Dari hasil pengelompokan analisis cluster, kemudian dilakukan analisis diskriminan untuk validasi hasil cluster Fuzzy c-means, Analisis diskrimianan digunakan untuk melihat tingkat keakuratan pengelompokan, implementasi metode ini digunaan untuk menganalisis kesejahteraan rakyat di Indonesia dengan menggunakan indikator: umur harapan hidup saat lahir (X_1), rata-rata lama sekolah (X_2), Angka Partisipasi Sekolah (APS) 16-18 Tahun (X_3), dan rumah tangga dengan akses air yang layak (X_4).

METODE

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari publikasi Badan Pusat Statistik pada tahun 2022 (BPS, 2022). Populasi yang digunakan dalam penelitian ini adalah 34 provinsi di Indonesia yang sekaligus digunakan sebagai sampel penelitian. Tujuan dari penelitian ini yaitu untuk menggelompokkan kesejahteraan rakyat yang dilihat berdasarkan 4 variabel (BPS, 2022), yaitu:

1. Umur/angka harapan hidup (X_1) merupakan perkiraan lama hidup rata-rata penduduk dengan asumsi tidak ada perubahan pola mortalitas menurut umur.
2. Rata-rata lama sekolah (X_2) merupakan Rata-rata jumlah tahun yang dihabiskan oleh penduduk berusia 15 tahun ke atas untuk menempuh semua jenis pendidikan formal yang pernah dijalani.
3. Angka partisipasi sekolah 16-18 tahun (X_3) merupakan Rasio anak yang sekolah pada kelompok umur tertentu terhadap jumlah penduduk pada kelompok umur yang sama.
4. Rumah tangga dengan akses air yang layak (X_4).merupakanAir yang bersumber dari ledeng, air kemasan, serta pompa, sumur terlindung dan mata air terlindung yang jarak ke tempat pembuangan limbah (*Septic Tank*) > 10 meter.

Langkah awal untuk tahapan analisis pada penelitian ini yaitu dengan menganalisis menggunakan analisis statistik deskriptif. Analisis statistik deskriptif dilakukan dengan tujuan untuk mendapatkan informasi umum berupa rata-rata, standar deviasi, nilai maksimum dan minimum dari data. Karena penelitian ini termasuk dalam analisis *cluster*, maka ada beberapa asumsi yang harus dipenuhi yaitu uji representatif yang bertujuan untuk mengetahui apakah data yang digunakan sudah mewakili populasi atau tidak dan uji multikolinearitas yang bertujuan untuk mengecek apakah dalam variabel terdapat korelasi. Data dapat dikatakan memenuhi asumsi jika data tersebut dapat mewakili populasi dan tidak adanya korelasi antara variabel, jika asumsi-asumsi ini tidak terpenuhi maka penelitian tidak dapat dilanjutkan (Santosa, 2008). Selain uji asumsi, perlu dilakukan standardisasi data yang bertujuan untuk menyeragamkan data agar tidak terjadi kekeliruan perhitungan akibat satuan dari data yang berbeda-beda. Setelah itu, dilakukan analisis *fuzzy c-means* untuk

menggelompokkan provinsi-provinsi yang ada di Indonesia. Fuzzy c-means (FCM) adalah modifikasi dari bentuk analisis cluster K-means yang fungsinya berdasar atas logika fuzzy yang berguna untuk penentuan nilai derajat keanggotaan dari setiap objek dan fungsi keanggotaan (Mohammadrezapour, Kisi, & Pourahmad, 2020). Konsep FCM ialah menentukan cluster, dimana pusat cluster ini masih belum tepat pada keaadaan awalnya dan tiap-tiap data mempunyai derajat keanggotaannya untuk setiap cluster. Dengan secara berulang melakukan perbaikan terhadap pusat cluster dan nilai keanggotaan dari setiap data menunjukkan bahwa pusat cluster bergerak kearah posisi yang tepat. Pengelompokan pada *Fuzzy clustering* dilakukan dengan mengelompokkan objek berdasarkan kedekatan dari nilai derajat keanggotaan yang dimiliki setiap objek yang mana nilai derajat keanggotaan tersebut terletak diantara rentang 0 dan 1 (Kakarla & Babu, 2019). Hasil pengelompokkan yang terbentuk kemudian di uji menggunakan analisis diskriminan untuk mendapatkan hasil penggelompokkan yang tepat dan memenuhi kriteria pengelompokkan yang baik. Dalam penelitian ini digunakan *software* R-*studio* untuk memudahkan analisis data.

HASIL DAN PEMBAHASAN

Analisis *Cluster*

Terdapat dua asumsi analisis *cluster* yaitu :

1. Sampel Representatif

Pengujian in dilakukan untuk melihat apakah sampel yang digunakan mewakili populasi yang ada. Pengujian dilakukan dengan Uji *Kaiser Mayer Olkin (KMO)*. Uji Kaiser-Mayer-Olkin (KMO) memiliki nilai 0 sampai dengan 1. Jika nilai KMO berkisar 0,5 sampai 1 maka sampel dapat dikatakan mewakili populasi atau sampel representatif Ningrat, dkk. (2016)Hasil Perhitungan Nilai KMO dengan menggunakan bantuan software *R-Studio* disajikan pada Tabel 1.

Tabel 1. Hasil Uji Kaiser Mayer Olkin (KMO)

Variabel	Nilai KMO
X_1, X_2, X_3, X_4	0,6773

Berdasarkan pada Tabel 1 di dapatkan hasil nilai KMO sebesar 0,6773 dimana nilai ini berada dalam rentang 0,5 sampai 1, sehingga dapat ditarik kesimpulan bahwa data sudah memenuhi asumsi sampel representatif dan dapat dikatakan sampel mewakili populasi.

2. Multikolinearitas

Pengujian multikolinearitas ini bertujuan untuk mendeteksi terjadinya multikolinearitas atau terdapatnya korelasi antar variabel bebas (Wibowo, Nisa, Faisol, & Setiawan, 2020). Multikolinearitas bertindak sebagai proses pembobotan yang tidak terlihat oleh pengamat yang dapat mengakibatkan hasil menjadi bias pada variabel-variabel yang saling berhubungan (Alwi & Hasrul, 2018). Uji multikolinearitas dilakukan dengan menghitungan nilai *Variance Inflation factor* (VIF). Apabila nilai VIF > 10 artinya terdapat multikolinearitas antar variabel bebas. Hasil dari pengujian VIF disajikan pada Tabel 2.

Tabel 2. Hasil Nilai VIF

Variabel	**X_1**	**X_2**	**X_3**	**X_4**
Nilai VIF	1,3057	1,7231	1,3899	1.3955

Berdasarkan Tabel 2 diketahui bahwa nilai VIF untuk semua variabel bebas memiliki nilai yang kurang dari 10 (VIF < 10). Hal ini menunjukkan bahwa tidak terjadi multikolinearitas pada variabel.

Standardisasi

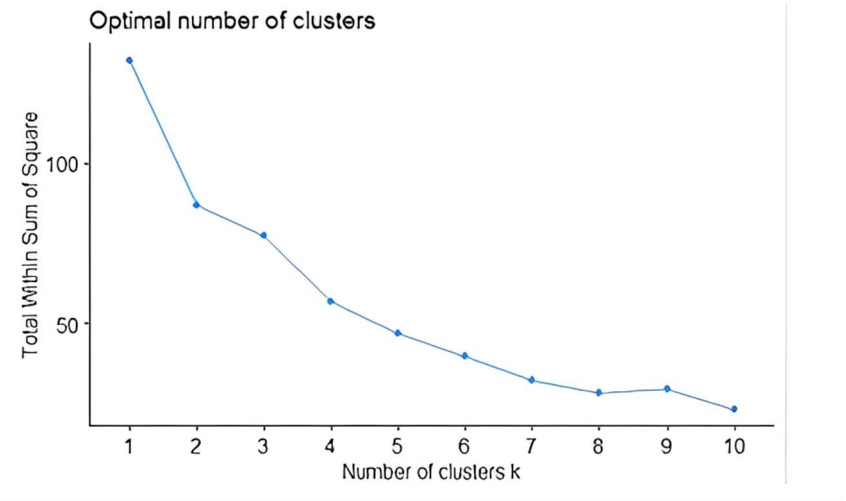
Standarisasi merupakan proses mengubah skala atribut data menjadi lebih. Dengan tujuan untuk menyetarakan jangkauan setiap variabel. Proses standarisasi data akan dilakukan apabila pada penelitian terdapat perbedaan ukuran satuan yang besar antar variabel-variabel yang diteliti (Silvi, 2018). Penelitian ini menggunakan data dengan satuan yang berbeda sehingga standardisasi dilakukan untuk membuat data menjadi seragam dengan cara mengubah data ke Z-Score. Setelah mendapatkan hasil standardisasi, data dianalisis menggunakan analisis *cluster Fuzzy c means*. Nilai standardisasi disajikan pada Tabel 3.

Tabel 3. Hasil Standardisasi data

No	**Provinsi**	**Z X_1**	**Z X_2**	**Z X_3**	**Z X_4**
1	Aceh	-0.07549982	0.70642597	1.34346705	0.24966040
2	Bali	-0.36602781	0.93149877	0.57355897	0.49786079
3	Banten	-0.22275373	0.38489340	1.47511800	-0.38738728
⋮	⋮	⋮	⋮	⋮	⋮
32	Sumatera Barat	-0.67645498	0.40632891	0.29859180	0.23429561
33	Sumatera Selatan	-1.59579697	-1.09415642	0.99850824	-0.59067523
34	Sumatera Utara	-1.67937352	-2.09090739	-1.87281562	-2.57155077

Penentuan Jumlah Cluster

Metode Elbow merupakan metode yang digunakan dalam menentukan jumlah cluster. Penggunaan metode Elbow dalam menentukan jumlah cluster yaitu dengan cara melihat selisih antara nilai dua cluster, apabila terdapat selisih yang paling besar atau signifikan maka nilai cluster tersebut merupakan nilai cluster yang baik. Penentuan jumlah *cluster* menggunakan metode elbow disajikan pada Gambar 1.



Gambar 1. Penentuan Jumlah Cluster

Berdasar dari Gambar 1 dapat diketahui dengan menggunakan metode elbow perubahan yang besar antara dua nilai cluster terjadi pada titik dua, maka dapat diartikan jumlah cluster yang optimal adalah 2 cluster. Sehingga dilanjutkan dengan menganalisis 2 cluster

Analisis *Cluster Fuzzy C-Means*

Analisis *Cluster Fuzzy C Means* merupakan salah satu metode yang dapat digunakan untuk pengelompokkan data. Dalam pengolahan data menggunakan FCM clustering menurut (Tao et al., 2018) perlu dilakukan beberapa langkah antara lain yaitu, langkah pertama,menentukan jumlah *cluster k* yang diinginkan, matriks partisi awal , dan nilai error yang diharapkan ε.

Pada penelitian ini jumlah *cluster* ditentukan menggunakan uji *elbow*, menghasilkan 2 cluster yang merupakan jumlah cluster yang optimal. Pem,bagian *cluster* dimaksudkan untuk mengelompokan provinsi di Indonesia menjadi 2 *cluster* yakni *cluster* 1 dan *cluster* 2. *Cluster* 1 di definisikan kelompok tidak sejahtera dan *cluster* 2 di definisikan kelompok sejahtera.

Langkah awal yang dilakukan adalah menentukan parameter awal yang akan di gunakan yaitu : bobot pangkat ($m = 2$), maksimum iterasi ($maxlter = 100$), eror terkecil yang di harapkan ($\zeta = 0{,}00001$), fungsi objektif awal (Po) dan iterasi awal ($t = 1$) serta membentuk matriks partisi U dengan dengan μ*'ik* = 34; dan *k* =2. Matriks partisi U yang telah terbentuk akan digunakan untuk menghitung pusat *cluster* dengan menggunakan bantuan software *R Studio* di dapatkan hasil untuk 2 *cluster*, iterasi berenti di iterasi yang ke 23 dan menghasilakan pusat *cluste*r yang disajikan pada Tabel 4.

Tabel 4. Hasil Nilai Centroid Untuk 2 *Cluster*

	X_1	X_2	X_3	X_4
Cluster 1	-0.2987030	-0.5911497	-0.5525593	-0.5555523
Cluster 2	0.3139036	0.5851880	0.5076072	0.4394711

Hasil pada Tabel 4 kemudian akan digunakan untuk menghitung jarak untuk melihat kemiripan antar objek yang akan dikelompokkan. Besarnya kesamaan antar objek tidak hanya dapat diukur menggunakan jarak tetapi juga dapat diukur melalui pola dan asosiasi.

Ukuran Kesamaan (*Measurement of Similarity*)

Konsep kesamaan objek merupakan hal mendasar dalam analisis cluster sebab pengelompokan objek didasarkan pada kesamaan karakteristik objek 7 tersebut Sartono dkk. (2003). Langkah awal dalam analisis cluster adalah mengukur jarak antar pengamatan, hal ini bersesuaian dengan prinsip cluster yaitu mengelompokan pengamatan yang menunjukkan kesamaan. Analisis cluster menggunakan jarak antar objek untuk mengukur kesamaan. Nilai jarak antar objek berbanding lurus dengan perbedaan antar objek artinya jika jarak antar objek nilainya semakin membesar, maka nilai perbedaan antar objek juga akan semakin membesar begitupun sebaliknya.

Salah satu metode untuk mengukur berdasarkan jarak adalah Squared euclidean distance. Terdapat kelebihan jika menggunakan jarak Squared euclidean distance, yaitu jarak antar objek yang dihasilkan memiliki nilai yang besar, sehingga terdapat perbedaan yang jelas jarak antar objek Maylana (2014). Jarak Squared euclidean distance merupakan variasi dari jarak euclidean. Cara membedakan dua rumus ini yakni pada jarak euclidean kuadrat akarnya dihilangkan atau dikuadratkan, sementara pada jarak euclidean akarnya tidak dikuadratkan. Berdasarkan hasil software *Rstudio* jarak ke centroid cluster 2 disajikan pada Tabel 5.

Tabel 5. Jarak ke Centroid 2 *Cluster*

	Cluster1	**Cluster2**
1	5,9768039	0,9010189
2	4,7008095	0,5899944
3	5,0981845	1,9478849
⋮	⋮	⋮
32	2,4859771	1,0985831
33	4,3425203	7,7693328
34	9,9628619	25,8673076

Hasil perhitungan jarak ke centroid cluster digunakan untuk mencari nilai fungsi objektif dengan menggunakan *Rstudio* yaitu sebesar 60,88064. Langkah selanjutnya setelah mendapatkan nilai pusat *cluster*, dilakukan perhitungan perubahan nilai matriks partisi U dengan derajat keanggotaan yang baru. Pengelompokkan masing-masing objek ke dalam *cluster* didasarkan pada nilai derajat keanggotaan terbesar. Hasil perubahan matriks partisi *U* dengan derajat keanggotaan yang baru yang diolah menggunakan *R Studio* disajikan pada Tabel 6.

Tabel 6. matriks partisi *U* dengan derajat keanggotaan yang baru

Ods	**Provinsi**	**Cluster 1**	**Cluster 2**	**Kelompok**
1	Aceh	0.13100388	0.86899612	2

Ods	Provinsi	Cluster 1	Cluster 2	Kelompok
2	Sumatera Utara	0.11151346	0.88848654	2
3	Sumatera Barat	0.27645034	0.72354966	2
⋮	⋮	⋮	⋮	⋮
32	Maluku Utara	0.30647673	0.69352327	2
33	Papua Barat	0.64146571	0.35853429	1
34	Papua	0.35853429	0.27805776	1

Berdasarkan Tabel 6 Pengelompokan masing-masing objek kedalam cluster atas dasar nilai derajat keanggotaan terbesar. Terlihat nilai derajat keanggotaan terbesar dari provinsi Aceh ditujukan oleh cluster 2. Dari hasil tersebut menunujukan bahwa provinsi Aceh masuk pada cluster 2. Hasil Pengelompokan yang di olah menggunakan *software Rstudio* disajikan pada Tabel 7.

Tabel 7. Hasil pengelompokan 2 *Cluster*

Cluster	Jumlah Anggota
1	16
2	18

Berdasarkan hasil yang tersaji pada Tabel 7 dapat diketahui bahwa terdapat 16 provinsi yang terdapat pada cluster 1 dan 18 provinsi pada cluster 2. Provinsi yang masuk pada cluster 1 adalah Provinsi Lampung, Sumatera Selatan, Gorontalo, Jambi, Bengkulu, Nusa Tenggara Timur, Jawa Tengah, Kalimantan Tengah, Nusa Tenggara Barat, Sulawesi Barat, Kalimantan Barat, Papua, Kalimantan Selatan, Sulawesi Selatan, Papua Barat, dan Kepulauan Bangka Belitung. Sementara yang masuk pada cluster 2 adalah Provinsi Aceh, Sulawesi Tenggara, Sumatera Utara, Jawa Barat, Riau, Sumatera Barat, DKI Jakarta, Kepulauan Riau, DI Yogyakarata, Jawa Timur, Banten, Kalimantan Timur, Maluku Utara, Bali, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah dan Maluku. Hasil dari 2 cluster dijadikan sebagai variabel terikat (Y) sehingga dapat dilakukan anaisis diskriminan

Analisis Diskriminan

Penggunaan analisis diskriminan yaitu untuk mengklasifikasi atau mengelompokkan suatu individu atau objek kedalam kelompok yang saling bebas dan menyeluruh berdasarkan beberapa faktor penjelas (Lembang et al., 2019). Pada penelitian ini, pilihan yang tepat menggunkan analisis diskriminan sebagai salah satu metode validasi dari hasil solusi analisis *cluster* yang telah didapat.

Sebelum melakukan uji analisis diskrimian diperlukan melakukan uji asumsi analisis diskriminan, uji analisis diskriman tersebut yaitu :

1. Uji normal multivariat dilakukan dengan menggunakan uji *Shapiro wilk* dengan hipotesis sebagai berikut:

   H_0 : data berdistribusi normal multivariat

   H_1 : data tidak berdistribusi normal multivariat

   Statistik uji metode uji *Shapiro wilk* untuk uji normal multivariat dapat dilihat pada persamaan berikut (Trimardiani, Akbar, & Wibawati, 2020).

$$W = \frac{1}{p}\sum_{j=1}^{p} w_{z_j}$$

H_0 akan ditolak jika nilai $p - value <$ taraf signifikansi α ($\alpha = 0{,}05$)

2. Berdasarkan bantuan software *Rstudio* hasil uji *Shapiro wilk* didapatkan p-value = 0.0569, dimana hal ini menunjukkan bahwa p -value > α ($\alpha = 0{,}05$) hal ini berarti data berdistribusi normal multivariat.
3. Asumsi homogenitas varians dilakukan untuk menguji matriks ragam-peragam antar kelompok adalah sama. Pengujian dilakukan dengan statsitik uji *Box'M*. Hasil pengujian asumsi homogenitas yang disajikan pada Tabel 8.

Tabel 8. Hasil Pengujian Asumsi Homogenitas Varians

Jumlah	**uji *Box'M***	**(*P-value*)**	**Keputusan**
2	16,424	0,0881	Homogen

Berdasarkan hasil uji homogenitas yang disajikan pada Tabel 8 menunjukkan bahwa *cluster* 2 menghasilkan nilai p-value sebesar 0,0881, artinya nilai p-value ini > dari taraf signifikansi α (0,05), hal ini berarti bahwa asumsi homogenitas varians untuk *cluster* 2 Terpenuhi.

Analisis Diskriminan Untuk 2 *Cluster*

Berdasarkan hasil uji asumsi didapatkan *cluster* 2 memenuhi asumsi, sehingga *cluster* 2 dapat dilanjutkan dengan analisis diskriminan. Model analisis diskriminan digunakan sebagai metode validasi dari hasil *cluster* yang telah didapat. Hasil dari kelompok diskriminan dapat dilihat pada model berikut.

$$D_1 = 0{,}03195955(X_1) + 1{,}14108621(X_2) + 0.04987572(X_3) + 0.07110288(X_4)$$

Model ini dibuat sehingga dapat menguji Ketepatan hasil analisis *Fuzzy c-means* dengan cara membuat tabel klasifikasi. Hasilnya disajikan pada Tabel 9.

Tabel 9. Confusion matrix and statistics

Kelompok Aktual	**Kelompok Prediksi**		**Jumlah Amatan**
	1	**2**	
1	17	1	18
2	0	16	16

Berdasarkan Tabel 9 dijelaskan bahwa kelompok aktual pertama pada kelompok prediksi pertama hasilnya sebanyak 17 amatan, sedangkan untuk kelompok aktual pertama pada kelompok prediksi kedua sebanyak 1 amatan. Untuk Kelompok aktual kedua pada kelompok prediksi pertama sebanyak 0 amatan, sedangkan untuk kelompok aktual ke dua pada kelompok prediksi kedua sebanyak 16 amatan. Setelah fungsi diskriminan terbentuk,

selanjutnya adalah menilai validitas dari analisis diskriminan. Hasil dapat didasarkan pada analisis sampel ataupun validitas sampel yang dapat dilakukan dengan cara sebagai berikut:

1. Hit ratio

Hit Ratio digunakan untuk menilai keakuratan dari fungsi diskriminan. Dinyatakan sebagai model yang baik jika model yang dihasilkan dapat memberi tingkat akurasi hit ratio lebih besar sama dengan 50% maka model tersebut dianggap baik, dan sebaliknya jika model yang dihasilkan memberi tingkat akurasi Hit Ratio kurang dari 50%, maka model tersebut dianggap kurang baik. Hit ratio dapat dirumuskan sebagai berikut.

$$\text{Hit Ratio} = \frac{n \text{ benar}}{\text{N}} \times 100\%$$

dimana n benar merupakan jumlah sampel yang diklasifikasi secara tepat dan N merupakan banyaknya sampel secara keseluruhan. Hasil persentase ketepatan pengelompokan *cluster Fuzzy C-means* pada penelitian ini dihitung menggunakan hit rasio sebagai berikut

$$\text{Hit Ratio} = \frac{17 + 16}{34} \times 100\% = 97,0588\%.$$

Hal ini menunjukan bahwa nilai hit ratio sebesar 97,0588%. Berdasarkan hasil nilai hit ratio tersebut dapat diketahui bahwa nilai hit ratio lebih besar dari 50% sehingga dapat disimpulkan bahwa pengelompokan *Fuzzy c-means* tingkat akurasi tinggi yaitu sebesar 97,0588%.

2. Akurasi Model

Validitas analisis diskriminan dapat ditentukan dengan melihat keakuratan fungsi diskriminan. Keakuratan fungsi diskriminan untuk mengetahui apakah hasil klasifikasi sudah akurat atau belum. Pengujian akurasi model dapat menggunakan uji press' Q. Rumus uji press Q sebagai berikut (Rismia, Widiharih, & Santoso, 2021).

$$\text{press } Q = \frac{[N - (nK)]^2}{N(K - 1)}$$

dimana N merupakan banyaknya sampel, n merupakan banyaknya sampel yang diklasifikasi secara tepat dan K nerupakan banyaknya *cluster*. klasifikasi yang akurat dapat ditunjukkan dari nilai Press Q yang lebih besar dari pada nilai pada tabel χ^2 dengan $\alpha = 0,05$ dan derajat kebebasan bernilai 1. Pada penelitian ini didapatkan nilai *Press'Q* sebagai berikut.

$$\text{Press}'\text{sQ} = \frac{(34 - (33 \times 2))^2}{34(2 - 1)} = 30,1176.$$

Hasil nilai press Q yang didapatkan sebesar 30.1176, kemudian nilai *Pres'Q* di bandingkan dengan nilai tabel $\chi^2_{k=1,5\%} = 3,84$. Maka didapatkan Nilai *Press'Q* $(30.1176) >$

dari $\chi^2_{k=1,5\%}$ (3,84) sehingga dapat dikatakan bahwa pengelompokan *Fuzzy c-means* dengan 2 *cluster* akurat.

KESIMPULAN

Hasil Pengelompokan Kesejahteraan rakyat di Indonesia dengan analisis *Fuzzy c-means* menggunakan validasi analisis diskriminan memberikan hasil yang baik berdasarkan nilai hit rasio sebesar 97,0588% dan nilai *Press'Q* lebih besar dari $\chi^2_{k=1,5\%}$. Oleh karena itu dapat disimpulkan bahwa pengelompokan dengan menggunakan analisis *cluster Fuzzy c-means* dengan menggunakan validasi analisis diskriminan sudah akurat dan dapat diterapkan. Hasil pengelompokkan kesejahteraan rakyat di Indonesia berdasarkan indikator kesejahteraan rakyat dengan 2 *cluster* di peroleh bahwa *cluster* pertama terdiri dari daerah-daerah yang memiliki indikator kesejahteraan yang rendah dan *cluster* kedua terdiri dari dearah-daerah yang memiliki nilai indikator kesejahteraan rakyat yang tinggi.

DAFTAR PUSTAKA

Alwi, W., & Hasrul, M. (2018). Analisis Klaster Untuk Pengelompokkan Kabupaten/Kota Di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat. *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)*, *6*(1), 35.

BPS. (2022). Statistik Indonesia 2022. *Badan Pusat Statistik: https://www.bps.go.id/publication/2022/02/25/0a2afea4fab72a5d052cb315/statistik-indonesia-2022.html*

BPS. (2022). Indikator Kesejahteraan Rakyat 2022. *Badan Pusat Statistik: https://www.bps.go.id/publication/2022/11/30/71ae912cc39088ead37c4b67/indikator-kesejahteraan-rakyat-2022.html*

Dani, A. T. R., Wahyuningsih, S., & Rizki, N. A. (2021). Pengelompokkan Data Runtun Waktu menggunakan Analisis Cluster. *EKSPONENSIAL*, *11*(1), 29–38.

Kakarla, N. M. L., & Babu, G. R. M. (2019). Application of fuzzy K-means (FKM) algorithms in identifying better clusters of few drugs from drugbank database. *International Journal of Innovative Technology and Exploring Engineering*, *8*(6), 655–660.

Lembang, F. K., Talakua, M. W., & Hasanudin, M. S. (2019). Misklasifikasi Penjurusan Mahasiswa FMIPA Universitas Pattimura Tahun Akademik 2016/2017 Menggunakan Metode Analisis Diskriminan Berganda. *Variance: Journal of Statistics and Its Applications*, *1*(2), 64–74.

Maghfiroh, A. W., Ulinnuha, N., & Fanani, A. (2019). Penerapan Fuzzy C-Means dalam Mengelompokkan Kabupaten/Kota Berdasarkan Fasilitas Pelayanan Kesehatan Di Jawa Timur. *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun*, *4*(1), 8-14.

Maylana, R. (2014). *Beberapa Metode Pautan Pada Analisis Kelompok Menggunakan Jarak Euclidean dan Square Euclidean*. Universitas Brawijaya.

Mohammadrezapour, O., Kisi, O., & Pourahmad, F. (2020). Fuzzy c-means and K-means clustering with genetic algorithm for identification of homogeneous regions of groundwater santosquality. *Neural Computing and Applications*, *32*, 3763–3775.

Ningrat, D. R., Di Asih, I. M., & Wuryandari, T. (2016). Analisis cluster dengan algoritma K-Means dan Fuzzy C-Means clustering untuk pengelompokan data obligasi korporasi. *Jurnal Gaussian*, *5*(4), 641–650.

Rismia, E. R., Widiharih, T., & Santoso, R. (2021). Klasifikasi Regresi Logistik Multinomial Dan Fuzzy K-Nearest Neighbor (Fk-Nn) Dalam Pemilihan Metode Kontrasepsi Di Kecamatan Bulakamba, Kabupaten Brebes, Jawa Tengah. *Jurnal Gaussian*, 10(4), 476-487.

Santosa, P. I. (2018). Metode penelitian kuantitatif: Pengembangan hipotesis dan pengujiannya menggunakan SmartPLS. *Yogyakarta: Andi Ofset*.

Sartono, B., Affendi, F. M., Syafitri, U. D., Sumertajaya, I. M., & Anggraeni, Y. (2003). Analisis peubah ganda. *Bogor: Fakultas Matematika Dan Ilmu Pengetahuan Alam IPB*.

Silvi, R. (2018). Analisis Cluster dengan Data Outlier Menggunakan Centroid Linkage dan K-Means Clustering untuk Pengelompokkan Indikator HIV/AIDS di Indonesia. *J. Mat.“Mantik*, *4*(1), 22–31.

Tao, Y., Zhang, Y., & Wang, Q. (2018). Fuzzy c-mean clustering-based decomposition with GA optimizer for FSM synthesis targeting to low power. *Engineering Applications of Artificial Intelligence*, *68*, 40–52.

Trimardiani, Y., Akbar, M. S., & Wibawati. (2020). Penerapan Diagram Kendali Multivariate Exponentially Weighted Moving Covariance Matrix (MEWMC) pada Pengendalian Kualitas Proses Produksi Air di PDAM Surya Sembada Kota Surabaya. *Jurnal Sains dan seni ITS*, 9(2), 2337-3520.

Wibowo, R. A., Nisa, K., Faisol, A., & Setiawan, E. (2020). Simulasi Pemilihan Metode Analisis Cluster Hirarki Agglomerative Terbaik Antara Average Linkage dan Ward pada Data yang Mengandung Masalah Multikolinearitas. *Jurnal Siger Matematika*, *1*(2), 49–55.

Wijayanti, W., HG, I. R., & Yanuar, F. (2021). Penggunaan Metode Fuzzy C-Means Untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Kesehatan Lingkungan. *Jurnal Matematika UNAND*, *10*(1), 129-136.