Jurnal Riset Matematika dan Sains Terapan
Volume 2, Nomor 1, Mei 2022, 21-28
ISSN: 2809-4581

Analisis Teknik non-Hierarki untuk Pengelompokan Kabupaten/Kota di Provinsi Jawa Barat Berdasarkan Indikator Kesejahteraan Rakyat 2020

Dini Andiani[a)], Siti Dwi Rahayu Septiani [b)] dan Alham Riana [c)]

[1]Program Studi Matematika, FMIPA, Universitas Bale Bandung, Bandung, Indonesia

[a)] diniandiani367@gmail.com
[b)]sitidwirahayu@unibba.ac.id
[c)]alhamriana26@gmail.com

Abstrak. Mewujudkan masyarakat yang sejahtera merupakan tujuan dari setiap pembangunan di berbagai wilayah. Tidak meratanya pembangunan baik dari sisi sarana atau pun prasarana menjadi salah satu alasan tidak meratanya kesejahteraan masyarakat. Selain itu, sasaran yang tidak tepat dalam pembangunan menjadi unsur lain yang ikut terlibat terbentuknya kesejahteraan rakyat. Untuk terciptanya masyarakat yang sejahtera secara merata, maka mengidentifikasi karakteristik berdasarkan tingkat kesejahteraan rakyat menjadi salah satu solusi. Penelitian ini bertujuan mengelompokkan kabupaten/kota berdasarkan sepuluh indikator kesejahteraan rakyat di Provinsi Jawa Barat dengan menggunakan *K-means Cluster*. Langkah-langkah yang digunakan dalam analisis *Cluster* non-hierarki *K-means* yaitu standarisasi data, menentukan *k* sebagai jumlah Cluster yang akan dibentuk, menentukan centroid, menghitung jarak setiap data ke setiap centroid,menentukan centroid baru, menghitung jarak setiap data ke setiap centroid baru, mengulangi langkah hingga nilai jarak data ke setiap centroid tidak berubah, interpretasi dan validasi cluster. Berdasarkan hasil analisis, Provinsi Jawa Barat dapat dibentuk menjadi 3 Cluster, yaitu Cluster 1 terdiri dari 4 kabupaten/kota, Cluster 2 terdiri dari 7 kabupaten/kota dan Cluster 3 terdiri dari 16 kabupaten/kota. Kemudian, berdasarkan Uji validasi Cluster diketahui bahwa terdapat 2 variabel yang membedakan dalam pengklasifikasian, yaitu variabel produk domestik Regional Bruto (PDRB) dan variabel kepadatan penduduk.

PENDAHULUAN

Kesejahteraan rakyat pada dasarnya merupakan suatu kondisi yang bentuknya dinamis atau dengan kata lain nilai kuantitatifnya tidak akan pernah berhenti karena akan terus berubah seiring dengan perkembangan kebutuhan hidup manusia. Dalam melaksanakan program pembangunan perlu adanya identifikasi berdasarkan karakteristik tingkat kesejahteraan rakyat tiap daerah agar dalam pengambilan kebijakan dan strategi pembangunan bisa tepat sasaran dan tepat guna. Berdasarkan proyeksi BPS Provinsi Jawa Barat, jumlah penduduk di Jawa Barat tahun 2020 adalah sebesar 48,27 juta jiwa dengan kepadatan penduduk 1.402 jiwa/km^2. Jumlah penduduk mengalami kenaikan dari tahun sebelumnya sebesar 0,18%. Ketika jumlah penduduk bertambah, berarti pemerintah juga harus terus menambah jumlah fasilitas hidup layak bagi masyarakatnya. Selain masalah tentang kependudukan, angka pengangguran di Provinsi Jawa Barat juga masih tinggi dan mengalami kenaikan dari tahun sebelumnya. Sesuai data Tahun 2020, angka pengangguran terbuka mencapai 3.922.500 jiwa atau sekitar 10,46% dibandingkan tahun sebelumnya dengan tingkat pengangguran terbuka 2.968.368 jiwa atau sekitar 8,04 %. Implikasi tingkat

Diterima 9 Mei 2022; Disetujui untuk publikasi 23 Mei 2022,
Diterbitkan oleh Program Studi Matematika UNIBBA.

pengangguran terbuka akan semakin meningkat jika tidak ada perubahan strategi dalam menciptakan lapangan kerja. (Badan Pusat Statistik Provinsi Jawa Barat, 2020). Oleh sebab itu, sangat penting mempertimbangkan pengelompokkan 27 kabupaten/kota di Jawa Barat berdasarkan 10 indikator kesejahteraan rakyat yaitu PDRB (X_1), Kepadatan Penduduk (X_2) , Jumlah Penduduk Miskin(X_3) , Rata-rata pengeluaran per kapita(X_4), Jumlah Angkatan Kerja(X_5), Umur Harapan Hidup(X_6), Rata-Rata Lama Sekolah(X_7), Angka Harapan Lama Sekolah(X_8), Tingkat Pengangguran Terbuka(X_9), Kepemilikan Rumah Sendiri(X_{10}).

Analisis multivariat merupakan analisis beberapa variabel dalam hubungan tunggal atau banyak hubungan. Analisis Multivariat juga didefinisikan sebagai analisis dimana masalah yang diteliti bersifat multidimensional dan menggunakan tiga atau lebih variabel [20]. Salah satu dari analisis multivariat yang dapat digunakan yaitu analisis Cluster. Analisis Cluster merupakan proses pengelompokan data ke dalam kelompok berdasarkan parameter tertentu sehingga objek-objek dalam sebuah Cluster/kelompok memiliki tingkat homogenitas internal dan heterogenitas eksternal yang tinggi. Analisis Cluster bertujuan untuk mengelompokkan n objek berdasarkan p variat yang memiliki kesamaan karakteristik diantara objek-objek tersebut. Objek tersebut akan diklasifikasikan ke dalam satu atau lebih Cluster (kelompok) sehingga objek-objek yang berada dalam satu Cluster akan mempunyai kemiripan atau kesamaan karakter [22]. Secara umum analisis Cluster dibagi menjadi dua metode yaitu metode hierarki dan metode non-hierarki. Pada metode hierarki memulai pengelompokan dengan dua atau lebih objek yang mempunyai kesamaan paling dekat, kemudian proses diteruskan ke objek lain yang mempunyai kedekatan kedua, demikian seterusnya sehingga Cluster akan membentuk semacam pohon dimana ada hierarki (tingkatan) yang jelas antar objek. Di dalam metode hierarki sendiri terdapat beberapa metode, diantaranya 2 metode pautan tunggal (*single linkage*), metode pautan lengkap (*complete linkage*), metode pautan rata-rata (average linkage), metode antar pusat (*centroid linkage*), dan metode Ward (*ward's merthod*). Sedangkan pada metode non-hierarki metode ini justru memulai dengan menentukan terlebih dahulu jumlah Cluster yang diinginkan. Setelah jumlah Cluster ditentukan, baru proses Cluster dilakukan tanpa mengikuti proses hierarki. Metode non-hierarki bertujuan mengelompokkan n objek ke dalam k kelompok ($< n$), metode yang termasuk dalam metode non-hierarki adalah metode K-means (C-means) [22] . Metode K-means digunakan sebagai alternatif metode Cluster untuk data dengan ukuran yang besar karena memiliki kecepatan yang lebih tinggi dibandingkan metode hierarki. Algoritma K-means sangat terkenal karena kemudahannya untuk meng-Cluster data yang besar dan data outlier dengan sangat cepat [11]. Dalam penentuan kesamaan atau kemiripan pada sebuah objek diperlukan sebuah parameter yang tepat dalam mengukur similaritas tersebut. Di antara jarak yang dapat digunakan dalam analisis Cluster adalah jarak Euclidean. Jarak Euclidean ini merupakan jarak yang dapat digunakan dalam pengukuran bidang pada dimensi banyak, jadi jarak ini adalah jarak yang cocok diaplikasikan pada hampir semua objek dalam penarikan sebuah Cluster baik dalam metode hierarki maupun non-hierarki.

TINJAUAN PUSTAKA

Beberapa teori terkait penelitian ini adalah analisis multivariat, analisis cluster, metode hierarki, metode non-hierarki, dan validasi cluster. Analisis multivariat merupakan analisis beberapa variabel dalam hubungan tunggal atau banyak hubungan. Analisis multivariat juga didefinisikan sebagai analisis dimana masalah yang diteliti bersifat multidimensional dan menggunakan 3 atau lebih variabel. Jika terdapat sebanyak n objek dan p variabel, maka observasi objek ke-i dan variabel ke-j yang dinotasikan xij dengan i = 1, 2, …, n dan j = 1, 2, … , p dapat ditampilkan pada tabel berikut:

TABEL 1. Hubungan beberapa variabel dan beberapa objek.

	Var 1	Var 2	…	Var j	…	Vara p
Objek 1	x_{11}	x_{12}	…	x_{1j}	…	x_{1p}
Objek 2	x_{21}	x_{22}	…	x_{2j}	…	x_{2p}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Objek i	x_{i1}	x_{i2}	…	x_{ij}	…	x_{ip}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Objek n	x_{n1}	x_{n2}	…	x_{nj}	…	x_{np}

Analisis Cluster adalah salah satu teknik dari analisis multivariat yang mempunyai tujuan utama yaitu untuk mengelompokan objek-objek yang berdasarkan kemiripan karakteristik yang dimilikinya. Analisis Cluster akan membagi sejumlah data pada satu atau beberapa Cluster tertentu. Analisis klaster adalah teknik analisis multivariat yang bertujuan untuk mengelompokkan hal-hal yang memiliki kesamaan berdasarkan sifat-sifat yang sebanding. Analisis klaster membagi sejumlah besar data menjadi satu atau lebih klaster yang berbeda. Sebuah Cluster yang baik adalah Cluster yang memiliki: 1). Homogenitas (kesamaan) yang tinggi antara anggota dalam satu Cluster (*within-Cluster*); 2). Heterogenitas (perbedaan) yang tinggi antara Cluster yang satu dengan Cluster yang lainnya (*between Cluster*).

Teknik hierarki (*hierarchical method*) adalah teknik Clustering membentuk konstruksi hierarki atau berdasarkan tingkatan tertentu seperti struktur pohon. Dengan demikian proses pengelompokannya dilakukan secara bertingkat atau bertahap. Hasil dari pengelompokan ini dapat disajikan dalam bentuk dendogram. Metode-metode yang digunakan dalam teknik hierarki: 1). *Agglomerative method* dan 2). *Divisive Methods*.

Agglomerative method ini dimulai dengan kenyataan bahwa setiap objek membentuk Cluster nya masing-masing. Setelah itu, 2 objek dengan jarak terdekat bergabung. Selanjutnya objek ketiga akan bergabung dengan Cluster yang yang yang ada atau bersama objek lain dan membentuk Cluster baru. Hal ini tetap memperhitungkan jarak kedekatan antara objek. Proses akan berlanjut hingga akhirnya terbentuk 1 Cluster yang terdiri dari keseluruhan objek. Ada beberapa teknik pengelompokan dalam metode agglomerative yaitu: a). *Single linkage* (jarak terdekat atau pautan tunggal), pengelompokan ini terjadi bila Cluster-Cluster digabungkan menurut jarak antara anggotaanggota yang terdekat di antara dua Cluster atau lebih [14]. b). *Complete linkage* (jarak terjauh atau pautan lengkap), pengelompokan ini terjadi bila Cluster-Cluster digabungkan menurut jarak antara anggotaanggota yang terjauh di antara dua Cluster [14]. c). *Average linkage* (jarak rata-rata atau pautan rata-rata) pengelompokan ini menggabungkan objek menurut jarak rata-rata pasangan pasangan anggota masing-masing pada himpunan antara dua Cluster (Eko Prasetyo 2012). d). *Ward's method*, metode ini menggunakan perhitungan yang lengkap dan memaksimumkan homogenitas di dalam satu Cluster [5].

Divisive Methods berlawanan dengan metode agglomerative. Metode ini diawali dengan satu kelas terbesar yang mencakup semua observasi (objek) detik selanjutnya objek yang mempunyai ketidakmiripan yang cukup besar akan dipisahkan sehingga membentuk Cluster yang lebih kecil. Pemisahan ini dilanjutkan sehingga mencapai sejumlah Cluster yang diinginkan. Metode yang sering digunakan adalah *splinter average distance methods* di mana pada perhitungan jarak rata-rata masing-masing objek dengan objek pada grup Splinter dan jarak rata-rata objek tersebut dengan objek lain pada grupnya. Proses tersebut dimulai dengan pemisahan dengan jarak terjauh sehingga terbentuklah dua grup. Kemudian dibandingkan dengan jarak rata-rata masing-masing objek dengan grup Splinter dengan grupnya sendiri. Apabila suatu objek mempunyai jarak yang lebih dekat ke grup Splinter daripada ke grup nya sendiri, maka objek tersebut haruslah dikeluarkan dari grupnya dan dipisahkan ke grup Splinter. Apabila komposisinya sudah stabil, yaitu jarak suatu objek ke grupnya selalu lebih kecil daripada jarak objek itu ke grup Splinter, maka proses berhenti dan dilanjutkan dengan tahap pemisahan dalam grup.

Metode nonHierarki

Prosedur pada metode non-hierarki dimulai dengan memilih sejumlah nilai Cluster awal sesuai dengan jumlah yang diinginkan, kemudian obyek pengamatan digabungkan ke dalam Clusters tersebut. Dua kelemahan dari prosedur non-hierarki ialah bahwa banyaknya klaster disebutkan/ditentukan sebelumnya dan pemilihan pusat Cluster

sembarang. Hasil Cluster mungkin tergantung pada bagaimana pusat dipilih. Banyak program non hierarki, memilih kobjek (kasus) yang pertama, tanpa ada nilai yang hilang sebagai pusat Cluster awal (k = banyaknya Cluster). Jadi hasil Cluster mungkin tergantung urutan observasi dalam data. Bagaimanapun juga, Cluster non hierarki lebih cepat daripada metode hierarki dan lebih menguntungkan kalua jumlah objek/kasus atau observasi besar sekali (sampel besar). Masalah utama dalam metoda non hierarki adalah bagaimana memilih bakal Cluster. Harus disadari pengaruh pemilihan bakal Cluster terhadap hasil akhir analisis Cluster. Bakal Cluster pertama adalah observasi pertama dalam set data tanpa missing value. Bakal kedua adalah observasi lengkap berikutnya (tanpa missing data) yang dipisahkan dari bakal pertama oleh jarak minimum khusus.

Ada tiga prosedur dalam metode non hierarki, yaitu: 1). Sequential threshold. Metode ini dimulai dengan memilih bakal Cluster dan menyertakan seluruh objek dalam jaraktertentu. Jika seluruh objek dalam jarak tersebut disertakan, bakal Cluster kedua terpilih, kemudian proses terus berlangsung seperti sebelumnya. 2). Parallel Threshold Metode ini memilih beberapa bakal Cluster secara simultan pada permulaannya dan menandai objek-objek dengan jarak permulaan ke bakal terdekat. 3). Optimalisasi Metode ketiga ini mirip dengan kedua metode sebelumnya kecuali pada penandaan ulang terhadap objek-objek. Hal penting lain dalam tahap keempat adalah menentukan jumlah Cluster yang akan dibentuk.Karena tidak ada kriteria statistik internal digunakan untuk inferensia, seperti tes signifikansipada teknik multivariat lainnya, para peneliti telah mengembangkan beberapa kriteria dan petunjuk sebagai pendekatan terhadap permasalahan ini dengan memperhatikan substansi dan aspek konseptual. 4). K-means Cluster Salah satu metode non-hierarki adalah K-means. Metode K-means ini digunakan sebagai alternatif metode Cluster untuk data dengan ukuran yang besar karena memiliki kecepatan yang lebih tinggi dibandingkan metode hierarki. (Sartono dkk. 2003) menyarankan penggunaan untuk menjelaskan algoritma dalam penentuan suatu objek ke dalam Cluster tertentu berdasarkan rataan terdekat.

Validasi Cluster

Setelah Cluster terbentuk, baik dengan metode hierarki maupun nonhierarki, langkah selanjutnya adalah melakukan interpretasi terhadap Cluster yang terbentuk, yang pada intinya memberi nama spesifik untuk menggambarkan isi Cluster tersebut. Pengelompokan tidak bermanfaat apabila kita tidak mengetahui profil setiap kelompok. Untuk menginterpretasikan Cluster dan membuat profil semua Cluster, digunakan rata-rata setiap Cluster pada setiap variabel [19].

Menurut Santoso [17], untuk menguji validasi Cluster digunakan uji parsial F dan signifikansi yang terdapat pada tabel ANOVA. Tabel ANOVA digunakan untuk melihat tingkat signifikansi antar Cluster. Teknik ANOVA akan menguji variabilitas dari observasi masing-masing kelompok dan variabilitas antar mean kelompok. Melalui kedua variabel tersebu, akan dapat ditarik kesimpulan mengenai means populasi. Jika diketahui bahwa xj merupakan variabel pembeda pengklasifikasian dan c adalah banyaknya Cluster yang terbentuk. Nilai F diperoleh dari rata-rata jumlah kuadrat (mean square) dengan rata-rata jumlah kuadrat dalam kelompok.

METODE PENELITIAN

Sumber data yang digunakan dalam penelitian ini berasal dari BPS (Badan Pusat Statistik) [3] Provinsi Jawa Barat. Data tersebut merupakan pendataan penduduk tingkat Provinsi selama periode 2020. Populasi dalam penelitian ini adalah seluruh data kependudukan di Indonesia, sedangkan sempel dalam penelitian ini adalah data kependudukan di Provinsi Jawa Barat selama tahun 2020. Variabel yang digunakan dalam penelitian ini adalah adalah PDRB, Kepadatan Penduduk, Jumlah Penduduk Miskin, Rata-rata Pengeluaran per Kapita, Jumlah Angkatan Kerja, Umur Harapan Hidup, Rata-Rata Lama Sekolah, Angka Harapan Lama Sekolah, Tingkat Pengangguran Terbuka, Kepemilikan Rumah Sendiri.

Penelitian ini menggunakan analisis teknik non-hierarki metode *K-means Cluster*.

HASIL DAN PEMBAHASAN

Data yang digunakan merupakan data indikator kesejahteraan rakyat penduduk Jawa Barat tahun 2020. Data tersebut selanjutnya diolah menggunakan metode non-hierarki *K-means cluster*. Langkah awal yang dilakukan adalah standarisasi data dengan melakukan transformasi (standarisasi) pada data asli sebelum dianalisis lebih lanjut. Hasil standarisasi tersebut dapat dilihat pada tabel berikut:

TABEL 2. Data indikator kesejahteraan rakyat setelah ditransformasi.

Objek	Z 1	Z 2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10
1	1,6398	-0,3678	3,2156	-0,488	3,048	-0,763	-0,1678	-0,39881	1,81051	0,2128
2	-0,1375	-0,6949	0,3001	-0,975	0,4152	-0,905	-1,0253	-0,72044	-0,1596	0,8397
3	-0,3749	-0,6961	0,896	-0,954	0,4795	-1,503	-0,9486	-1,02919	0,44948	0,3158
4	0,4329	-0,3834	1,188	-0,305	1,3618	0,9172	0,2923	-0,12865	-0,5881	-0,147
5	-0,2628	-0,643	1,1799	-0,922	0,4307	-0,592	-0,7116	-1,13211	-0,4285	0,5563
…										
25	-0,5306	2,3193	-1,1397	1,27	-1	1,2732	1,6865	1,299356	1,39464	-0,594
26	-0,6444	-0,046	-0,5929	-0,125	-0,918	-0,065	0,5502	0,849084	-0,8359	0,7109
27	-0,845	-0,4929	-1,3443	-0,405	-1,325	-0,891	0,0622	0,566056	-1,3652	-0,053

Berdasarkan output pada tabel di atas, data tersebut adalah hasil proses standarisasi yang mengacu pada z-score dengan ketentuan, angka negatif (-) berarti data dibawah rata-rata total, angka positif (+) berarti diatas ratarata total. Selanjutnya dilakukan uji multikolinearitas untuk mengetahui adanya variabel independen yang memiliki kesamaan karakteristik antar variabel independen lain, yaitu dengan cara mencari nilai VIF dengan asumsi bahwa jika nilai VIF dari suatu variabel memiliki nilai lebih dari 10, maka variabel tersebut mengindikasikan terjadi multikolinieritas. Dari hasil pengujian multikolinearitas, terdapat beberapa variabel yang memiliki nilai VIF lebih dari 10 yaitu, Z3,Z4 dan Z5. Dalam hal ini dapat diasumsikan bahwa variabel tersebut mengalami multikolinearitas. Dalam mengatasi masalah multikolinearitas dilakukan dengan menggunakan metode Principal Componen Analysis (PCA). Hasil output analisis PCA disajikan pada tabel berikut.

TABEL 3 . Data output analisis PCA

Objek	CS 1	CS 2	CS3	CS4	CS5	CS6	CS7	CS8	CS9	CS10
1	-0,9182	-4,89	0,2298	-1,11	-0,0492	-0,804	0,1698	-0,2348	-0,0985	-0,06529
2	-2,1168	-0,339	0,0158	-0,096	0,09737	-0,138	0,0842	0,16358	-0,0748	0,40275
3	-2,2317	-0,791	0,7731	-0,821	0,30994	-0,01	0,2629	0,01637	-0,1735	0,1215
4	-0,0646	-1,528	-0,872	-0,142	-0,9796	0,4378	-04515	-0,521	0,506	0,05058
5	-2,09	-0,737	0,0669	-0,353	-0,4674	0,3041	0,1698	-0,0509	0,3571	-0,17354
…										
25	4,4033	0,3481	-0,502	-0,588	-0,4564	0,7733	0,1395	-0,3168	-0,2588	0,16559
26	3,7241	1,1859	1,105	0,1767	-0,3098	-0,959	0,4567	0,44799	0,2654	0,0413
27	0,0313	1,6146	-0,657	0,0318	-0,2312	-0,784	-0,0767	-0,2252	0,1575	-0,26304

Setelah dilakukan analisis dengan menggunakan metode PCA, maka akan diperoleh data baru yang akan digunakan untuk analisis selanjutnya, yaitu analisis K-means Cluster. Pertama, tentukan jumlah cluster, dalam hal ini sebanyak 3 cluster sehingga terdapat tiga buah centroid yang masing-masing centroid merupakan nilai dari tiap

variabel untuk Kabupaten Bogor, Kota Bekasi, dan Kabupaten Pangandaran. Selanjutnya, alokasikan semua objek ke dalam cluster terdekat berdasarkan jarak cluster sehingga diperoleh hasil sebagai berikut.

TABEL 4 . Iterasi ke-2

No	Kabupaten	Klaster 1	Klaster 2	Klaster 3	Jarak terdekat	Cluster
1	Bogor	2,898338	7,010958	5,585086	2,898338	1
2	Sukabumi	3,2926	5,599185	1,153898	1,153898	3
3	Cianjur	3,40026	5,832589	1,945524	1,945524	3
…	…	…				…
23	Bekasi	5,451089	2,329718	6,376792	2,329718	2
24	Depok	5,623908	1,687279	6,008351	1,687279	2
25	Cimahi	5,841892	3,811562	2,179145	2,179145	2
26	Tasik	4,474741	3,811562	2,179145	2,179145	3
27	Banjar	5,339552	4,621841	2,462364	2,462364	3

Berdasarkan hasil di atas diketahui bahwa, hasil dari tahapan pertama dan kedua tidak berubah, maka hasil sudah sesuai dengan pengelompokan Cluster. Berikut adalah hasil dari pengelompokan tersebut :

TABEL 5 . Anggota dari cluster yang terbentuk

Cluster 1	Terdiri dari 4 kabupaten/kota yaitu: Kabupaten Bogor, Kabupaten Bandung, Kabupaten Karawang, Kabupaten Bekasi.
Cluster 2	Terdiri dari 7 kabupaten/kota yaitu: Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kotas Cimahi.
Cluster 3	Terdiri dari 16 kabupaten/kota yaitu: Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Majalengka, Kabupaten Sumedang, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Purwakarta, Kabupaten Bandung Barat, Kabupaten Pangandaran, Kota Tasikmalaya, Kota Banjar

Setelah Cluster terbentuk maka langkah selanjutnya adalah memberi ciri spesifik untuk menggambarkan isi Cluster tersebut. Berikut adalah tabel final Cluster.

TABEL 5. Final Cluster Centers

Variabel	Klaster 1	Klaster 2	Klaster 3
pdrb	-0,0602	3,3933	-1,4695
Kepadatan penduduk	-2,7581	0,4860	0,4769
Jumlah penduduk miskin	-0,4239	0,2998	-0,0252
Rata-rata pengeluaran per kapita	0,2893	-0,2133	0,0210
Jumlah angkatan kerja	0,1715	0,0575	-0,0680
umur harapan hidup	-0,0455	0,0529	-0,0118
Rata-rata lama sekolah	-0,1081	0,0741	-0,0054
Angka harapan lama sekolah	-0,2431	0,0486	0,0395
Tingkat pengangguran terbuka	0,0794	0,0229	-0,0299
Kepemilikan rumah sendiri	-0,0204	0,0116	0,0000

Berdasarkan tabel di atas diperoleh bahwa: 1) Cluster 1 beranggotakan 4 kabupaten dimana pada Cluster pertama memiliki nilai variabel kepadatan penduduk (X_2) dan jumlah penduduk miskin (X_3) paling rendah diantara Cluster lain artinya, tingkat kepadatan penduduk dan kemiskinan di daerah pada Cluster 1 paling rendah. Pada variabel PDRB (X_1), umur harapan hidup (X_6), rata-rata lama sekolah (X_7), angka harapan lama sekolah (X_8) dan kepemilikan rumah sendiri berada dibawah rata-rata, sedangkan pada variabel rata-rata pengeluaran per kapita (X_4) dan jumlah angkatan kerja (X_5) memiliki nilai tertinggi; 2) Cluster 2 beranggotakan 7 kabupaten/kota dimana pada Cluster ini memiliki nilai variabel kepadatan penduduk (X_2) dan jumlah penduduk miskin (X_3) paling tinggi di antara Cluster lain artinya, tingkat kepadatan penduduk dan kemiskinan di daerah pada Cluster 2 paling tinggi. Pada variabel rata-rata pengeluaran per kapita (X_4) memiliki nilai paling rendah di antara Cluster lain, sedangkan variabel lainnya PDRB (X_1), jumlah angkatan kerja (X_5), umur harapan hidup (X_6), rata-rata lama sekolah (X_7), angka harapan lama sekolah (X_8), tingkat pengangguran terbuka(X_9) dan kepemilikan rumah sendiri (X_{10}) memiliki nilai di atas rata-rata; 3) Cluster 3 beranggotakan 16 kabupaten/kota dimana pada Cluster ini memiliki nilai variabel PDRB (X_1), jumlah penduduk miskin ($X3$), jumlah angkatan kerja (X_5), umur harapan hidup ($X6$), rata-rata lama sekolah ($X7$), tingkat pengangguran terbuka (X_9) dan kepemilikan rumah sendiri (X_{10}) di bawah rata-rata sedangkan variabel kepadatan penduduk(X_2), rata-rata pengeluaran per kapita(X_4), angka harapan lama sekolah (X_8) memiliki nilai di atas rata-rata.

Variabel pembeda berdasarkan Uji F diperoleh bahwa hanya variabel X_1 dan X_2 yang merupakan variabel pembeda dalam pengklasifikasian.

SIMPULAN

Berdasarkan uraian dan perhitungan yang telah dilakukan sebelumnya, maka dapat diambil kesimpulan dari hasil pengelompokan kabupaten/kota di Provinsi Jawa Barat berdasarkan Indikator Kesejahteraan Rakyat menggunakan metode K-means Cluster yaitu sebagai berikut: 1). Berdasarkan hasil uji VIF (Variance Inflation Factor) di atas, dapat diketahui bahwa terdapat beberapa variabel yang memiliki nilai VIF lebih dari 10 yaitu, Z3,Z4 dan Z5. Maka dapat di asumsikan bahwa variabel tersebut mengalami multikolinearitas. Dalam mengatasi masalah multikolinearitas dalam penelitian ini dilakukan dengan menggunakan metode Principal Componen Analysis (PCA); 2). Penentuan jumlah Cluster dalam penelitian ini ditentukan terlebih dahulu sebagai bagian dari prosedur pengelompokan non-hierarki yaitu 3 Cluster; 3). Cluster yang terbentuk dengan metode K-means diperoleh hasil Cluster yang pertama yaitu 4 kabupaten/kota, Cluster 2 yaitu 7 kabupaten/kota dan Cluster 3 yaitu 16 kabupaten/kota. Dari Cluster yang terbentuk diperoleh urutan kelompok kabupaten/kota dengan tingkat kesejahteraan yaitu sejahtera , cukup sejahtera dan prasejahtera berturut-turut adalah Cluster 1, Cluster 2, Cluster 3.

UCAPAN TERIMA KASIH

Terima kasih penulis sampaikan kepada Pak Fadli Azis, M.Mat selaku Dewan Redaksi Jurnal Riset Matematika dan Sains Terapan (JRMST) yang telah sabar hingga penulis dapat menyelesaikan artikel ini hingga terbit.

REFERENSI

1. A. Yasid, "Implementasi Automatic Clustering Menggunakan Differential Evolution Dan CS Measure untuk Analisis Data Kemahasiswaan," *Jurnal Ilmiah NERO*, vol. 1, No. 2, 2014.
2. A Setiadi, E Delima, Sikumbang, "K-means Clustering Dalam Penerimaan Karyawan Baru" *Jurnal Informatics For Educators And Profesionals*, vol. 4, No. 2, 103-112, E-ISSN:2548-3412, STMIK Nusa Mandiri, 2020.
3. Badan Pusat Statistik, "Indikator Kesejahteraan Rakyat". Badan Pusat Statistik Povinsi Jawa Barat, 2020.
4. Badan Pusat Statistik, "Jawa Barat Dalam Angka". Badan Pusat Statistik Povinsi Jawa Barat. 2021.
5. D. William R, G. Matthew, Multivariate Analysis: Method And Application, John Wiley & Sons Inc., Canada, 1984.
6. Ediyanto, dkk, "Pengklasifikasian Karakteristik Dengan Metode K-means Cluster Analysis", Buletin Ilmiah Mat. Stat dan Terapannya (Bimaster), vol. 02, No. 2, hal 133-136, 2013.
7. Hair, dkk, Multivariate Data Analysis Pearson International Edition Edition 6. New Jersey, 2006.
8. Johnson, R. A., dan Wichern, D. W., Applied Multivariate Statistical Analysis 6th edition, Pearson Education Inc. United States of America, 2007.
9. M. W. Talakua, Z. A. Leleury, A. W. Talluta, "Analisis Cluster dengan Menggunakan Metode K-means untuk Pengelompokan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014" *Jurnal Ilmu Matematika Dan Terapan* vol. 1, No. 2, 2014.
10. M. Goreti, Yuki, Novia, S. Wahyuningsih. "Perbandingan Hasil Analisis Cluster Dengan Menggunakan Metode Single Linkage Dan Motode C-Means", *Jurnal Eksponensial*, vol.7, no.2, ISSN 2085-7829, 2016.
11. Nasari, F., & Sianturi, C. J. M, Penerapan Algoritma K-means Clustering untuk Pengelompokkan Penyebaran Diare di Kabupaten Langkat. CogITo Smart Journal, 2(2), 108. https://doi.org/10.31154/cogito.v2i2.19.108-119, 2016.
12. Nengsi, Anggraini; Jasmin; Pareza Alam, Jusia, "Penerapan Metode K-means Clustering Untuk Menentukan Persediaan Stok Barang Pada Toko Pensmart Jambi" *Jurnal Ilmiah Mahasiswa Teknik Informatika*, vol. 1, No. 1, STIKOM Dinamika Bangsa, Jambi, 2019.
13. Portal Hukum Dan Peraturan Indonesia, "Kesejahteraan ", https://pararegal.id/pengertian/kesejahteraan/, 2008.
14. Prasetyo, E., Data Mining Konsep dan Aplikasi Menggunakan Matlab, Andi Offset, Yogyakarta, 2012.
15. R. Herawaty, B. Bangun, "Analisis Klaster Non-Hierarki dalam Pengelompokan Kabupaten/Kota di Sumatra Utara Berdasarkan Faktor Produksi Padi", ISSN: 1979-8164, Agrica, 2016.
16. S. Yulianto, K. Hilya, Hidayatullah, "Analisis Klaster Untuk Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Rakyat", Jurnal *Statistika*, vol. 2, No. 1, Semarang, 2014.
17. S. Singgih, Mengatasi Berbagai Masalah Statistik dengan SPSS Versi 11.5. Jakarta: Elex Media Komputindo, 2014.
18. Sartono dkk, Analisis Peubah Ganda, Jurusan Statistika FMIPA IPB, Bogor, 2003.
19. Simamora, B., Analisis Multivariat Pemasaran, Jakarta: Gramedia Pustaka Utama 41, 2005.
20. S. Machfudhoh, N. Wahyuningsih, "Analisis Cluster Kabupaten/Kota Berdasarkan Pertumbuhan Ekonomi Jawa Timur", *Jurnal Sains Dan Seni POMITS*, vol. 2, No. 1, ITS, 2013.
21. Sudjana, Metode Statistika, Bandung: Tarsito, 2005.
22. Supranto, J, Analisis Multivariat: Arti dan interpretasi, Jakarta, PT. Rineka Cipta, 2004.
23. W. Alwi; M. Hasrul, "Analisis Klaster Untuk Pengelompokan Kabupaten/Kota Di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat", *Jurnal MSA*, vol. 6 No. 1 ED. Jan-Juni 2018.