

KESEJAHTERAAN RAKYAT INDONESIA BERDASARKAN PROVINSI MENGUNAKAN ANALISIS FAKTOR DAN FUZZY C-MEAN CLUSTERING

Disusun oleh :

Akbar Arta Putra

Muhammad Naufal Agisya

I. PENDAHULUAN

Indonesia Emas 2045 merupakan gagasan bangsa yang ingin dicapai, dimana pada tahun 2045 Indonesia memasuki masa emasnya yang berusia 100 tahun. Gagasan ini dipersiapkan dimana nantinya Indonesia pada masa 2045 memiliki sumber daya unggul, beretika, dan bebas dari korupsi. Selain itu, sesuai dengan Visi Indonesia 2045, Indonesia menjadi pusat pendidikan dan teknologi, negara yang mandiri dan paling berpengaruh di Asia Pasifik, dan barometer pertumbuhan ekonomi dunia [1] .

Indonesia harus mempersiapkan sumber daya tersebut sedari sekarang karena diprediksi oleh BPS pada tahun 2030-2040 Indonesia mendapatkan bonus demografi yang artinya sekitar 70 persen penduduk di Indonesia dalam usia yang produktif dimana nantinya mereka yang akan menjadi peran utama untuk negara ini. Namun, untuk mempersiapkan hal tersebut, kesejahteraan negara mengambil peran yang cukup besar karena kesejahteraan dapat menjadi salah satu indikator dalam pembangunan suatu negara [2]

Kesejahteraan merupakan salah satu hal yang sangat ingin dicapai baik setiap individu maupun negara. Kesejahteraan menurut UU No. 11 Tahun 2009, kesejahteraan masyarakat merupakan kondisi terpenuhinya kebutuhan baik dari material, spiritual, dan sosial warga negara agar dapat hidup layak dan mampu mengembangkan diri sehingga dapat melaksanakan fungsi sosialnya. Menurut Bappenas, rumah tangga dikategorikan sejahtera jika proporsi pengeluaran kebutuhan pokok setara atau lebih rendah dari proporsi kebutuhan non-pokok, vice versa. Menurut Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K), indikator-indikator yang menjadi tolak ukur kesejahteraan adalah pendidikan, perumahan, ketenagakerjaan [3].

Untuk itu, jika Indonesia mengusung gagasan Indonesia Emas 2045, hal tersebut perlu diperhatikan kembali apakah kesejahteraan Indonesia sudah menyebar dan merata ke seluruh provinsi yang ada di Indonesia atau belum dan faktor-faktor apa saja yang berpengaruh terhadap kesejahteraan masyarakat. Hal tersebut perlu dikaji kembali agar Indonesia dapat

memberikan perhatian yang lebih kepada provinsi-provinsi yang belum menerima kesejahteraan yang setara dengan provinsi lainnya. Dengan demikian, penelitian ini ingin mengkaji provinsi-provinsi mana saja yang tingkat kesejahteraannya memiliki kemiripan dimana nantinya dapat mempermudah atau dapat menjadi rekomendasi kepada pemerintah untuk memberikan suatu kebijakan dari klaster yang telah terbentuk melalui faktor-faktor yang berpengaruh signifikan sehingga Indonesia sejahtera secara menyebar dan merata dan pada akhirnya siap untuk menjadi Indonesia Emas 2045.

II. METODE

A. Data

Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistika (BPS) pada tahun 2021. Data tersebut memiliki sebanyak 9 variabel dimana satu variabel berbentuk kategorik yaitu provinsi dan delapan lainnya berbentuk numerik yang terdiri atas PDRB per kapita, jumlah penduduk miskin, kepadatan penduduk, pengeluaran per kapita yang disesuaikan, angka harapan hidup, rata-rata lama sekolah, gini ratio, dan tingkat pengangguran terbuka. Data tersebut memiliki 34 observasi dimana observasinya adalah provinsi yang ada di Indonesia.

B. Eksplorasi Data dan Preprocessing Data

Sebelum melakukan analisis data, perlu dilakukan eksplorasi data dan preprocessing data. Data preprocessing adalah proses yang mengubah data mentah ke dalam bentuk yang lebih mudah dipahami. Proses ini penting dilakukan karena data mentah sering kali tidak memiliki format yang teratur. Pada proses ini dilakukan eksplorasi data untuk melihat karakteristik data tersebut melalui statistika deskriptif. Selanjutnya, setelah dilakukan eksplorasi data, akan dilakukan data cleaning dimana proses ini untuk memilah kembali data yang telah diperoleh. Pada tahap ini akan diseleksi kembali data yang sekiranya kurang relevan atau tidak akurat dengan cara menghapus atau menghilangkan data tersebut. Selain data cleaning, juga dilakukan transformasi data yang digunakan untuk mengubah dimensi data yang dimiliki dikarenakan data yang dimiliki tidak memiliki dimensi yang sama.

C. Analisis Faktor

Analisis faktor (factor analysis) merupakan suatu metode untuk mereduksi variabel pengamatan secara keseluruhan menjadi beberapa variabel atau dimensi. Dimensi baru tersebut harus mampu merepresentasikan variabel-variabel utamanya. Analisis faktor merupakan

perluasan dari analisis komponen utama. Dalam analisis faktor ini, tidak dibedakan variabel dependen maupun independennya. Proses analisis faktor sendiri mencoba menemukan hubungan (interrelationship) antar sejumlah variabel-variabel yang saling dependen dengan yang lain sehingga bisa dibuat satu atau beberapa kumpulan variabel yang lebih sedikit dari jumlah awal.

D. Fuzzy K-Means Clustering

Fuzzy C-Means adalah suatu teknik multivariat yaitu teknik cluster data yang posisi setiap titik data dalam suatu clusternya ditentukan oleh derajat keanggotaan. Derajat keanggotaan tersebut merupakan bilangan real pada interval $[0,1]$. Semakin tinggi nilai dari keanggotaannya maka semakin besar juga kemiripannya dengan kelompok. Hal pertama yang dilakukan adalah menentukan pusat klaster, yang nantinya dapat menandai lokasi rata-rata untuk setiap klaster. Pusat klaster dan derajat keanggotaan setiap titik data dilakukan secara berulang dan diperbaiki sehingga nantinya dapat dilihat bahwa pusat klaster akan bergerak menuju ke lokasi yang tepat [4].

E. Kruskal Wallis Test

Uji kruskal-wallis akan digunakan untuk melihat signifikansi dari variabel-variabel yang digunakan pada setiap klasternya. Ketika suatu data dinyatakan signifikan, maka rata-rata dari variabel tersebut untuk masing-masing klaster berhubungan dengan kesejahteraan masyarakat.

III. HASIL

A. Eksplorasi Data

Eksplorasi data merupakan tahap untuk memahami karakteristik data sebelum dilakukannya data preprocessing. Pada tahap ini akan dilihat karakteristik data menggunakan statistik deskriptif.

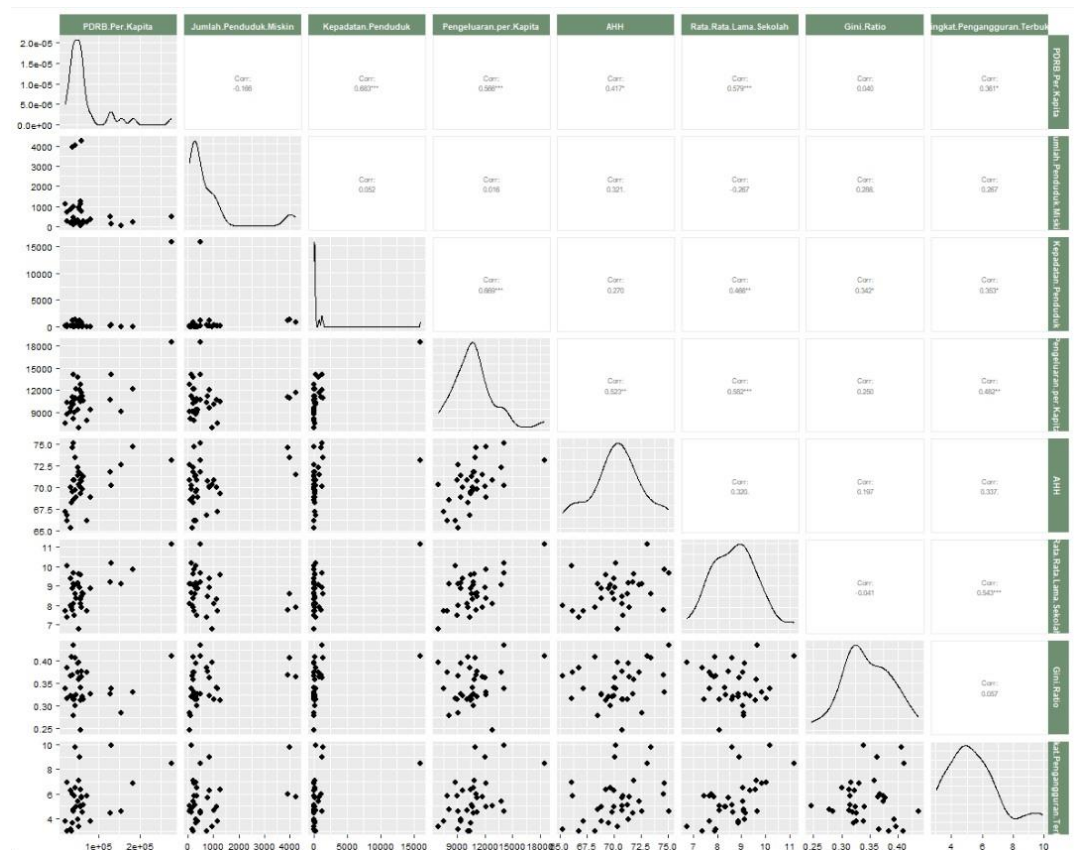
1. Gambaran Umum Variabel Penyusun Indeks Kesejahteraan Rakyat Indonesia

Tabel 1. Statistika Deskriptif

Variabel	Min	Kuartil 1	Median	Mean	Kuartil 3	Maks
PDRB Perkapita	20580	40297	53248	66891	62101	274519

Jumlah Penduduk Miskin	49.49	201.19	346.96	779.52	851.77	4259.60
Kepadatan Penduduk	9	54.25	103.50	744.26	261	15978
Pengeluaran Perkapita	6955	9379	10625	10759	11576	18520
Angka Harapan Hidup	65.29	68.98	70.24	70.33	71.65	75.08
Rata-Rata Lama Sekolah	6.760	7.990	8.755	8.716	9.175	11.170
<i>Gini Ratio</i>	0.2470	0.3170	0.3390	0.3458	0.3748	0.4360
Tingkat Pengangguran Terbuka	3.010	4.447	5.060	5.492	6.322	9.910

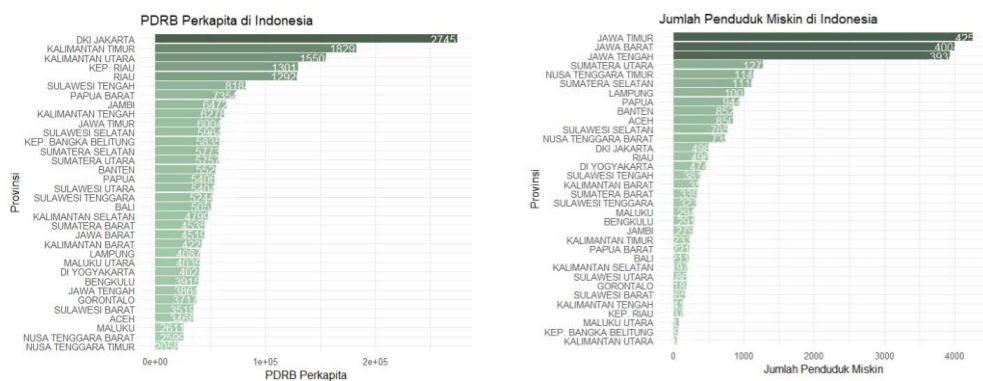
Tabel 1 atas merupakan tabel pemusatan data dari masing-masing variabel. Dari tabel tersebut, kita dapat melihat gambaran data secara sekilas. Contohnya pada tahun 2021 variabel PDRB Perkapita memiliki terdapat provinsi di Indonesia dengan nilai PDRB terkecil sebesar 20580 juta rupiah, sedangkan terdapat provinsi di Indonesia dengan nilai PDRB terbesar sebesar 274519 juta rupiah. Selain itu rata-rata provinsi di Indonesia pada tahun 2021 memiliki rata-rata PDRB Perkapita sebesar 66891 juta rupiah. 2. Melihat Visualisasi antar Beberapa variabel



Gambar 1. Visualisasi Variabel

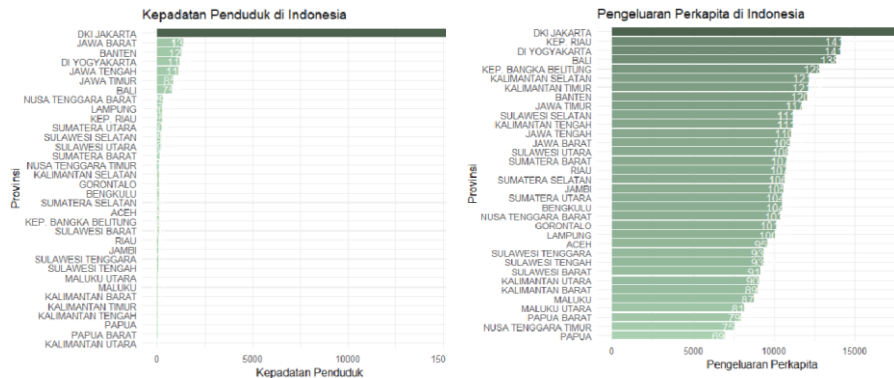
Gambar 1 atas merupakan kumpulan informasi dari data kesejahteraan rakyat Indonesia berdasarkan nilai korelasi dan scatter plot. Gambar ini dapat digunakan sebagai informasi sebaran data dan juga besarnya hubungan antarvariabel.

3. Melihat visualisasi antar beberapa variabel Gambaran variabel data kesejahteraan rakyat berdasarkan provinsi di Indonesia.



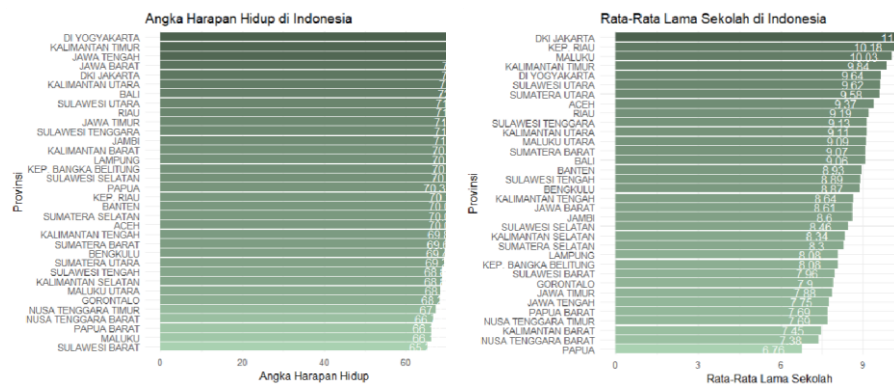
Gambar 2. Visualisasi variabel PDRB per Kapita dan Jumlah Penduduk Miskin di Indonesia berdasarkan provinsi

Berdasarkan **Gambar 2** di atas, PDRB Perkapita di Indonesia tahun 2021 paling tinggi adalah DKI Jakarta, sedangkan paling rendah adalah Nusa Tenggara Timur. Selanjutnya, jumlah penduduk miskin di Indonesia tahun 2021 paling banyak adalah Jawa Timur, sedangkan paling sedikit adalah Kalimantan Utara.



Gambar 3. Visualisasi variabel Kepadatan Penduduk dan Pengeluaran per Kapita di Indonesia berdasarkan provinsi

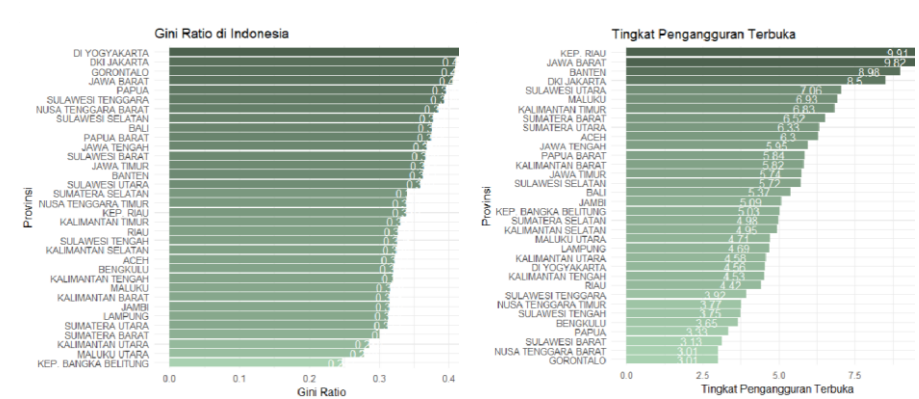
Berdasarkan **Gambar 3** di atas, kepadatan penduduk di Indonesia tahun 2021 paling padat adalah DKI Jakarta, sedangkan paling tidak padat adalah Kalimantan Utara. Selanjutnya, pengeluaran perkapita di Indonesia tahun 2021 paling banyak adalah DKI Jakarta, sedangkan paling sedikit adalah Papua.



Gambar 4. Visualisasi variabel Angka Harapan Hidup dan Rata-Rata Lama Sekolah di Indonesia berdasarkan provinsi

Berdasarkan **Gambar 4** di atas, angka harapan hidup di Indonesia tahun 2021 paling lama adalah DI Yogyakarta, sedangkan paling sebentar adalah Sulawesi

Barat. Selanjutnya, rata-rata lama sekolah di Indonesia tahun 2021 paling lama adalah DKI Jakarta, sedangkan paling sebentar adalah Papua.



Gambar 5. Visualisasi variabel *Gini Ratio* dan Tingkat Pengangguran Terbuka di Indonesia berdasarkan provinsi

Berdasarkan **Gambar 5** di atas *gini ratio* di Indonesia tahun 2021 paling tinggi adalah DI Yogyakarta, sedangkan yang paling rendah adalah Kepulauan Bangka Belitung. Selanjutnya, tingkat pengangguran terbuka di Indonesia tahun 2021 paling banyak adalah Kepulauan Riau, sedangkan paling sedikit adalah Gorontalo.

B. Data Cleaning

Data cleaning adalah suatu proses persiapan data untuk memilah kembali data yang telah diperoleh. Pada tahap ini akan diseleksi kembali data yang sekiranya kurang relevan atau tidak akurat dengan cara menghapus atau menghilangkan data tersebut.

Selanjutnya akan dilakukan pengecekan *missing value* pada data kesejahteraan rakyat Indonesia tahun 2021. Berikut hasil dari masing-masing variabel.

Tabel 2. Jumlah *Missing Value* per Variabel

Variabel	Jumlah <i>Missing Value</i>	Variabel	Jumlah <i>Missing Value</i>
PDRB Perkapita	0	Rata-rata lama sekolah	0
Jumlah Penduduk Miskin	0	Gini Ratio	0

Kepadatan Penduduk	0	Tingkat Pengangguran Terbuka	0
Pengeluaran Perkapita	0	Angka Harapan Hidup	1

Berdasarkan **Tabel 2**, terdapat missing value atau data hilang pada variabel angka harapan hidup. Untuk itu data hilang tersebut akan diisi oleh rata-rata dari variabel angka harapan hidup. Sehingga *missing value* dapat diatasi.

C. Transformasi Data

Transformasi data merupakan suatu proses untuk mengubah dimensi suatu data. Fungsi dari transformasi ini adalah untuk menyamakan satuan dari masing-masing variabel, sehingga hal ini akan mempermudah proses analisis data. Pada tahap ini transformasi yang digunakan adalah mengubah bentuk data ke dalam bentuk standar atau *z-score*.

D. Analisis Faktor

Untuk melakukan analisis faktor diperlukan beberapa pengujian asumsi. Hal itu diperlukan untuk memutuskan apakah data memenuhi syarat untuk dilakukan analisis faktor,

1. Pendeteksian Multikolinieritas

Pendeteksian multikolinieritas menggunakan nilai VIF, jika nilai VIF lebih besar dari 10, maka terindikasi adanya gejala multikolinieritas pada data. Berikut nilai VIF pada masing-masing variabel.

Tabel 3. Nilai VIF per Variabel

Variabel	Nilai VIF	Variabel	Nilai VIF
PDRB Perkapita	2.917188	Rata-rata lama sekolah	2.613293
Jumlah Penduduk Miskin	2.091005	Gini Ratio	1.364503

Kepadatan Penduduk	3.317038	Tingkat Pengangguran Terbuka	2.036210
Pengeluaran Perkapita	2.878454	Angka Harapan Hidup	2.084751

Berdasarkan **Tabel 3** di atas dapat disimpulkan bahwa tidak adanya multikolinieritas pada data kesejahteraan rakyat Indonesia tahun 2021. 2. *Kaiser–Meyer–Olkin (KMO) - Measure of Sampling Adequacy(MSA)* Nilai KMO yang didapat adalah sebesar 0.6156487. Syarat nilai KMO untuk analisis faktor adalah lebih dari 0.5. Maka nilai tersebut memenuhi persyaratan. Selanjutnya nilai MSA dari masing-masing variabel.

	data_KMO\$MSAi
PDRB.Per.Kapita	0.6664956
Jumlah.Penduduk.Miskin	0.2829568
Kepadatan.Penduduk	0.5997201
Pengeluaran.per.Kapita	0.7552744
AHH	0.5206239
Rata.Rata.Lama.Sekolah	0.7202195
Gini.Ratio	0.5973474
Tingkat.Pengangguran.Terbuka	0.6231728

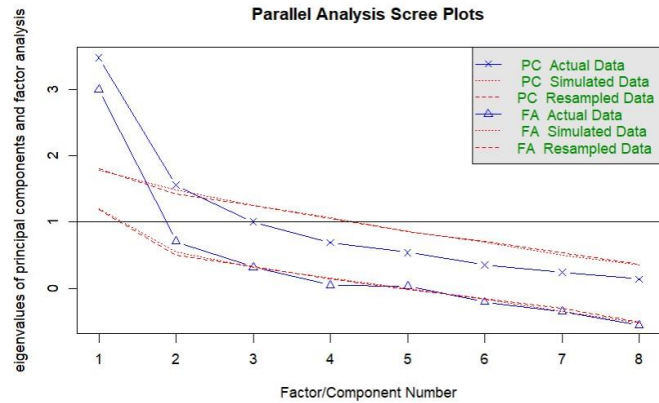
Gambar 6. Nilai MSA Masing-Masing Variabel

Nilai MSA pada masing-masing variabel lebih dari 0.5, maka dapat disimpulkan bahwa semua variabel dapat diprediksi dan dianalisis lebih lanjut.

2. *The Bartlett's Test Statistics*

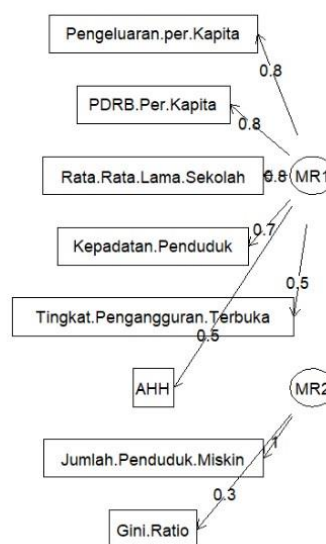
Bartlett test of homogeneity of variances memenuhi persyaratan karena signifikansi di bawah 0,05 (5%), yaitu 2.2e-16, sehingga variance tidak homogen.

Berikut hasil Faktor-Faktor Pembentuk kesejahteraan rakyat di Indonesia tahun 2021.



Gambar 7. Scree Plot

Factor Analysis



Gambar 8. Output Factor Analysis

Dengan menggunakan *parallel analysis* didapatkan bahwa faktor optimum adalah 2 faktor PCA. Dengan kombinasi faktor 1, yaitu pengeluaran perkapita, PDRB perkapita, rata-rata lama sekolah, kepadatan penduduk, tingkat pengangguran terbuka, angka harapan hidup, sedangkan faktor 2, yaitu jumlah penduduk miskin dan *gini ratio*.

Nilai PCA yang telah terbentuk akan dianalisis untuk dijadikan indeks kesejahteraan rakyat. Hasil indeks kesejahteraan rakyat tersebut akan dikategorikan menjadi tidak sejahtera, kurang sejahtera, cukup sejahtera, dan sejahtera. Berikut hasilnya

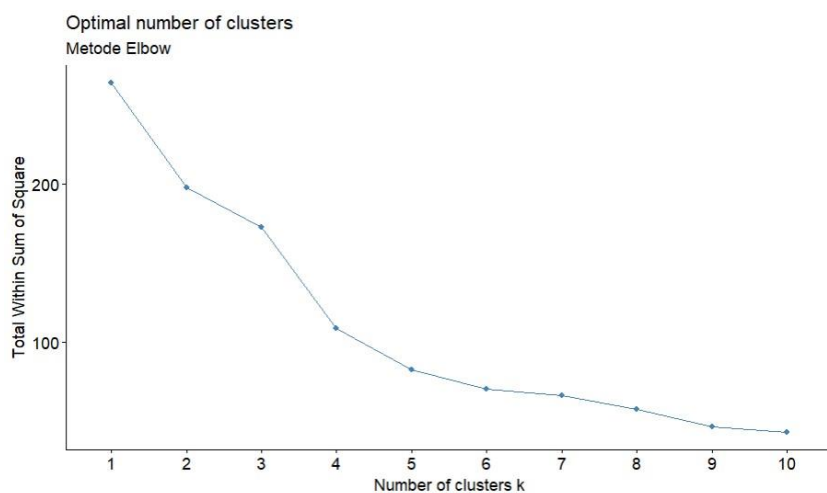
Provinsi	ikr	Jenks_ikr	Kategori
ACEH	0.3075500	1	Tidak Sejahtera
SUMATERA UTARA	0.2479787	1	Tidak Sejahtera
SUMATERA BARAT	0.5180711	1	Tidak Sejahtera
RIAU	0.6307084	1	Tidak Sejahtera
JAMBI	0.4068745	1	Tidak Sejahtera
SUMATERA SELATAN	0.4083343	1	Tidak Sejahtera
BENGKULU	0.8212185	1	Tidak Sejahtera
LAMPUNG	0.5375812	1	Tidak Sejahtera
KEP. BANGKA BELITUNG	0.8392327	1	Tidak Sejahtera
KEP. RIAU	1.8649884	2	Kurang Sejahtera
DKI JAKARTA	4.4692076	4	Sejahtera
JAWA BARAT	2.4608032	3	Cukup Sejahterah
JAWA TENGAH	1.7482598	2	Kurang Sejahtera
DI YOGYAKARTA	1.4636815	2	Kurang Sejahtera
JAWA TIMUR	1.5042918	2	Kurang Sejahtera
BANTEN	0.7273559	1	Tidak Sejahtera
BALI	0.5911393	1	Tidak Sejahtera
NUSA TENGGARA BARAT	1.2582273	2	Kurang Sejahtera
NUSA TENGGARA TIMUR	1.3634987	2	Kurang Sejahtera
KALIMANTAN BARAT	0.6400143	1	Tidak Sejahtera
KALIMANTAN TENGAH	0.6136779	1	Tidak Sejahtera
KALIMANTAN SELATAN	0.5575536	1	Tidak Sejahtera
KALIMANTAN TIMUR	1.6732381	2	Kurang Sejahtera
KALIMANTAN UTARA	0.9007432	1	Tidak Sejahtera
SULAWESI UTARA	0.5144364	1	Tidak Sejahtera
SULAWESI TENGAH	0.8538766	1	Tidak Sejahtera
SULAWESI SELATAN	0.2786196	1	Tidak Sejahtera
SULAWESI TENGGARA	0.2842906	1	Tidak Sejahtera
GORONTALO	0.9279715	1	Tidak Sejahtera
SULAWESI BARAT	1.3369516	2	Kurang Sejahtera
MALUKU	1.0402862	2	Kurang Sejahtera
MALUKU UTARA	1.4527901	2	Kurang Sejahtera
PAPUA	0.8753134	1	Tidak Sejahtera
PAPUA	1.7392722	2	Kurang Sejahtera

Gambar 9. Nilai Indeks Kesejahteraan Masyarakat per Provinsi

E. Fuzzy C-Means Clustering

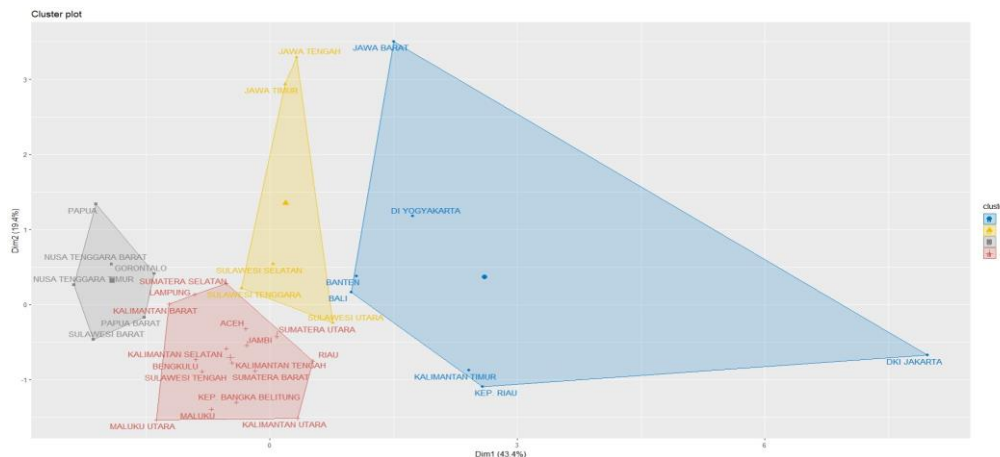
Proses klasterisasi menggunakan metode *Fuzzy C-Means* diawali dengan melihat klaster optimumnya terlebih dahulu dengan menggunakan

Metode Elbow. Berikut grafik Metode Elbow



Gambar 10. Plot Metode Elbow

Berdasarkan pada **Gambar 10**, dapat dilihat bahwa nilai k optimum, yaitu pada saat nilainya 4. Selanjutnya prosesklasterisasi dengan metode *Fuzzy C-Means*. Berikut hasil visualisasi cluster dengan metode *Fuzzy C-Means*.



Gambar 11. Visualisasi Klaster

Berdasarkan hasil yang diperoleh dapat dilihat 4 klaster yang telah terbentuk.

1. Klaster 1, yaitu Jawa Barat, DI Yogyakarta, Banten, Bali, Kalimantan Timur, Kepulauan Riau, dan DKI Jakarta.
2. Klaster 2, yaitu Jawa Tengah, Jawa Timur, Sulawesi Selatan, Sulawesi Tenggara, dan Sulawesi Utara.
3. Klaster3, yaitu Papua, Nusa Tenggara Barat, Gorontalo, Nusa Tenggara Timur, Papua Barat, dan Sulawesi Barat.
4. Klaster 4, yaitu Sumatera Selatan, Lampung, Kalimantan Barat, Aceh, Jambi, Sumatera Utara, Kalimantan Selatan, Bengkulu, Kalimantan Tengah, Sulawesi Tengah, Sumatera Barat, Kepulauan Bangka Belitung, Maluku, Maluku Utara, Riau, dan Kalimantan Utara.

F. Uji Bartlett terhadap klaster yang terbentuk

Berdasarkan *output* melalui Uji *Kruskal-Wallis*, dapat dilihat bahwa pengujian signifikan ($p\text{-value} = 5.66e-14 < 0,05$). Hal ini menunjukkan hipotesis nol ditolak yang menunjukkan terdapat perbedaan rata-rata antar kluster. Masing-masing kluster memiliki variabel yang dominan dalam mempengaruhi kesejahteraan rakyat.

IV. KESIMPULAN

Indonesia Emas 2045 merupakan gagasan bangsa yang ingin dicapai, dimana pada tahun 2045 Indonesia memasuki masa emasnya yang berusia 100 tahun. Indonesia harus mempersiapkan sumber daya sedari sekarang karena diprediksi oleh BPS pada tahun 2030-2040 Indonesia mendapatkan bonus demografi. Namun, untuk mempersiapkan hal tersebut,

kesejahteraan negara mengambil peran yang cukup besar karena kesejahteraan dapat menjadi salah satu indikator dalam pembangunan suatu negara. Perlu diperhatikan kembali apakah kesejahteraan Indonesia sudah menyebar dan merata ke seluruh provinsi yang ada di Indonesia atau belum dan faktor-faktor apa saja yang berpengaruh terhadap kesejahteraan masyarakat. Hal tersebut perlu dikaji kembali agar Indonesia terutama pemerintah dapat memberikan perhatian yang lebih kepada provinsi-provinsi yang belum menerima kesejahteraan yang setara dengan provinsi lainnya.

Berdasarkan analisis yang telah dilakukan diperoleh kesimpulan bahwa terdapat 2 faktor yang berpengaruh terhadap kesejahteraan rakyat Indonesia dengan kombinasi dari faktor-faktor tersebut adalah faktor 1, yaitu pengeluaran perkapita, PDRB perkapita, rata-rata lama sekolah, kepadatan penduduk, tingkat pengangguran terbuka, angka harapan hidup, sedangkan faktor 2, yaitu jumlah penduduk miskin dan gini ratio.

Selanjutnya, berdasarkan klaster yang telah terbentuk berdasarkan tingkat kemiripannya diharapkan dapat mempermudah atau menjadi rekomendasi kepada pemerintah untuk memberikan suatu kebijakan dari klaster yang telah terbentuk, agar setiap provinsi memiliki kesejahteraan yang merata.

REFERENSI

- [1] Ringkasan Eksekutif Visi Indonesia 2045, 2019 diakses pada https://perpustakaan.bappenas.go.id/e-library/file_upload/koleksi/migrasi-data-publik/asi/file/Policy_Paper/Ringkasan%20Eksekutif%20Visi%20Indonesia%202045_Final.pdf
- [2] Dwitianti, N., Selvia, N., & Andrari, F. R. (2019). Penerapan fuzzy c-means cluster Dalam Pengelompokkan Provinsi Indonesia menurut indikator Kesejahteraan Rakyat. Faktor Exacta, 12(3), 201. <https://doi.org/10.30998/faktorexacta.v12i3.4526>
- [3] Gustyarini, NID. 2019. ANALISIS FAKTOR YANG MEMPENGARUHI KESEJAHTERAAN MASYARAKAT DI PROVINSI JAWA TENGAH. Skripsi. Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang.
- [4] Siringoringo, R, dkk (2018). PENINGKATAN PERFORMA CLUSTER FUZZY C-MEANS PADA PENGKLASTERAN SENTIMEN MENGGUNAKAN PARTICLE SWARM OPTIMIZATION. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK), 6(4), 349-354. <https://doi.org/10.30998/faktorexacta.v12i3.4526>
<https://rpubs.com/ranydc/kerentanansosialCov-19DKI>