
Cluster Analysis and Discriminant Analysis for Grouping Provinces Based on Factors Affecting Poverty Levels in Indonesia 2018-2020

Kariyam^{1*}, Baiq Jasmin Sabhira Safwa V.R², Juan Latif Alifia³, Larasati Oktarani⁴, Putri Pratista Andanitya⁵, Willia Diva Ikhsani⁶

^{1,2} Statistika, Universitas Islam Indonesia, Yogyakarta, Indonesia,

¹ kariyam@uii.ac.id, ² 23611003@students.uui.ac.id, ³ 23611020@students.uui.ac.id,

⁴ 23611022@students.uui.ac.id, ⁵ 23611073@students.uui.ac.id, ⁶ 23611076@students.uui.ac.id

*Corresponding author

A B S T R A K

Kemiskinan merupakan kondisi yang terjadi akibat ketidakmampuan seseorang atau kelompok dalam memenuhi kebutuhan pokok minimum, seperti pangan, sandang, kesehatan, perumahan, dan pendidikan, yang diperlukan untuk mempertahankan kelangsungan hidup. Tingkat kemiskinan suatu daerah dipengaruhi oleh berbagai faktor, di antaranya Tingkat Pengangguran Terbuka (TPT), Upah Minimum Provinsi (UMP), dan Indeks Pembangunan Manusia (IPM). Penelitian ini bertujuan untuk mengelompokkan provinsi di Indonesia berdasarkan faktor-faktor yang memengaruhi kemiskinan serta menentukan fungsi diskriminan dari kelompok yang terbentuk. Metode analisis yang digunakan adalah analisis *cluster* untuk mengelompokkan provinsi ke dalam beberapa kelompok tingkat kemiskinan dan analisis diskriminan untuk membentuk fungsi pemisah antar kelompok tersebut. Hasil analisis *cluster* menunjukkan terbentuknya tiga kelompok, yaitu kelompok dengan tingkat kemiskinan tertinggi yang terdiri atas 7 provinsi, kelompok dengan tingkat kemiskinan sedang yang terdiri atas 8 provinsi, dan kelompok dengan tingkat kemiskinan terendah yang mencakup provinsi lainnya. Selanjutnya, analisis diskriminan menghasilkan fungsi diskriminan yang mampu membedakan antar kelompok tingkat kemiskinan secara cukup baik. Hasil penelitian ini diharapkan dapat menjadi bahan pertimbangan bagi pemerintah dalam perumusan kebijakan penanggulangan kemiskinan yang lebih tepat sasaran.

Kata Kunci:

kemiskinan, Tingkat Pengangguran Terbuka, Upah Minimum Provinsi, Indeks Pembangunan Manusia, analisis *cluster*, analisis diskriminan.

A B S T R A C T

Poverty is a condition that occurs due to the inability of a person or group to meet the minimum basic needs, such as food, clothing, health, housing, and education, which are necessary to maintain survival. The poverty level of an area is influenced by various factors, including the Open Unemployment Rate (TPT), the Provincial Minimum Wage (UMP), and the Human Development Index (IPM). This research aims to group provinces in Indonesia based on factors that affect poverty and determine the discriminatory function of the group formed. The analysis method used is cluster analysis to group provinces into several poverty level groups and discriminatory analysis to form a separating function between the groups. The results of cluster analysis show the formation of three groups, namely the group with the highest poverty level consisting of 7 provinces, the group with moderate poverty level consisting of 8 provinces, and the group with the lowest poverty level which includes other provinces. Furthermore, discriminant analysis produces a discriminant function that is able to distinguish between poverty levels quite well. The results of this research are expected to be considered by the government in formulating poverty alleviation policies that are more on target.

Keywords: Poverty, Open Unemployment Rate, Provincial Minimum Wage, Human Development Index, Cluster Analysis, Discriminatory Analysis

1. PENDAHULUAN

Kemiskinan merupakan kondisi ketidakmampuan individu atau rumah tangga dalam memenuhi kebutuhan dasar minimum yang meliputi kebutuhan pangan, sandang, perumahan, pendidikan, dan kesehatan yang layak. Badan Pusat Statistik (BPS) mendefinisikan kemiskinan menggunakan pendekatan kebutuhan dasar (*basic needs approach*), yaitu dengan membandingkan pengeluaran penduduk terhadap garis kemiskinan yang ditetapkan berdasarkan kebutuhan minimum makanan dan nonmakanan. Pendekatan ini digunakan karena mampu menggambarkan kondisi kesejahteraan penduduk secara lebih menyeluruh dan objektif.

Berdasarkan pendekatan kebutuhan dasar tersebut, BPS menggunakan tiga indikator utama dalam mengukur tingkat kemiskinan, yaitu Indeks Kemiskinan (*Head Count Index*) yang menunjukkan persentase penduduk miskin, Indeks Kedalaman Kemiskinan (*Poverty Gap Index*) yang menggambarkan rata-rata kesenjangan pengeluaran penduduk miskin terhadap garis kemiskinan, serta Indeks Keparahan Kemiskinan (*Poverty Severity Index*) yang mencerminkan tingkat ketimpangan pengeluaran di antara penduduk miskin. Di antara ketiga indikator tersebut, *Head Count Index* merupakan ukuran yang paling sederhana dan paling umum digunakan karena secara langsung menunjukkan proporsi penduduk miskin di suatu wilayah (BPS, 2025).

Salah satu faktor ekonomi yang berperan penting dalam memengaruhi tingkat kemiskinan adalah upah minimum. Upah minimum merupakan standar upah terendah yang ditetapkan pemerintah sebagai jaring pengaman bagi pekerja untuk memenuhi kebutuhan hidup layak. Badan Pusat Statistik mencatat bahwa Upah Minimum Regional (UMR) atau Upah Minimum Provinsi (UMP) ditetapkan berbeda-beda di setiap provinsi berdasarkan kondisi ekonomi, inflasi, dan tingkat kebutuhan hidup. Perbedaan UMR antarprovinsi mencerminkan ketimpangan kondisi ekonomi wilayah yang berpotensi memengaruhi kesejahteraan masyarakat (BPS, 2017).

Berdasarkan latar belakang tersebut, penelitian ini menggunakan data Upah Minimum Regional (UMR) provinsi di Indonesia yang bersumber dari Badan Pusat Statistik untuk mengelompokkan provinsi berdasarkan karakteristik ekonomi yang dimiliki. Metode analisis yang digunakan adalah analisis *cluster* untuk membentuk kelompok provinsi yang memiliki kemiripan karakteristik, serta analisis diskriminan untuk menguji dan mengklasifikasikan provinsi ke dalam kelompok yang terbentuk serta mengetahui variabel yang paling berperan dalam membedakan kelompok tersebut. (BPS, 2017).

Sebelum dilakukan analisis *cluster* dan analisis diskriminan, terlebih dahulu dilakukan pengujian asumsi terhadap variabel yang digunakan, yaitu uji kenormalan dan uji multikolinearitas. Uji kenormalan dilakukan untuk memastikan bahwa data berdistribusi normal, yang dapat diuji melalui *normal probability plot*, uji *Anderson–Darling*, dan uji *Kolmogorov–Smirnov* dengan kriteria nilai signifikansi lebih besar dari 0,05. Sementara itu, uji multikolinearitas dilakukan untuk memastikan tidak terdapat hubungan yang kuat antarvariabel bebas, yang dapat dideteksi melalui nilai koefisien korelasi, nilai toleransi, dan *Variance Inflation Factor* (VIF) (Ansari , Muhammad Novi, 2020).

Analisis *cluster* bertujuan untuk mengelompokkan objek atau wilayah ke dalam dua atau lebih kelompok berdasarkan kesamaan karakteristik yang dimiliki. Dalam penelitian ini, ukuran jarak yang digunakan adalah jarak Euclidean karena merupakan ukuran jarak yang paling umum dan mudah diinterpretasikan dalam analisis pengelompokan. Selanjutnya, analisis diskriminan digunakan untuk mengevaluasi hasil pengelompokan tersebut serta membangun fungsi diskriminan yang dapat menjelaskan perbedaan antar kelompok provinsi berdasarkan variabel yang digunakan (Hidayat , 2024).

Kebaruan penelitian ini terletak pada fokus variabel, periode data, serta pendekatan analisis yang digunakan dibandingkan dengan penelitian-penelitian sebelumnya. Berbeda dengan studi Widiyatna dan Rahardjo yang menganalisis kemiskinan provinsi di Indonesia menggunakan kombinasi Tingkat Pengangguran Terbuka (TPT), Upah Minimum Provinsi (UMP), dan Indeks Pembangunan Manusia (IPM) pada periode 2009–2011, penelitian ini secara khusus menitikberatkan pada peran Upah Minimum Regional (UMR) sebagai variabel utama dalam pengelompokan provinsi berdasarkan karakteristik ekonomi yang berkaitan dengan kemiskinan.

Selain itu, penelitian ini menggunakan data UMR yang lebih mutakhir dari Badan Pusat Statistik, sehingga hasil pengelompokan yang diperoleh diharapkan mampu merepresentasikan kondisi ekonomi dan kesejahteraan provinsi di Indonesia secara lebih relevan dengan situasi terkini. Pendekatan ini menjadi penting mengingat kebijakan upah minimum terus mengalami penyesuaian sebagai respons terhadap inflasi, pertumbuhan ekonomi, dan dinamika pasar tenaga kerja.

Dari sisi metodologi, penelitian ini tidak hanya membentuk kelompok provinsi melalui analisis cluster, tetapi juga menggunakan analisis diskriminan secara sistematis sebagai alat validasi hasil pengelompokan, termasuk pengujian tingkat ketepatan klasifikasi menggunakan hit ratio dan *Press's Q Statistic*. Dengan demikian, penelitian ini tidak sekadar bersifat deskriptif dalam pengelompokan wilayah, melainkan juga memberikan pengujian statistik yang kuat terhadap keakuratan struktur kelompok yang terbentuk.

Dengan kombinasi fokus pada satu indikator ekonomi utama (UMR), penggunaan data terbaru, serta validasi statistik yang komprehensif melalui analisis diskriminan, penelitian ini diharapkan dapat memberikan kontribusi baru dalam pemetaan karakteristik ekonomi provinsi di Indonesia, khususnya dalam konteks keterkaitannya dengan kemiskinan dan ketimpangan kesejahteraan antarwilayah.

2. METODE

Metode yang digunakan dalam penelitian ini adalah analisis *cluster* dengan metode *ward* dan analisis diskriminan. Variabel yang digunakan pada penelitian ini adalah :

1. X_1 = Tingkat Pengangguran Terbuka (TPT) dalam persen.
2. X_2 = Jumlah Upah Minimum Regional (UMR) dalam ribuan rupiah.
3. X_3 = Indeks Pembangunan Manusia (IPM) dalam persen.

Sampel data dalam penelitian ini adalah penduduk di Indonesia yang mencakup 24 provinsi pada tahun 2018, 2019, dan 2020. Dalam penelitian ini menggunakan teknik pengumpulan data dengan metode dokumentasi. Pengumpulan data diambil dari BPS pusat melalui [web www.bps.go.id](http://www.bps.go.id). Langkah-langkah yang dilakukan dalam menganalisis pada penelitian ini adalah:

1. Pengujian asumsi kenormalan menggunakan uji *Anderson Darling*.
2. Uji multikolinearitas, jika tidak terjadi multikolinearitas, maka dilakukan analisis *cluster* dengan metode *ward*.
3. Analisis *cluster* dengan metode *ward*
 - a. Menentukan jumlah kelompok.
 - b. Mengelompokkan anggota tiap kelompok.
 - c. Menentukan profil dari tiap kelompok yang terbentuk.
4. Pengujian Homogenitas, jika bersifat homogen maka dilakukan analisis diskriminan.
5. Analisis Diskriminan
 - a. Menentukan jumlah akar ciri, yaitu pangkat dari $W^{-1}B$ yang diketahui adalah minimum $(g - l, p)$
 - b. Mengui kesamaan vektor nilai rata-rata dengan statistik *V-Bartlett*

- c. Menentukan persamaan diskriminan yaitu kombinasi linier dari variabel independen yang dirumuskan:

$$D = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + BX_p$$

- d. Uji validasi fungsi diskriminan dengan menggunakan *Hit Ratio* dan Akurasi Statistik.

3. HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini menggunakan satuan yang berbeda, sehingga harus dilakukan standarisasi ke dalam bentuk baku. Variabel yang digunakan adalah TPT(X1), UMP(X2), dan IPM(X3).

A. Tahun 2018

Uji Kenormalan Data Tahun 2018

Uji kenormalan data dengan menggunakan uji *Anderson-Darling*. Hasil pengujian menunjukkan bahwa nilai *p-value* untuk variabel X1, X2, dan X3 masing-masing sebesar 0.0858, 0.3357, dan 0.07643, yang seluruhnya lebih besar dari 0.05. dengan demikian, H_0 diterima, yang menunjukkan bahwa variabel X1, X2, dan X3 mengikuti sebaran normal.

Uji Multikolinearitas Data Tahun 2018

Nilai *tolerance* berada diantara 0-2, nilai VIF < 10. Hal ini menunjukan bahwa tidak terja di multikolinearitas.

Analisis Kelompok Tahun 2018 (*Cluster Analysis*)

Berdasarkan hasil analisis *cluster* menggunakan metode *Ward*, diperoleh 3 *cluster* sebagai solusi terbaik. Setelah anggota *cluster* terbentuk, langkah selanjutnya adalah menentukan profil masing-masing *cluster* dengan menghitung rata-rata setiap variabel pada tiap *cluster*.

Tabel 3. 1 Rata-rata Tiap Kelompok

Cluster	X1	X2	X3
1	1,1550920	0,2533139	0,7385114
2	-0,7709315	-0,7709315	-0,6635133
3	-0,1082685	-0,1082685	0,9651496

Berdasarkan **Tabel 3.1**, *Cluster* 1 memiliki nilai rata-rata paling tinggi pada variabel X1, X2, dan X3, sehingga daerah yang termasuk dalam *Cluster* 1 dapat dikategorikan sebagai daerah dengan kondisi sosial ekonomi paling baik. Sebaliknya, *Cluster* 2 memiliki nilai rata-rata paling rendah pada ketiga variabel, sehingga daerah pada *Cluster* 2 tergolong daerah dengan kondisi sosial ekonomi paling rendah. Sementara itu, *Cluster* 3 memiliki nilai rata-rata yang berada di antara *Cluster* 1 dan *Cluster* 2, sehingga daerah pada *Cluster* 3 dapat dikategorikan sebagai daerah dengan kondisi sosial ekonomi menengah dibandingkan dua *cluster* lainnya.

Uji Homogenitas

Nilai *p-value* dari hasil analisis menggunakan *Box's M-test* adalah 0.3151, yang jauh lebih besar dari tingkat signifikansi = 0.05. Oleh karena itu, hipotesis nol yang menyatakan homogenitas matriks kovarians antar *cluster* tidak ditolak, dan dapat disimpulkan bahwa asumsi homogenitas matriks kovarians telah terpenuhi.

Analisis Diskriminan Tahun 2018

Dikarenakan terdapat 3 *cluster* yang terbentuk, maka fungsi diskriminan yang didapatkan adalah 2.

Tabel 3. 2 Nilai Eigen

<i>Fuction</i>	<i>Nilai Eigen</i>	<i>Varian</i>	<i>Korelasi Konikal</i>
1	62.848	89.2	0.992
2	7.644	10.8	0.940

Berdasarkan **Tabel 3.2**, terlihat fungsi diskriminan pertama memiliki nilai *eigen* sebesar 62.848 dan mampu menjelaskan 89.2% varians. Hal ini menunjukkan bahwa fungsi diskriminan pertama memiliki kekuatan diskriminan yang sangat kuat dalam membedakan antar kelompok. Nilai korelasional kanonikal pada fungsi pertama sebesar 0.992, dengan nilai $r^2 = 0.984$, angka ini menunjukkan bahwa sekitar 98.4% varians pada variabel dependen (kelompok) dapat dijelaskan model diskriminan fungsi pertama.

Sementara itu, fungsi diskriminan kedua hanya menjelaskan 7.644 dan hanya mampu menjelaskan 10.8% varians, sehingga kontribusinya dalam membedakan kelompok lebih relatif kecil dibandingkan fungsi diskriminan pertama.

Model fungsi diskriminan dapat dituliskan sebagai berikut:

$$D_1 = b_1X_1 + b_2X_2 + b_3X_3$$

$$D_2 = b_1X_1 + b_2X_2 + b_3X_3$$

Fungsi Diskriminan 1 (LD1):

$$D_1 = -1.056X_1 - 0.00000144X_2 - 0.208X_3$$

Fungsi Diskriminan 1 (LD2):

$$D_2 = -0.204X_1 + 0.00000321X_2 - 0.156X_3$$

(dengan nilai koefisien b diperoleh dari hasil koefisien fungsi diskriminan).

Validasi Fungsi Diskriminan Tahun 2018

1. Hit Ratio

27 data diprediksi secara tepat keanggotaannya. Sehingga dapat disimpulkan bahwa model diskriminan yang terbentuk mempunyai tingkat validasi 100%.

2. Akurasi Statistik

$$Press'Q = \frac{[31 - (30 \times 3)]^2}{31(3 - 1)} = \frac{3481}{68} = 56.51$$

Dengan menggunakan $\alpha = 0.05$, dan $df = 1$, nilai X^2 tabel adalah 5.991.

Dikarenakan nilai $Q > X^2$ maka dapat disimpulkan bahwa fungsi diskriminan akurat.

B. Tahun 2019

Uji Kenormalan Data Tahun 2019

Uji kenormalan data dilakukan dengan menggunakan *uji Anderson-Darling*. Hasil pengujian menunjukkan bahwa nilai *p-value* untuk variabel X1, X2, dan X3 masing-masing sebesar 0.232, 0.3974, dan 0.117, yang seluruhnya lebih besar dari 0.05. Dengan demikian, H_0 diterima, yang menunjukkan bahwa variabel X1, X2, dan X3 mengikuti sebaran normal.

Uji Multikolinearitas Data Tahun 2019

Nilai *tolerance* berada diantara 0-2, nilai VIF < 10. Hal ini menunjukan bahwa tidak terjadi multikolinearitas.

Analisis Kelompok Tahun 2019 (*Cluster Analysis*)

Berdasarkan hasil analisis *cluster* menggunakan metode *Ward*, diperoleh 3 *cluster* sebagai solusi terbaik. Setelah anggota *cluster* terbentuk, langkah selanjutnya adalah menentukan profil masing-masing *cluster* dengan menghitung rata-rata setiap variabel pada tiap *cluster*.

Tabel 3. 3 Rata-rata Tiap Kelompok

Cluster	X1	X2	X3
1	1.34731533	0.6543210	0.8656966
2	-0.04538778	0.5263541	-0.0303412
3	-0.74432847	-0.8641785	-0.4771769

Berdasarkan **Tabel 3.3**, *Cluster* 1 memiliki nilai rata-rata paling tinggi pada variabel X1, X2, dan X3, sehingga daerah yang termasuk dalam *Cluster* 1 dapat dikategorikan sebagai daerah dengan kondisi sosial ekonomi paling baik. Sebaliknya, *Cluster* 3 memiliki nilai rata-rata paling rendah pada ketiga variabel, sehingga daerah pada *Cluster* 2 tergolong daerah dengan kondisi sosial ekonomi paling rendah. Sementara itu, *Cluster* 2 memiliki nilai rata-rata yang berada di antara *Cluster* 1 dan *Cluster* 3, sehingga daerah pada *Cluster* 3 dapat dikategorikan sebagai daerah dengan kondisi sosial ekonomi menengah dibandingkan dua *cluster* lainnya.

Uji Homogenitas

Nilai *p-value* dari hasil analisis menggunakan *Box's M-test* adalah 0.762, yang jauh lebih besar dari tingkat signifikansi $\alpha = 0.05$. Oleh karena itu, hipotesis nol yang menyatakan homogenitas matriks kovarians antar *cluster* tidak ditolak, dan dapat disimpulkan bahwa asumsi homogenitas matriks kovarians telah terpenuhi.

Analisis Diskriminan Tahun 2019

Dikarenakan terdapat 3 *cluster* yang terbentuk, maka fungsi diskriminan yang didapatkan adalah 2.

Tabel 3. 4 Nilai Eigen Tahun 2019

Fuction	Nilai Eigen	Varian	Korelasi Konikal
1	72.601	95.4	0.993
2	3.530	4.6	0.883

Berdasarkan **Tabel 3.4**, terlihat bahwa fungsi diskriminan pertama memiliki nilai eigen sebesar 72,601 dan mampu menjelaskan 95,4% varians total. Hal ini menunjukkan bahwa fungsi diskriminan pertama memiliki kekuatan diskriminan yang sangat kuat dalam membedakan antar kelompok. Nilai korelasi kanonikal pada fungsi pertama sebesar 0,993, sehingga diperoleh nilai $r^2 = 0,993 = 0,986$. Angka ini menunjukkan bahwa sekitar 98,6% variasi pada variabel dependen (kelompok) dapat dijelaskan oleh model diskriminan pada fungsi pertama.

Sementara itu, fungsi diskriminan kedua memiliki nilai eigen sebesar 3,530 dan hanya mampu menjelaskan 4,6% varians. Nilai korelasi kanonikalnya sebesar 0,883, dengan $r^2 = 0,780$. Meskipun masih menunjukkan hubungan yang cukup kuat, kontribusi fungsi diskriminan kedua dalam membedakan kelompok relatif lebih kecil dibandingkan fungsi diskriminan pertama.

Dengan demikian, dapat disimpulkan bahwa fungsi diskriminan pertama merupakan fungsi yang paling dominan dan paling penting dalam proses klasifikasi dan pembedaan antar kelompok, sedangkan fungsi diskriminan kedua hanya berperan sebagai pelengkap.

Model fungsi diskriminan dapat dituliskan sebagai berikut:

$$D_1 = b_1X_1 + b_2X_2 + b_3X_3$$

$$D_2 = b_1X_1 + b_2X_2 + b_3X_3$$

Fungsi Diskriminan 1 (LD1):

$$LD1 = -1.076477X_1 - 0.000002352X_2 - 0.2738043X_3$$

Fungsi Diskriminan 2 (LD2):

$$LD2 = 0.4931886X_1 - 0.000002912X_2 - 0.09874397X_3$$

(dengan nilai koefisien b diperoleh dari hasil koefisien fungsi diskriminasi).

Validasi Fungsi Diskriminan Tahun 2019

1. Hit Ratio

27 data diprediksi secara tepat keanggotaannya. Sehingga dapat disimpulkan bahwa model diskriminan yang terbentuk mempunyai tingkat validasi 100%.

2. Akurasi Statistik

$$Press'Q = \frac{[30 - (28 \times 3)]^2}{30(3 - 1)} = \frac{2500}{68} = 48.6$$

Dengan menggunakan $\alpha = 0.05$, dan $df = 1$, nilai X^2 adalah 5.991. dikarenakan nilai $Q > X^2$ dapat disimpulkan bahwa fungsi diskriminan akurat.

C. Tahun 2020

Uji Kenormalan Data Tahun 2020

Uji kenormalan data dilakukan dengan menggunakan *uji Anderson-Darling*. Hasil pengujian menunjukkan bahwa nilai *p-value* untuk variabel X1, X2, dan X3 masing-masing sebesar 0.5199, 0.7568, dan 0.5836, yang seluruhnya lebih besar dari 0.05. Dengan demikian, H_0 diterima, yang menunjukkan bahwa variabel X1, X2, dan X3 mengikuti sebaran normal.

Uji Multikolinearitas Data Tahun 2020

Nilai *tolerance* berada diantara 0-2, nilai VIF < 10. Hal ini menunjukkan bahwa tidak terjadi multikolinearitas.

Analisis Kelompok Tahun 2020 (*Cluster Analysis*)

Berdasarkan hasil analisis *cluster* menggunakan metode *Ward*, diperoleh 2 *cluster* sebagai solusi terbaik. Setelah anggota *cluster* terbentuk, langkah selanjutnya adalah menentukan profil masing-masing *cluster* dengan menghitung rata-rata setiap variabel pada tiap *cluster*.

Tabel 3. 5 Rata-rata Tiap Kelompok

Cluster	X1	X2	X3
1	0.5667542	0.6425478	0.6154414
2	0.6122129	-2.0670672	0.2258283
3	-0.8326350	-0.4419577	-0.8243488

Berdasarkan **Tabel 3.5**, terlihat bahwa *cluster* 1 memiliki rata-rata paling tinggi pada variabel X1, X2, dan X3, sehingga daerah yang termasuk dalam *Cluster* 1 dapat dikategorikan sebagai daerah dengan kondisi sosial paling baik. Sebaliknya, *Cluster* 2 memiliki nilai rata-rata paling rendah pada ketiga variabel, sehingga daerah pada *Cluster* 2 tergolong daerah dengan kondisi sosial ekonomi yang paling rendah dibandingkan *Cluster* 1.

Uji Homogenitas

Nilai *p - value* dari hasil analisis menggunakan *Box's M-test* adalah 0.9527, yang jauh lebih besar dari tingkat signifikansi $\alpha = 0.05$. Oleh karena itu, hipotesis nol yang menyatakan homogenitas matriks kovarians antar *cluster* tidak ditolak, dan dapat disimpulkan bahwa asumsi homogenitas matriks kovarians telah terpenuhi.

Analisis Diskriminan Tahun 2020

Dikarenakan terdapat 2 *cluster* yang terbentuk, maka fungsi diskriminan yang didapatkan adalah sebanyak 1.

Tabel 3. 6 Nilai Eigen Tahun 2020

<i>Fuction</i>	<i>Nilai Eigen</i>	<i>Varian</i>	<i>Korelasi Konikal</i>
1	72.777	100	0.993

Berdasarkan **Tabel 3.6**, terlihat bahwa fungsi diskriminan memiliki nilai eigen sebesar 72,777 dan mampu menjelaskan 100% varians total. Hal ini menunjukkan bahwa fungsi diskriminan pertama memiliki kekuatan diskriminan yang sangat kuat dan membedakan antar kelompok. Nilai korelasi kanonikal pada fungsi pertama sebesar 0,993 sehingga diperoleh nilai $r^2 = 0,993^2 = 0,986$. Angka ini menunjukkan bahwa sekitar 98,6% variasi pada variabel dependen (kelompok) dapat dijelaskan oleh modek diskriminan pada fungsi pertama.

Model fungsi diskriminan dapat dituliskan sebagai berikut:

$$D_1 = b_1X_1 + b_2X_2 + b_3X_3$$

Fungsi Diskriminan 1 (LD1):

$$LD1 = -5.520633X_1 + 0.0000032X_2 - 0.2041701X_3$$

Fungsi Diskriminan 2 (LD2):

$$LD2 = 2.167368X_1 + 0.000025X_2 + 0.1238626X_3$$

(dengan nilai koefisien b diperoleh dari hasil koefisien fungsi diskriminan).

Validasi Fungsi Diskriminan Tahun 2020

1. Hit Ratio

26 data diprediksi secara tepat keanggotaannya. Sehingga dapat disimpulkan bahwa model diskriminan yang terbentuk mempunyai tingkat validasi 96.2963%.

2. Akurasi Statistik

$$Press'Q = \frac{[27 - (27 \times 2)]^2}{27(2 - 1)} = \frac{625}{27} = 23.148$$

Dengan menggunakan $\alpha = 0.05$, dan $df = 1$, nilai X^2 adalah 5.991. dikarenakan nilai $Q > X^2$ dapat disimpulkan bahwa fungsi diskriminan akurat.

D. Gabungan Tahun 2018, 2019, 2020

Uji Kenormalan Data Gabungan

Uji kenormalan data dilakukan dengan menggunakan *uji Anderson-Darling*. Hasil pengujian menunjukkan bahwa nilai *p-value* untuk variabel X1, X2, dan X3 masing-masing sebesar 0.2174, 0.8249, dan 0.1111, yang seluruhnya lebih besar dari 0.05. Dengan demikian, diterima, yang menunjukkan bahwa variabel X1, X2, dan X3 mengikuti sebaran normal.

Uji Multikolinearitas Data Gabungan

Nilai *tolerance* berada diantara 0-2, nilai VIF < 10. Hal ini menunjukkan bahwa tidak terjadi multikolinearitas.

Analisis Kelompok Data Gabungan (*Cluster Analysis*)

Berdasarkan hasil analisis *cluster* menggunakan metode *Ward*, diperoleh 3 *cluster* sebagai solusi terbaik. Setelah anggota *cluster* terbentuk, langkah selanjutnya adalah menentukan profil masing-masing *cluster* dengan menghitung rata-rata setiap variabel pada tiap *cluster*.

Tabel 3. 7 Rata-rata Tiap Kelompok

Cluster	X1	X2	X3
1	1.42381503	0.8695877	1.18288191
2	0.02107592	0.5004844	0.02521847
3	-0.75264040	-0.9029552	-0.63228929

Berdasarkan **Tabel 3.7**, terlihat bahwa *cluster* 1 memiliki nilai rata-rata tertinggi pada seluruh variabel, khususnya pada variabel X1 (Tingkat Pengangguran Terbuka), X2 (Upah Minimum Regional), dan X3 (Indeks Pembangunan Manusia). Hal ini menunjukkan bahwa *cluster* 1 merepresentasikan wilayah dengan tingkat pengangguran, upah minimum, dan IPM yang relatif lebih tinggi dibandingkan *cluster* lainnya. *Cluster* 2 memiliki nilai rata-rata berada di sekitar nol pada seluruh variabel, yang menunjukkan kondisi sosial ekonomi yang berada pada tingkat sedang atau mendekati rata-rata keseluruhan data. Sementara itu, *cluster* 3 memiliki nilai rata-rata terendah pada seluruh variabel, yang mencerminkan wilayah dengan tingkat pengangguran terbuka, upah minimum, dan indeks pembangunan manusia yang relatif lebih rendah dibandingkan *cluster* lainnya.

Interpretasi Data

Dari keempat analisis data disimpulkan bahwa daerah yang konstan berada di kelompok yang sama adalah:

Tabel 3. 8 Daerah Yang Konstan Dalam Kelompok

Tingkat Kemiskinan Tertinggi	Bengkulu Lampung Nusa Tenggara Barat Nusa Tenggara Timur Kalimantan Barat Sulawesi Tengah Sulawesi Barat Maluku Utara
Tingkat Kemiskinan Sedang	Kalimantan Tengah Kalimantan Selatan Sulawesi Selatan Papua Barat
Tingkat Kemiskinan Terendah	Aceh Riau Kepulauan Riau Banten Maluku

Uji Homogenitas

Nilai dari hasil analisis menggunakan *Box's M-test* adalah 0.1762, yang jauh lebih besar dari tingkat signifikansi = 0.05. Oleh karena itu, hipotesis nol yang menyatakan homogenitas matriks kovarians antar *cluster* tidak ditolak, dan dapat disimpulkan bahwa asumsi homogenitas matriks kovarians telah terpenuhi.

Analisis Diskriminan Data Gabungan

Dikarenakan terdapat 3 *cluster* yang terbentuk, maka fungsi diskriminan yang didapatkan adalah sebanyak 2.

Tabel 3. 9 Nilai Eigen Data Gabungan

Fuction	Nilai Eigen	Varian	Korelasi Konikal
----------------	--------------------	---------------	-------------------------

1	178.597	96.11	0.997
2	7.235	3.89	0.937

Berdasarkan **Tabel 3.9**, terlihat bahwa terbentuk dua fungsi diskriminan karena terdapat tiga cluster. Fungsi diskriminan pertama memiliki nilai eigen sebesar 94,624 dan mampu menjelaskan 70,8% variasi data, sedangkan fungsi diskriminan kedua memiliki nilai eigen sebesar 39,055 dengan kontribusi sebesar 29,2%. Hal ini menunjukkan bahwa fungsi diskriminan pertama merupakan fungsi yang paling dominan dalam membedakan antar cluster. Nilai korelasi kanonikal yang tinggi pada kedua fungsi, yaitu 0,995 pada fungsi pertama dan 0,987 pada fungsi kedua, mengindikasikan bahwa fungsi diskriminan memiliki kemampuan yang sangat kuat dalam memisahkan kelompok yang terbentuk.

Berdasarkan koefisien konikal diperoleh model diskriminan, yaitu:
Fungsi Diskriminan 1 (LD1):

$$D_1 = -1.042156X_1 - 0.00000275X_2 - 0.120301X_3$$

Fungsi Diskriminan 2 (LD2):

$$D_2 = 0.569469X_1 - 0.00000265X_2 - 0.143789X_3$$

Validasi Fungsi Diskriminan data Gabungan

1. *Hit Ratio*

Dari total seluruh data yang dianalisis, sebagian besar observasi berhasil diklasifikasikan dengan benar sehingga diperoleh nilai *hit ratio* sebesar 95%. Nilai tersebut mengindikasikan bahwa fungsi diskriminan memiliki kemampuan klasifikasi yang tinggi.

2. Akurasi Statistik

$$Press'Q = \frac{(80 - 76 \times 3)^2}{80(3 - 1)} = \frac{(80 - 228)^2}{160} = \frac{21.904}{160} = 136.9$$

Dengan menggunakan $\alpha = 0.05$, dan $df = 1$, nilai X^2 adalah 5.991. dikarenakan nilai $Q > X^2$ dapat disimpulkan bahwa fungsi diskriminan akurat.

4. KESIMPULAN

1. Berdasarkan hasil analisis cluster dan analisis diskriminan yang telah dilakukan terhadap data Tingkat Pengangguran Terbuka (TPT), Upah Minimum Regional (UMR), dan Indeks Pembangunan Manusia (IPM) pada provinsi-provinsi di Indonesia tahun 2018–2020, dapat disimpulkan bahwa provinsi di Indonesia dapat dikelompokkan ke dalam beberapa kelompok tingkat kemiskinan dengan karakteristik sosial ekonomi yang berbeda. Pengelompokan menggunakan metode *Ward* secara konsisten membentuk kelompok dengan tingkat kemiskinan tertinggi, sedang, dan terendah, yang mencerminkan adanya ketimpangan kondisi kesejahteraan antarwilayah di Indonesia.
2. Hasil analisis diskriminan menunjukkan bahwa fungsi diskriminan yang terbentuk memiliki kemampuan klasifikasi yang sangat baik dalam membedakan provinsi ke dalam kelompoknya masing-masing. Hal ini ditunjukkan oleh nilai eigen dan korelasi kanonikal yang tinggi serta tingkat ketepatan klasifikasi yang besar, sehingga variabel TPT, UMR, dan IPM terbukti merupakan indikator yang kuat dalam menjelaskan perbedaan tingkat kemiskinan antarprovinsi.
3. Berdasarkan hasil tersebut, penelitian ini merekomendasikan agar penelitian selanjutnya mengembangkan model dengan menambahkan variabel sosial ekonomi lainnya serta menggunakan periode data yang lebih panjang untuk menangkap dinamika perubahan

tingkat kemiskinan secara lebih komprehensif. Selain itu, hasil pengelompokan ini dapat dijadikan sebagai salah satu acuan bagi pemerintah dalam merumuskan kebijakan penanggulangan kemiskinan yang lebih terarah dan berbasis karakteristik wilayah, sehingga upaya pengentasan kemiskinan dapat dilakukan secara lebih efektif dan berkelanjutan.

5. REFERENSI

- [1] Ansari , Muhammad Novi. (2020). Metode Analisis Diskriminan dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Indeks Pembangunan Manusia. *Journal of Statistics and Its Application on Teaching and Research* , 6. Retrieved from ojs.unm.ac.id.
- [2] BPS. (2017 , Maret 23). *Upah Minimum Regional/Propinsi (Rupiah), 2020*. Retrieved from bps.go.id: <https://www.bps.go.id/id/statistics-table/2/MjIwIzI=/upah-minimum-regional-propinsi.html>
- [3] BPS. (2025, Desember 12). *Kependudukan dan Migrasi*. Retrieved from bps.go.id: <https://www.bps.go.id/id/statistics-table/2/MjIwIzI=/upah-minimum-regional-propinsi.html>
- [4] Hidayat , A. (2024). *Analisis Cluster*. Retrieved from statistikian.com: https://www.statistikian.com/2014/03/analisis-cluster_27.html
- [5] Simamora, B. (2005). Analisis Multivariat Pemasaran. In B. Simamora, *Analisis Multivariat Pemasaran* (p. 166). Jakarta: Penerbit PT Gramedia Pustaka Utama.