

ANALISIS FAKTOR TERHADAP INDEKS PEMBANGUNAN MANUSIA DAN PENGELOMPOKAN PROVINSI DI INDONESIA TAHUN 2023

Wildan 'Izza Nur Khabib*, Pismia Sylvi

Program Studi Statistika, Universitas Terbuka, Tangerang Selatan, Indonesia

*Penulis korespondensi: wildanizza29@gmail.com

ABSTRAK

Isu Sumber Daya Manusia (SDM) di Indonesia menjadi perhatian utama dalam beberapa tahun terakhir, pengembangan SDM dapat dilihat dari Indeks Pembangunan Manusia (IPM). Penelitian ini bertujuan untuk menganalisis perbedaan IPM setiap provinsi di Indonesia pada tahun 2023 serta mengidentifikasi faktor-faktor yang memengaruhinya. Metode analisis yang digunakan adalah analisis regresi linear berganda untuk mengetahui pengaruh variabel independen terhadap variabel dependen. Korelasi tersebut menghasilkan nilai koefisien R Square sebesar 0,995. Uji asumsi residual yang mencakup normalitas, linearitas, dan homoskedastisitas menunjukkan bahwa model regresi memenuhi semua asumsi statistik yang diperlukan. Berdasarkan analisis menggunakan metode k-means clustering, provinsi-provinsi di Indonesia dapat dikelompokkan ke dalam empat kategori, yaitu rendah, sedang, tinggi, dan sangat tinggi. Provinsi dengan IPM rendah memerlukan intervensi kebijakan untuk memperbaiki infrastruktur, meningkatkan akses terhadap layanan dasar, dan memberdayakan ekonomi masyarakat. Sementara itu, provinsi dengan IPM sedang dan tinggi perlu mempertahankan serta meningkatkan kemajuan yang telah dicapai melalui upaya berkelanjutan di sektor-sektor penting. DKI Jakarta, sebagai provinsi dengan IPM sangat tinggi, dapat dijadikan model bagi strategi pembangunan di provinsi lain. Hasil penelitian ini menunjukkan pentingnya pendekatan yang disesuaikan dengan kondisi setiap provinsi untuk mencapai pembangunan manusia yang lebih merata dan berkelanjutan di Indonesia.

Kata kunci: indeks pembangunan manusia, *k-means clustering*, korelasi, regresi linear berganda

1 PENDAHULUAN

Isu Sumber Daya Manusia (SDM) di Indonesia beberapa tahun terakhir menjadi topik yang menarik untuk dibicarakan di publik. Berbagai permasalahan mulai dari kualitas pendidikan di Indonesia guna menciptakan kesiapan tenaga kerja dalam menghadapi tantangan era globalisasi. Situasi ini semakin parah dengan adanya pandemi COVID-19 yang memberikan dampak signifikan terhadap perkembangan SDM di Indonesia. Kesejahteraan umum dan kualitas pengembangan SDM suatu wilayah dapat dilihat dari Indeks Pembangunan Manusia (IPM). IPM dijadikan sebagai gambaran berhasil atau tidaknya suatu program pembangunan infrastruktur di suatu daerah (Widodo *et al.*, 2020). Infrastruktur bukan sekedar pembangunan jalan raya, jalan tol, jembatan, rel kereta api, bandar udara, namun juga dilihat dari fasilitas pelayanan dasar seperti ekonomi, kesehatan, dan pendidikan (Afriyana *et al.*, 2023).

Indonesia adalah negara berkembang yang memiliki jangka panjang untuk menjadi negara maju, keberadaan sumber daya alam yang melimpah harus diimbangi dengan pembangunan infrastruktur untuk memajukan SDM yang ada. Saat ini pemerintah sedang fokus membangun infrastruktur untuk menunjang SDM yang berkualitas (Tubani & Hasanah, 2022). SDM adalah pemanfaatan manusia sebagai tenaga kerja yang memiliki keahlian dan

kemampuan baik secara fisik dan psikis guna mencapai tujuan organisasi yang telah ditetapkan (Kurniawan *et al.*, 2023). Pembangunan manusia sekarang menjadi tujuan pembangunan nasional yang di atur dalam Undang-Undang Nomor 25 Tahun 2004 tentang Sistem Perencanaan Pembangunan Nasional (Ummah, 2024).

IPM dipengaruhi oleh tiga dimensi, yaitu kesehatan, pendidikan, dan ekonomi (BPS, 2023). Kualitas masyarakat Indonesia saat ini masih belum berjalan dengan baik ditandai dengan angka kemiskinan yang masih tinggi, rendahnya pendidikan dan tingginya kasus gizi buruk (Kusuma & Muta'ali, 2019). Jika dilihat dari Berita Resmi Statistik No. 80/11/Th.XXVI, 15 November 2023 dari Badan Pusat Statistik (BPS) menunjukkan bahwa IPM antar provinsi di Indonesia masih terdapat kesenjangan. Penelitian Widodo *et al.* (2020) menggunakan metode *average linkage* menyatakan bahwa 20 dari 35 kabupaten/kota di Jawa Tengah memiliki IPM yang masih rendah. Penelitian Tubani & Hasanah (2022) menggunakan metode *k-medoids clustering* menyatakan bahwa 16 dari 22 kabupaten/kota di Nusa Tenggara Timur memiliki IPM yang masih rendah. Oleh karena itu, pentingnya menganalisis korelasi antara variabel dependen, yaitu nilai Indeks Pembangunan Manusia (Y), dengan empat variabel independen, yaitu Usia Harapan Hidup (X_1), Rata-Rata Lama Sekolah (X_2), Harapan Lama Sekolah (X_3), dan Pengeluaran per Kapita (X_4). Penelitian ini bertujuan mengelompokkan provinsi-provinsi ke dalam beberapa kelompok yang mirip berdasarkan variabel-variabel tersebut. Hasil dari penelitian ini diharapkan dapat memberikan gambaran tentang pembangunan infrastruktur yang ada di Indonesia dan kesenjangan yang terjadi di Indonesia.

2 METODE

Data penelitian adalah data sekunder dari Berita Resmi Statistik No. 80/11/Th.XXVI, 15 November 2023 dari <https://www.bps.go.id/id/pressrelease/2023/11/15/2033>. Data yang digunakan mencakup 34 provinsi, sementara pada tahun 2024 jumlah provinsi di Indonesia sebanyak 38 provinsi. Berikut adalah provinsi yang belum terdaftar pada data BPS IPM tahun 2023, yaitu Papua Tengah, Papua Pegunungan, Papua Selatan, Papua Barat Daya. Penelitian ini menggunakan dua metode, yaitu analisis regresi linear berganda diolah menggunakan aplikasi SPSS dan *k-means clustering* diolah menggunakan bahasa pemrograman Python.

Data yang digunakan memiliki lima variabel, yaitu X_1, X_2, X_3, X_4 , dan Y . Adapun penjelasan variabel yang digunakan dalam penelitian ini adalah sebagaimana tercantum pada Tabel 1.

Tabel 1. Definisi Variabel Penelitian

Variabel	Peran	Definisi
IPM	Variabel Dependen (Y)	Indikator untuk mengukur kesejahteraan masyarakat suatu wilayah
Usia Harapan Hidup SP2020	Variabel Independen (X_1)	Perkiraan rata-rata usia yang diharapkan hidup berdasarkan data Sensus Penduduk tahun 2020
Rata-Rata Lama Sekolah	Variabel Independen (X_2)	Menggambarkan tingkat pendidikan dari lama sekolah yang ditempuh oleh masyarakat daerah tersebut
Harapan Lama Sekolah	Variabel Independen (X_3)	Harapan jumlah tahun masyarakat dalam menempuh pendidikan formal
Pengeluaran Perkapita	Variabel Independen (X_4)	Jumlah pengeluaran rata-rata yang dialokasikan untuk kebutuhan pokok di suatu daerah

Adapun hipotesis yang digunakan adalah

H_0 : Terdapat pengaruh variabel independen, yaitu Usia Harapan Hidup (X_1), Rata-Rata Lama Sekolah (X_2), Harapan Lama Sekolah (X_3), dan Pengeluaran per Kapita (X_4), terhadap variabel dependen, yaitu Indeks Pembangunan Manusia (Y).

H_1 : Tidak ada pengaruh variabel independen, yaitu Usia Harapan Hidup (X_1), Rata-Rata Lama Sekolah (X_2), Harapan Lama Sekolah (X_3), dan Pengeluaran per Kapita (X_4), terhadap variabel dependen, yaitu Indeks Pembangunan Manusia (Y).

2.1 Uji Normalitas

Pengujian untuk membuktikan normal atau tidaknya suatu data dilakukan dengan analisis uji Shapiro-Wilk. Uji Shapiro-Wilk adalah metode perhitungan sebaran data yang dibuat oleh Shapiro dan Wilk (Quraisy, 2020). Uji normalitas residual dilakukan sebelum analisis regresi linear berganda untuk memastikan asumsi-asumsi yang mendasari model regresi terpenuhi, yaitu bahwa residual berdistribusi normal. Asumsi ini penting karena jika residual berdistribusi normal, estimasi koefisien regresi menjadi lebih akurat dan efisien. Berikut rumus uji Shapiro-Wilk:

$$W = \frac{(\sum_{i=1}^n a_i y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

dimana y_i adalah nilai data urutan ke- i dan \bar{y} adalah rata-rata sampel.

2.2 Asumsi Residual

Uji asumsi residual adalah bagian penting dalam proses penelitian karena hasil uji ini memastikan bahwa estimasi parameter tidak bias, yang berarti hasil dari pemodelan tersebut valid dan dapat dipertanggungjawabkan (Hermanto *et al.*, 2021). Dalam analisis regresi, penting untuk memeriksa asumsi residual meskipun variabel independennya sudah berdistribusi normal. Beberapa asumsi penting yang harus dipenuhi adalah normalitas, homoskedastisitas, dan independensi residual. Normalitas residual berarti bahwa sisa atau error dari model regresi harus berdistribusi normal agar hasil uji statistik, seperti uji t dan uji F, menjadi valid. Homoskedastisitas berarti bahwa variasi residual harus tetap sama di seluruh rentang nilai variabel independen. Jika variasi ini tidak konstan, atau disebut heteroskedastisitas, maka bisa terjadi masalah dalam estimasi parameter dan interpretasi hasil. Selain itu, residual juga harus independen satu sama lain, artinya tidak boleh ada pola atau ketergantungan yang bisa diprediksi di antara residual. Berikut rumus uji Durbin Watson:

$$d = \frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2} \quad (2)$$

dimana ε_t adalah residual urutan ke- t .

2.3 Analisis Regresi Linear Berganda

Regresi linear berganda adalah metode regresi yang digunakan untuk memahami hubungan antara variabel dependen dengan lebih dari satu variabel independen (Ernita *et al.*, 2022). Analisis regresi linear berganda bermanfaat untuk mengembangkan pemahaman yang lebih mendalam terkait hubungan antar variabel dan membuat prediksi yang lebih akurat. Berikut adalah langkah-langkah untuk melakukan analisis regresi linear berganda, yaitu:

1. Pengumpulan data
Langkah pertama adalah mengumpulkan data yang terdiri dari variabel dependen (Y) dan empat variabel independen (X_1, X_2, X_3 , dan X_4).

2. Menyusun persamaan model
Persamaan umum untuk regresi linear berganda adalah sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (3)$$

3. Menghitung koefisien regresi (β)
Koefisien regresi dapat dihitung dengan:

$$\beta = (X^T X)^{-1} X^T Y \quad (4)$$

4. Menghitung *sum of squares*
 - a. *Total Sum of Squares (SST)*

$$SST = \sum (Y_i - \bar{Y})^2 \quad (5)$$

- b. *Regression Sum of Squares (SSR)*

$$SSR = \sum (\hat{Y}_i - \bar{Y})^2 \quad (6)$$

- c. *Residual Error Sum of Squares (SSE)*

$$SSE = \sum (Y_i - \hat{Y}_i)^2 \quad (7)$$

- d. *Standar Error (SE)*

$$SE = \sqrt{\text{diag}(X^T X)^{-1} \left(\frac{SSE}{n - k - 1} \right)} \quad (8)$$

5. Menghitung uji statistik

- a. Koefisien r

Koefisien korelasi dilambangkan dengan huruf r , yang digunakan untuk mengukur seberapa kuat hubungan linear antar variabel yang ada. Nilai r yang mendekati 0 menandakan hubungan linear antar variabel lemah dan nilai r yang mendekati 1 menandakan hubungan linear antar variabel kuat. Adapun rumus dari nilai koefisien r , yaitu:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}} \sqrt{S_{yy}}} \quad (9)$$

- b. Uji F

Digunakan untuk menguji keberartian model secara khusus. Adapun rumus nilai F, yaitu:

$$F = \frac{SSR/k}{SSE/n - k - 1} \quad (10)$$

dimana k adalah jumlah variabel independen dan n adalah jumlah observasi. Hasil dari nilai F dibandingkan dengan F tabel, yaitu $F_{(k; n-k-1)}$.

c. Uji t

Digunakan untuk menguji korelasi masing-masing variabel independen dengan variabel dependen. Adapun rumus nilai t , yaitu:

$$t = \frac{\beta_i}{SE(\beta_i)} \quad (11)$$

hasil dari nilai t dibandingkan dengan t tabel, yaitu $t_{(\alpha/2; n-k-1)}$.

2.4 K-means Clustering

K-means clustering merupakan algoritma *unsupervised machine learning* yang digunakan untuk melakukan pengelompokan yang memisahkan data-data yang mirip menjadi satu kelompok (Rosida & Wijaya, 2023). Proses *k-means clustering* melibatkan langkah-langkah sebagai berikut:

1. Memuat data

Langkah pertama adalah mengumpulkan data yang terdiri dari variabel dependen (Y) dan empat variabel independen (X_1, X_2, X_3 , dan X_4).

2. Standardisasi data

Standardisasi data berguna untuk memastikan semua variabel memiliki bobot yang sama dalam analisis. Standardisasi dilakukan dengan rumus sebagai berikut:

$$Z_{score} = \frac{x - \bar{x}}{\sigma} \quad (12)$$

3. Melakukan *elbow method*

Elbow method adalah metode untuk menentukan jumlah cluster yang dikelompokkan dengan data yang mirip satu sama lain (Saputra *et al.*, 2020).

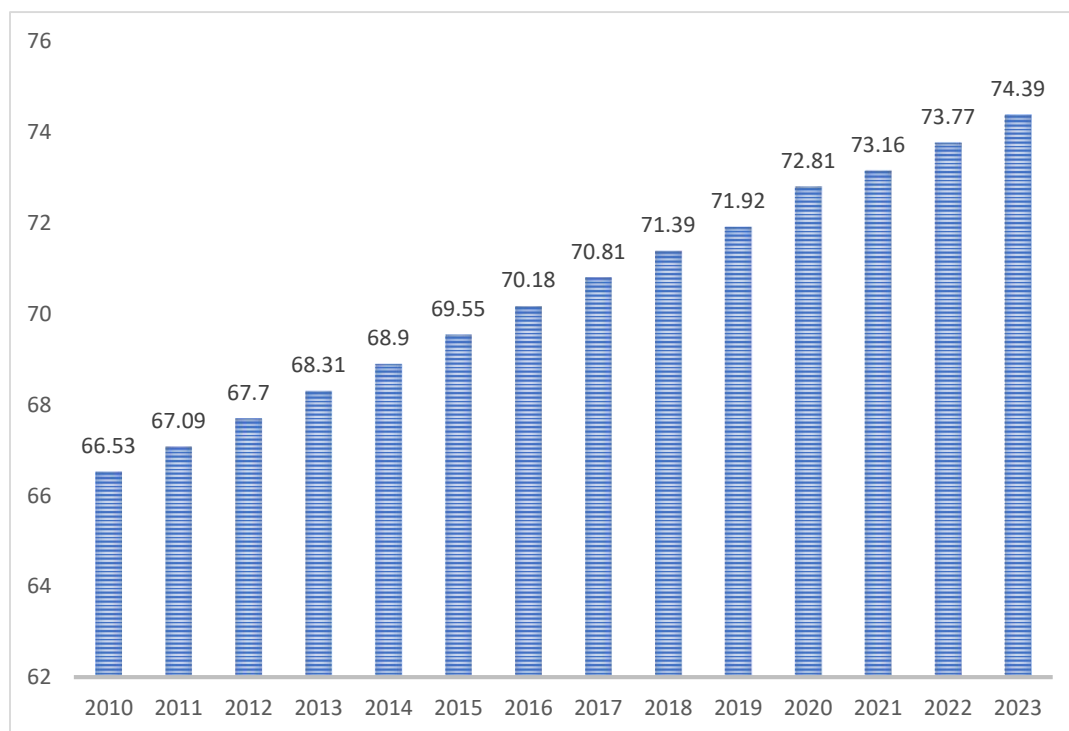
4. Melakukan *Principal Component Analysis* (PCA)

PCA digunakan karena dataset yang dimiliki memiliki lebih dari dua variabel. PCA adalah metode statistik yang digunakan untuk mengurangi dimensi dataset yang kompleks dengan mempertahankan sebanyak mungkin varians yang ada dalam data asli. Melalui transformasi linear, PCA menghasilkan serangkaian komponen utama, di mana setiap komponen utama ini merupakan kombinasi linear dari variabel asli dan mencerminkan dimensi yang mewakili informasi terpenting dari dataset tersebut (Yafi *et al.*, 2023).

3 HASIL DAN PEMBAHASAN

IPM adalah ukuran yang digunakan untuk menilai kualitas hidup penduduk di suatu wilayah berdasarkan tiga dimensi, yaitu: kesehatan, pendidikan, dan ekonomi. Pemahaman tentang IPM berguna untuk kemajuan pembangunan manusiadi Indonesia. Penelitian ini memberikan gambaran tentang faktor-faktor yang harus diperhatikan dalam perencanaan kebijakan untuk meningkatkan IPM. Hasil ini dapat menjadi landasan untuk penelitian lebih

lanjut dalam mengidentifikasi faktor-faktor alternatif yang dapat memengaruhi IPM di Indonesia.



Gambar 1. Perkembangan IPM Indonesia 2010 – 2023

Berdasarkan Gambar 1 dapat dilihat bahwa IPM Indonesia selalu mengalami kenaikan dari tahun ke tahun, itu menandakan bahwa Indonesia berupaya berbenah di segala bidang infrastruktur guna menciptakan pembangunan manusia yang lebih baik.

3.1 Uji Normalitas

Uji normalitas penting untuk memastikan interval prediksi yang akurat dan membantu mendeteksi outlier yang dapat memengaruhi hasil analisis. Oleh karena itu, uji normalitas residual menjadi langkah penting untuk memastikan model regresi linear berganda menghasilkan hasil yang valid dan dapat diandalkan. Jika asumsi normalitas tidak terpenuhi, perlu dilakukan langkah-langkah perbaikan seperti transformasi data atau menggunakan metode statistik yang lebih tepat.

Tabel 2. *Tests of Normality*

	Shapiro-Wilk		
	Statistic	df	Sig.
Usia Harapan Hidup SP2020	0,875	34	0,001
Rata - rata Lama Sekolah	0,979	34	0,726
Harapan Lama Sekolah	0,928	34	0,027
Pengeluaran Perkapita	0,920	34	0,016
Indeks Pembangunan Manusia	0,949	34	0,113

Berdasarkan Tabel 2 uji Shapiro-Wilk dari masing-masing variabel yaitu:

- a. Variabel X_1
Berdasarkan Tabel 2, diketahui nilai signifikansi dari variabel Usia Harapan Hidup (X_1) sebesar 0,001. Karena nilai signifikansi $< 0,05$, artinya data Usia Harapan Hidup tidak berdistribusi normal.
- b. Variabel X_2
Berdasarkan Tabel 2, diketahui nilai signifikansi dari variabel Rata-Rata Lama Sekolah (X_2) sebesar 0,726. Karena nilai signifikansi $> 0,05$, artinya data Rata-Rata Lama Sekolah berdistribusi normal.
- c. Variabel X_3
Berdasarkan Tabel 2, diketahui nilai signifikansi dari variabel Harapan Lama Sekolah (X_3) sebesar 0,027. Karena nilai signifikansi $< 0,05$, artinya data Harapan Lama Sekolah tidak berdistribusi normal.
- d. Variabel X_4
Berdasarkan Tabel 2, diketahui nilai signifikansi dari variabel Pengeluaran per Kapita (X_4) sebesar 0,016. Karena nilai signifikansi $< 0,05$, artinya data Pengeluaran per Kapita tidak berdistribusi normal.
- e. Variabel Y
Berdasarkan Tabel 2, diketahui nilai signifikansi dari variabel Indeks Pembangunan Manusia (Y) sebesar 0,113. Karena nilai signifikansi $< 0,05$, artinya data Indeks Pembangunan Manusia berdistribusi normal.

Dari lima variabel yang diuji, tiga variabel, yaitu Usia Harapan Hidup, Harapan Lama Sekolah, dan Pengeluaran per Kapita, tidak berdistribusi normal, sedangkan dua variabel lainnya, yaitu Rata-Rata Lama Sekolah dan IPM, berdistribusi normal. Hal ini perlu diperhatikan dalam analisis lanjutan karena variabel yang tidak berdistribusi normal dapat memengaruhi hasil regresi atau analisis statistik lainnya.

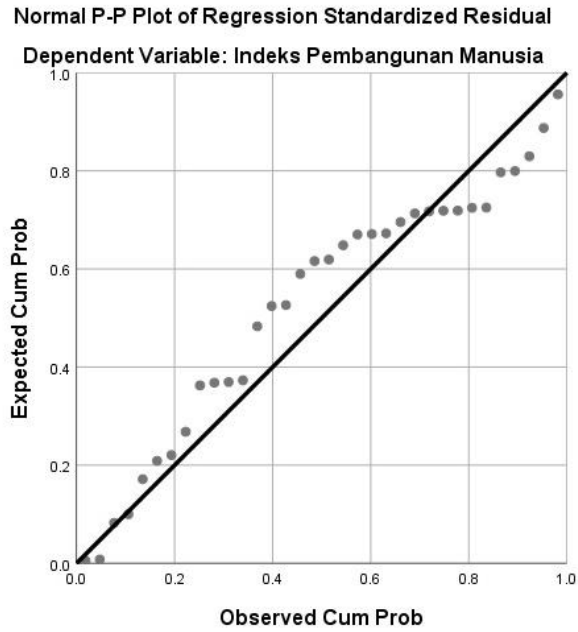
3.2 Asumsi Residual

Asumsi residual perlu diuji setelah melakukan uji normalitas dan sebelum melakukan analisis regresi linear berganda untuk memastikan bahwa model yang digunakan memenuhi asumsi-asumsi klasik regresi linear, seperti normalitas residual, homoskedastisitas, tidak adanya autokorelasi, dan linearitas.

Tabel 3. Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	0,998 ^a	0,995	0,995	0,27006	1,640

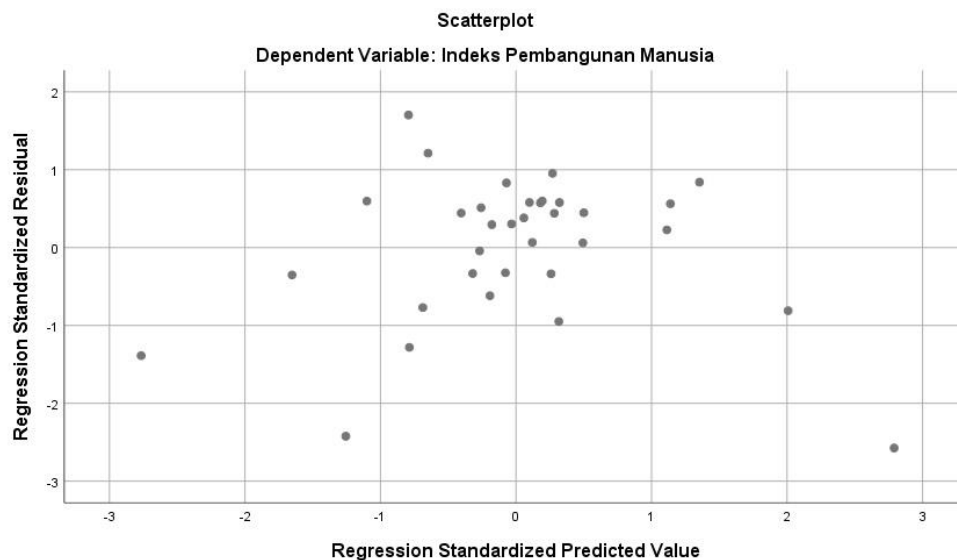
Berdasarkan Tabel 3, nilai Durbin-Watson sebesar 1,640. Nilai 1,640 cukup dekat dengan 2, yang menunjukkan tidak ada autokorelasi positif yang signifikan di antara residual, tetapi sedikit lebih rendah dari 2, yang bisa menunjukkan sedikit autokorelasi positif. Secara umum, nilai ini berada dalam batas yang dapat diterima, sehingga asumsi independensi residual dapat dianggap terpenuhi.



Gambar 2. Normalitas Residual (Normal P-P Plot)

Berdasarkan Gambar 2, titik-titik pada grafik sebagian besar mengikuti garis diagonal, yang menunjukkan bahwa residual cenderung normal. Beberapa penyimpangan dari garis diagonal di ujung bawah dan atas mungkin menunjukkan sedikit penyimpangan dari normalitas, tetapi ini umumnya masih dalam batas yang dapat diterima.

Selanjutnya, dari Gambar 3, plot menunjukkan distribusi titik-titik yang relatif acak tanpa pola yang jelas. Ini menunjukkan bahwa varians residual tampaknya konstan di seluruh rentang nilai prediksi, yang berarti asumsi homoskedastisitas terpenuhi. Dengan demikian, berdasarkan hasil yang diperoleh, model ini telah memenuhi asumsi-asumsi penting dalam regresi linear, yaitu normalitas residual dan homoskedastisitas.



Gambar 3. Scatterplot Standardized Residuals vs. Standardized Predicted Values

3.3 Analisis Regresi Linear Berganda

Data yang digunakan telah memenuhi asumsi-asumsi penting dalam regresi linear, yaitu normalitas residual dan homoskedastisitas, sehingga analisis regresi linear berganda dapat dilakukan. Penelitian dilakukan dengan metode analisis regresi linear berganda dengan memodelkan hubungan antara variabel dependen dan empat variabel independen.

Berdasarkan Tabel 3 bahwa nilai koefisien *R Square* sebesar 0,995. Itu artinya bahwa variabel independen berpengaruh terhadap variabel dependen sebesar 99,5%, sisanya yaitu 0,5% dipengaruhi oleh variabel lain diluar persamaan regresi ini.

Tabel 4. ANOVA

	<i>Model</i>	<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig.</i>
1	Regression	465,174	4	116293	1594,550	0,000 ^b
	Residual	2,115	29	0,073		
	Total	467,289	33			

Berdasarkan Tabel 4, metode yang digunakan untuk melakukan uji hipotesis dan uji F adalah sebagai berikut:

- Berdasarkan nilai signifikansi
Berdasarkan Tabel 4, diketahui nilai signifikansi sebesar 0,000. Karena nilai signifikansi $< 0,05$, maka H_0 diterima yang berarti terdapat pengaruh signifikan variabel independen terhadap variabel dependen.
- Berdasarkan perbandingan nilai F hitung dengan F tabel
Berdasarkan Tabel 4, diketahui nilai F_{hitung} sebesar 1594,55 dan nilai $F_{(4;29)}$ yaitu 2,701. Karena nilai $F_{hitung} > F_{(4;29)}$, maka H_0 diterima yang berarti terdapat pengaruh signifikan variabel independen terhadap variabel dependen.

Tabel 5. Nilai t dan nilai signifikansi

	<i>Model</i>	<i>Unstandardized Coefficients</i>		<i>Standardized Coefficients</i>	<i>t</i>	<i>Sig.</i>
		<i>B</i>	<i>Std. Error</i>	<i>Beta</i>		
1	(<i>Constant</i>)	-4,408	2,550		-1,729	0,094
	Usia Harapan Hidup SP2020	0,610	0,037	0,311	16,337	0,000
	Rata-Rata Lama Sekolah	1,108	0,073	0,268	15,181	0,000
	Harapan Lama Sekolah	1,127	0,074	0,221	15,187	0,000
	Pengeluaran Perkapita	0,001	0,000	0,457	21,450	0,000

Berdasarkan Tabel 5, diperoleh persamaan regresi linear berganda sebagai berikut:

$$Y = -4,408 + 0,61 X_1 + 1,108 X_2 + 1,127 X_3 + 0,001 X_4 \quad (13)$$

Model regresi ini menunjukkan hubungan linier antara variabel dependen Y dan 4 variabel independen $X_1, X_2, X_3,$ dan X_4 . Koefisien pada masing-masing variabel independen menggambarkan tingkat perubahan nilai Y terkait dengan perubahan satu unit pada masing-masing variabel independen, dengan asumsi variabel lainnya tidak berubah.

Metode yang digunakan untuk mengambil keputusan uji t parsial dalam analisis regresi linear berganda adalah sebagai berikut:

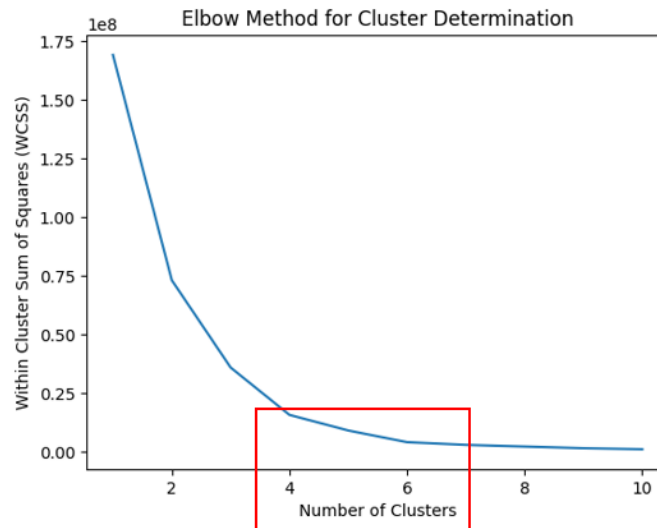
1. Uji t parsial variabel X_1
 - a) Berdasarkan nilai signifikansi
Berdasarkan Tabel 5, diketahui nilai signifikansi dari variabel Usia Harapan Hidup (X_1) sebesar 0,000. Karena nilai signifikansi $< 0,05$, artinya terdapat pengaruh signifikan Usia Harapan Hidup terhadap IPM.
 - b) Berdasarkan perbandingan nilai t_{hitung} dengan $t_{(0,025;29)}$
Berdasarkan Tabel 5, diketahui nilai t_{hitung} dari variabel Usia Harapan Hidup (X_1) sebesar 16,337 dan nilai $t_{(0,025;29)}$ yaitu 2,045. Karena nilai $t_{hitung} > t_{(0,025;29)}$, artinya terdapat pengaruh signifikan Usia Harapan Hidup terhadap IPM.
2. Uji t parsial variabel X_2
 - a) Berdasarkan nilai signifikansi
Berdasarkan Tabel 5, diketahui nilai signifikansi dari variabel Rata-Rata Lama Sekolah (X_2) sebesar 0,000. Karena nilai signifikansi $< 0,05$, artinya terdapat pengaruh signifikan Rata-Rata Lama Sekolah terhadap IPM.
 - b) Berdasarkan perbandingan nilai t_{hitung} dengan $t_{(0,025;29)}$
Berdasarkan Tabel 5, diketahui nilai t_{hitung} dari variabel Rata-Rata Lama Sekolah (X_2) sebesar 15,181 dan nilai $t_{(0,025;29)}$ yaitu 2,045. Karena nilai $t_{hitung} > t_{(0,025;29)}$, artinya terdapat pengaruh signifikan Rata-Rata Lama Sekolah terhadap IPM.
3. Uji t parsial variabel X_3
 - a) Berdasarkan nilai signifikansi
Berdasarkan Tabel 5, diketahui nilai signifikansi dari variabel Harapan Lama Sekolah (X_3) sebesar 0,000. Karena nilai signifikansi $< 0,05$, artinya terdapat pengaruh signifikan Harapan Lama Sekolah terhadap IPM.
 - b) Berdasarkan perbandingan nilai t_{hitung} dengan $t_{(0,025;29)}$
Berdasarkan Tabel 5, diketahui nilai t_{hitung} dari variabel Harapan Lama Sekolah (X_3) sebesar 15,187 dan nilai $t_{(0,025;29)}$ yaitu 2,045. Karena nilai $t_{hitung} > t_{(0,025;29)}$, artinya terdapat pengaruh signifikan Harapan Lama Sekolah terhadap IPM.
- c) Uji t parsial variabel X_4
 - a) Berdasarkan nilai signifikansi
Berdasarkan Tabel 5, diketahui nilai signifikansi dari variabel Pengeluaran per Kapita (X_4) sebesar 0,000. Karena nilai signifikansi $< 0,05$, artinya terdapat pengaruh signifikan Pengeluaran per Kapita terhadap variabel IPM.
 - b) Berdasarkan perbandingan nilai t_{hitung} dengan $t_{(0,025;29)}$
Berdasarkan Tabel 5, diketahui nilai t_{hitung} dari variabel Pengeluaran per Kapita (X_4) sebesar 21,45 dan nilai $t_{(0,025;29)}$ yaitu 2,045. Karena nilai $t_{hitung} > t_{(0,025;29)}$, artinya terdapat pengaruh signifikan Pengeluaran per Kapita terhadap variabel IPM.

Berdasarkan hasil uji t parsial bahwa semua variabel independen secara signifikan memengaruhi variabel dependen dalam model regresi linear berganda yang digunakan untuk menganalisis IPM.

Model regresi linier berganda yang dikembangkan berhasil menjelaskan variabilitas IPM dengan baik berdasarkan variabel-variabel independen yang digunakan. Semua asumsi regresi yaitu normalitas residual, independensi residual, dan homoskedastisitas telah terpenuhi. Secara keseluruhan, model ini signifikan dan memiliki kemampuan prediktif yang sangat baik, seperti yang ditunjukkan oleh nilai *R Square* dan *Adjusted R Square* yang tinggi. Semua variabel independen dalam model terbukti signifikan dan memberikan dampak positif terhadap IPM. Hal ini menunjukkan bahwa peningkatan pada variabel-variabel independen tersebut dapat berkontribusi pada peningkatan IPM.

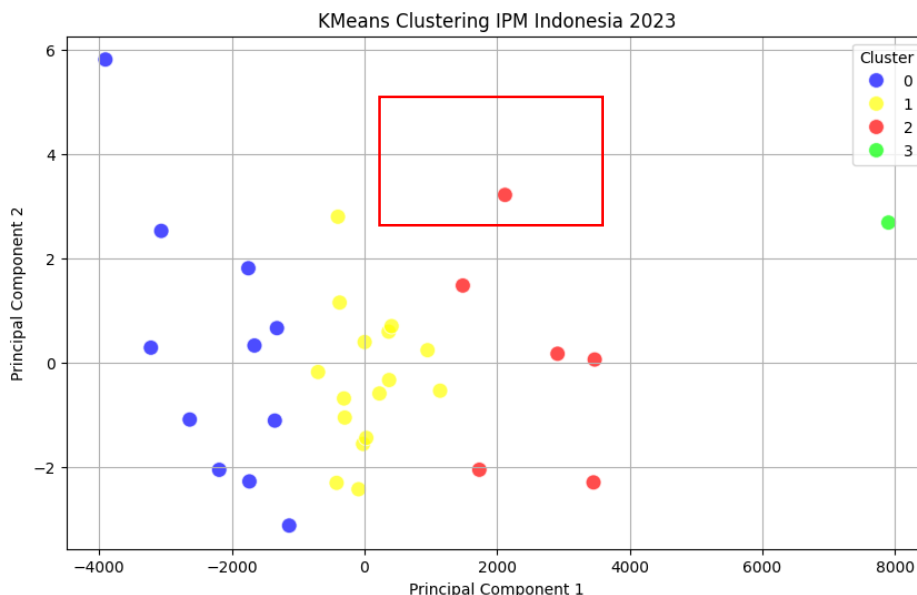
3.4 Pengelompokan Provinsi

Pengelompokan provinsi menggunakan metode *k-means clustering* dijelaskan secara rinci di bagian berikut. Melalui metode ini, provinsi-provinsi dikelompokkan menjadi beberapa kelompok yang menunjukkan kesamaan dalam variabel-variabel yang dianalisis.



Gambar 4. Hasil *Elbow Method*

Berdasarkan Gambar 4, hasil pencarian jumlah kelompok menggunakan *elbow method* didapat bahwa titik empat adalah titik di mana penurunan *Within-Cluster Sum of Squares* (WCSS) mulai berkurang secara drastis. Titik empat menunjukkan jumlah kelompok yang optimal, di mana penambahan kelompok lebih lanjut tidak memberikan peningkatan yang signifikan dalam homogenitas kelompok. Maka dari itu akan dibagi ke dalam empat kelompok sesuai dengan kemiripan dari variabel-variabel yang di analisis.



Gambar 5. Persebaran hasil dari *k-means clustering* dengan PCA

Berdasarkan Gambar 5 hasil *clustering* menggunakan PCA terdapat empat kelompok yang dibedakan dengan warna yang berbeda, yaitu biru, kuning, merah, dan hijau. Kelompok warna biru menggambarkan provinsi dengan kategori rendah, kelompok warna kuning

menggambarkan provinsi dengan kategori sedang, kelompok warna merah menggambarkan provinsi dengan kategori tinggi, dan kelompok warna hijau menggambarkan provinsi dengan kategori sangat tinggi.

Tabel 6. Pengelompokan Provinsi

Kelompok Rendah	Kelompok Sedang	Kelompok Tinggi	Kelompok Sangat Tinggi
Aceh	Sumatera Utara	Kep. Bangka Belitung	DKI Jakarta
Nusa Tenggara Timur	Sumatera Barat	Kep. Riau	
Kalimantan Barat	Riau	DI Yogyakarta	
Kalimantan Utara	Jambi	Bali	
Sulawesi Tengah	Sumatera Selatan	Kalimantan Selatan	
Sulawesi Tenggara	Bengkulu	Kalimantan Timur	
Sulawesi Barat	Lampung		
Maluku	Jawa Barat		
Maluku Utara	Jawa Tengah		
Papua Barat	Jawa Timur		
Papua	Banten		
	Nusa Tenggara Barat		
	Kalimantan Tengah		
	Sulawesi Utara		
	Sulawesi Selatan		
	Gorontalo		

Pada Tabel 6 terlihat bahwa pengelompokan provinsi di atas memberikan gambaran kesenjangan pembangunan manusia di Indonesia. Peningkatan kualitas layanan dasar seperti kesehatan, pendidikan, dan ekonomi di seluruh provinsi sangat penting untuk mencapai pembangunan yang lebih merata. Provinsi dalam kelompok rendah memiliki beberapa kendala seperti akses yang terbatas terhadap layanan kesehatan dan pendidikan berkualitas, infrastruktur yang kurang memadai, dan ekonomi yang kurang berkembang. Tantangan geografis seperti wilayah yang terpencil atau terpisah oleh laut juga dapat mempersulit pengembangan SDM di daerah tersebut. Provinsi dalam sedang menunjukkan perkembangan yang lebih baik dibandingkan dengan kelompok rendah. Namun, masih menghadapi tantangan seperti ketimpangan akses layanan kesehatan dan pendidikan antara daerah perkotaan dan pedesaan, atau masalah ekonomi yang tidak merata. Provinsi dalam kelompok tinggi sudah menunjukkan keberhasilan yang signifikan dalam pembangunan SDM, karena kombinasi akses yang lebih baik terhadap pendidikan dan layanan kesehatan, serta infrastruktur yang lebih maju dan ekonomi yang lebih kuat. DKI Jakarta berada pada kelompok sangat tinggi, yang menunjukkan kualitas SDM yang unggul, karena akses yang sangat baik ke layanan pendidikan dan kesehatan berkualitas, serta ekonomi yang sangat berkembang.

4 KESIMPULAN

Berdasarkan hasil yang diperoleh, bahwa data yang digunakan telah memenuhi asumsi-asumsi penting dalam regresi linear, yaitu normalitas residual dan homoskedastisitas, sehingga analisis regresi linear berganda dapat dilakukan. Hasil analisis menunjukkan bahwa variabel Usia Harapan Hidup (X_1), Rata-Rata Lama Sekolah (X_2), Harapan Lama Sekolah (X_3), Pengeluaran per Kapita (X_4) memiliki pengaruh signifikan terhadap Indeks Pembangunan Manusia (Y) sebesar 99,5%. Dari segi nilai signifikansi, nilai F , dan nilai t semua variabel independen memiliki pengaruh signifikan Indeks Pembangunan Manusia. Hubungan yang kuat

antara variabel-variabel tersebut mengindikasikan bahwa upaya peningkatan kualitas kesehatan, pendidikan, dan ekonomi sangat penting untuk mencapai pembangunan manusia yang lebih baik.

Pengelompokan provinsi menggunakan *elbow method* dalam *k-means clustering* menghasilkan empat kelompok, yaitu rendah, sedang, tinggi, dan sangat tinggi. Provinsi dalam kelompok rendah menghadapi kendala dalam akses ke layanan kesehatan dan pendidikan, infrastruktur, serta ekonomi yang kurang berkembang. Provinsi dalam kelompok sedang menunjukkan perkembangan yang lebih baik tetapi masih menghadapi ketimpangan layanan antara perkotaan dan pedesaan. Kelompok tinggi telah mencapai kemajuan signifikan dalam pembangunan SDM dengan akses yang lebih baik terhadap layanan dan ekonomi yang lebih kuat. DKI Jakarta, dalam kelompok sangat tinggi, menunjukkan kualitas SDM unggul dengan akses yang baik ke layanan dan ekonomi yang sangat berkembang.

Pemerintah perlu mengimplementasikan strategi peningkatan kualitas SDM berdasarkan kebutuhan masing-masing provinsi. Untuk kelompok rendah, perlu ditingkatkan infrastruktur dan layanan dasar. Kelompok sedang membutuhkan pengurangan ketimpangan layanan dan dukungan ekonomi lokal. Kelompok tinggi perlu menjaga dan meningkatkan kualitas layanan serta inovasi ekonomi. Sementara DKI Jakarta harus fokus pada menangani tantangan urbanisasi dan memastikan manfaat pembangunan dirasakan oleh semua lapisan masyarakat. Partisipasi aktif masyarakat dan dukungan terhadap kebijakan yang inklusif juga penting untuk meningkatkan kualitas hidup di seluruh wilayah Indonesia.

Meskipun penelitian ini menunjukkan pengaruh signifikan dari beberapa variabel, diperlukan penelitian lebih lanjut untuk mengeksplorasi faktor lain seperti kebijakan pemerintah, kondisi global, dan faktor geografis serta sosial budaya yang mungkin juga memengaruhi IPM. Kebijakan pembangunan manusia di Indonesia harus mempertimbangkan kondisi masyarakat setempat untuk mencapai hasil yang lebih efektif.

UCAPAN TERIMAKASIH

Penulis ingin menyampaikan rasa terima kasih kepada Ibu Dr. Pismia Sylvi, S.Si., M.Si. dari Universitas Terbuka, selaku dosen pembimbing mata kuliah Karya Ilmiah, atas bimbingan, dukungan, dan nasihat yang sangat berharga selama penyusunan karya ilmiah ini. Penulis juga mengucapkan terima kasih kepada seluruh keluarga yang telah memberikan motivasi dan dukungan moral yang tak ternilai sepanjang proses penyusunan karya ilmiah ini. Ucapan terima kasih juga ditujukan kepada teman-teman yang telah memberikan semangat kontribusi dalam bentuk saran dan masukan yang membangun. Karya Ilmiah “Analisis Faktor Terhadap Indeks Pembangunan Manusia dan Pengelompokan Provinsi di Indonesia Tahun 2023” ini tentu tidak akan terwujud tanpa bantuan dan dukungan dari berbagai pihak. Setiap kekurangan dalam karya ilmiah ini sepenuhnya menjadi tanggung jawab penulis. Penulis berharap hasil karya ilmiah ini dapat memberikan kontribusi positif bagi perkembangan ilmu pengetahuan dan kebijakan pembangunan di Indonesia.

DAFTAR PUSTAKA

- Afriyana, L., Salmah, E., Sriningsih, S., & Harsono, I. (2023). Analisis Dampak Pembangunan Infrastruktur Terhadap Pertumbuhan Ekonomi Inklusif Pada Kabupaten/Kota Di Provinsi Nusa Tenggara Barat Tahun 2016-2021. *Elastisitas: Jurnal Ekonomi Pembangunan*, 5(1), 1-12.
- BPS. (2023). *Berita Resmi Statistik No. 80/11/Th.XXVI*.
- Ernita, T., Sutan Nasution, A., Aditya, B., Tinggi Teknologi Industri Padang, S., Hamka, J., Tabing, P., & Koto Tangah, K. (2022). Analisis keselamatan dan kesehatan kerja terhadap produktivitas menggunakan regresi linear berganda di PT. Lembah Karet. *Jurnal Sains dan Teknologi Keilmuan dan Aplikasi Teknologi Industri*, 22(2), 408-416. DOI:

10.36275/stsp.v22i2.553.

- Hermanto, E. M. P., Athoillah, M., Hamidah, W., & Putra, D. P. (2021). Pelatihan Penggunaan Software R Untuk Menguji Perbandingan Berganda dan Asumsi Residual pada Rancangan Percobaan. *Jurnal Pengabdian Kepada Masyarakat*, 1, 449–458. <https://doi.org/https://doi.org/10.53625/jabdi.v1i4.238>
- Kanio Tubani, D., & Hadijah Hasanah, S. (2022). *Pengelompokan Kabupaten/Kota di Provinsi Nusa Tenggara Timur Menurut Tingkat Kesejahteraan, Tahun 2022*.
- Kurniawan, B., Hartono, S., Kosasih, Putra, S. A., Yuliyanti, E., Arianty, D., Sumiati, Metris, D., Librianty, N., Mustahidda, R., Kartadiwirya, K. B., Badrianto, Y., & Turmudhi, A. (2023). *Manajemen Sumber Daya Manusia*, 1(1) (2023). <https://repo.alungcipta.com/index.php/repo/article/view/11/5>
- Kusuma, M. E., & Muta'ali, L. (2019). *Hubungan Pembangunan Infrastruktur dan Perkembangan Ekonomi Wilayah Indonesia*. Skripsi. Universitas Gadjah Mada: Yogyakarta.
- Quraisy, A. (2020). Normalitas Data Menggunakan Uji Kolmogorov-Smirnov dan Saphiro-Wilk. *Journal of Health, Education, Economics, Science, and Technology*, 3, 7–11.
- Rosida, W., & Wijaya, Y. A. (2023). Klasterisasi Penyakit HIV/AIDS di Jawa Barat Menggunakan Algoritma K-Means Clustering. *Blend Sains Jurnal Teknik*, 1(4), 306–315. <https://doi.org/10.56211/blendsains.v1i4.235>
- Saputra, D. M., Saputra, D., & Oswari, L. D. (2020). *Effect of Distance Metrics in Determining K-Value in KMeans Clustering Using Elbow and Silhouette Method. Proceedings of the Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)*. 10.2991/aisr.k.200424.051
- Ummah, S. (2024). Analisis Deskriptif Indeks Pembangunan Manusia Indonesia Periode Tahun 2020-2023. *Jurnal Mahasiswa Humanis*, 4(1), 41.
- Widodo, E., Mashita, S. N., & Prasetyowati, Y. G. (2020). Perbandingan Metode Average Linkage, Complete Linkage, dan Ward'S pada Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Indeks Pembangunan Manusia. *Faktor Exacta*, 13(2), 81. <https://doi.org/10.30998/faktorexacta.v13i2.6581>
- Yafi, M., Goejantoro, R., & Tri Rian Dani. (2023). *Pengelompokan Algoritma K-Medoids Dengan Principal Component Analysis (PCA) (Studi Kasus : Kabupaten/Kota di Pulau Kalimantan Berdasarkan Indikator Kemiskinan Tahun 2021)*. *Prosiding Seminar Nasional Matematika dan Statistika*, 3(1), 183-195.