Klasterisasi Provinsi di Indonesia Berdasarkan Faktor Pendukung Transformasi Digital Pendidikan Tinggi Menggunakan Metode *K-Means Clustering*

N K Juliarini[*1], A A G W S Erlangga[2], N N E Smrti[3]

[1,2,3] STMIK Bandung Bali, Jl. Gatot Subroto Timur No. 507, Denpasar Timur, Indonesia

E-mail: juliarini.kdk@gmail.com[1], wahyujungde@gmail.com[2], smrti@yacanet.com[3]

Abstrak. Transformasi digital dalam pendidikan tinggi berkembang dengan kondisi yang berbeda antarprovinsi di Indonesia. Penelitian ini bertujuan mengelompokkan provinsi berdasarkan karakteristik pendukung transformasi digital pendidikan tinggi menggunakan metode *K-Means Clustering*. Data yang digunakan merupakan data tahun 2024 dari 34 provinsi dengan lima variabel, yaitu penggunaan internet, penggunaan komputer, rasio PDRB sektor informasi dan komunikasi, jumlah dosen, dan rata-rata lama sekolah. Hasil analisis membentuk tiga *cluster*, yaitu *Cluster* 1 yang terdiri atas DI Yogyakarta, *Cluster* 2 yang terdiri atas 29 provinsi, dan *Cluster* 3 yang terdiri atas DKI Jakarta, Jawa Barat, Jawa Tengah, dan Jawa Timur. *Cluster* 1 memiliki nilai relatif tinggi pada seluruh variabel, *Cluster* 2 relatif rendah, sedangkan *Cluster* 3 menonjol pada rasio PDRB sektor informasi dan komunikasi serta jumlah dosen, dengan penggunaan internet dan komputer yang relatif lebih tinggi. Evaluasi menggunakan *Silhouette Coefficient* menghasilkan nilai rata-rata sebesar 0,6254, yang menunjukkan pemisahan *cluster* yang cukup baik. Hasil penelitian menggambarkan perbedaan karakteristik antarprovinsi sebagai dasar pengembangan strategi transformasi digital pendidikan tinggi sesuai kondisi wilayah.

Kata kunci: Transformasi Digital; Pendidikan Tinggi; Variabel Digital; Klasterisasi; *K-Means*

Abstract. Digital transformation in higher education develops under different conditions across provinces in Indonesia. This study aims to classify provinces based on the characteristics supporting digital transformation in higher education using the *K-Means Clustering* method. The data consist of 2024 data from 34 provinces with five variables: internet usage, computer usage, the share of the information and communication sector in GRDP, the number of lecturers, and mean years of schooling. The analysis produces three clusters: *Cluster* 1 consisting of DI Yogyakarta, *Cluster* 2 consisting of 29 provinces, and *Cluster* 3 consisting of DKI Jakarta, West Java, Central Java, and East Java. *Cluster* 1 has relatively high values across all variables, *Cluster* 2 has relatively low values, while *Cluster* 3 is more prominent in the share of the information and communication sector in GRDP and the number of lecturers, with relatively higher internet and computer usage. Evaluation using the *Silhouette Coefficient* yields an average value of 0.6254, indicating reasonably good cluster separation. The findings illustrate differences in provincial characteristics that can serve as a basis for developing digital transformation strategies in higher education according to regional conditions.

Keywords: Digital Transformation; Higher Education; Digital Indicators; Clustering; K-Means

1. Pendahuluan

Transformasi digital semakin banyak diterapkan dalam pendidikan tinggi karena teknologi dapat mendukung proses pembelajaran melalui berbagai perangkat digital dan *platform e-learning* [1]. Pemanfaatan teknologi juga mencakup pengelolaan data akademik, komunikasi, pelayanan mahasiswa, dan kegiatan administrasi [2]. Dengan demikian, teknologi menjadi salah satu bagian yang mendukung berbagai aktivitas di perguruan tinggi. Namun, kondisi yang mendukung pemanfaatan teknologi tersebut dapat berbeda antarwilayah. Perbedaan akses dan penggunaan teknologi, ketersediaan tenaga pengajar, kondisi ekonomi, dan pendidikan dapat membentuk karakteristik wilayah yang berbeda [3]. Penelitian mengenai transformasi digital pada pendidikan tinggi juga menunjukkan bahwa penerapannya berkaitan dengan beberapa aspek dalam lingkungan perguruan tinggi, sehingga tidak dapat dilihat hanya dari penggunaan teknologi [4].

Perbedaan karakteristik tersebut dapat dilihat pada tingkat provinsi di Indonesia. Setiap provinsi memiliki kondisi yang berbeda dalam penggunaan teknologi, pendidikan, dan ekonomi [5]. Data yang digunakan dalam penelitian ini menunjukkan bahwa persentase penduduk yang menggunakan internet berkisar antara 54,84% hingga 90,10%, sedangkan persentase penduduk yang menggunakan komputer berada pada kisaran 10,89% hingga 51,55%. Jumlah dosen juga menunjukkan variasi yang cukup besar, yaitu antara 480 hingga 32.099 orang [6]. Variasi tersebut menunjukkan bahwa kondisi pendukung pemanfaatan teknologi pada pendidikan tinggi tidak seragam antarprovinsi.

Perbedaan tersebut menjadi lebih kompleks ketika beberapa indikator dipertimbangkan secara bersamaan. Nilai yang tinggi pada satu indikator tidak selalu diikuti oleh nilai yang tinggi pada indikator lainnya. Oleh karena itu, karakteristik suatu provinsi perlu dilihat berdasarkan kombinasi beberapa variabel agar pola kemiripan dan perbedaannya dapat diketahui dengan lebih baik. Pendekatan ini relevan karena pemanfaatan teknologi dalam lingkungan pendidikan tinggi juga berkaitan dengan ketersediaan sumber daya manusia dan kondisi wilayah yang mendukung penggunaannya [7].

Berdasarkan pertimbangan tersebut, penelitian ini menggunakan lima variabel yang mewakili beberapa aspek pendukung pemanfaatan teknologi pada pendidikan tinggi. Aspek teknologi direpresentasikan oleh persentase penduduk yang menggunakan internet dan komputer [8]. Penggunaan internet memberikan gambaran mengenai pemanfaatan teknologi digital, sedangkan penggunaan komputer memberikan informasi mengenai penggunaan perangkat komputer [9]. Aspek pendidikan tinggi direpresentasikan oleh jumlah dosen sebagai salah satu indikator ketersediaan tenaga pengajar. Selain itu, aspek ekonomi wilayah dilihat melalui rasio PDRB sektor informasi dan komunikasi terhadap total PDRB [10]. Sedangkan aspek pendidikan masyarakat direpresentasikan oleh rata-rata lama sekolah [11].

Kelima variabel tersebut tidak dimaksudkan untuk mengukur atau menetapkan tingkat transformasi digital suatu provinsi. Variabel-variabel tersebut digunakan untuk menggambarkan beberapa kondisi yang berkaitan dengan dukungan terhadap pemanfaatan teknologi dalam lingkungan pendidikan tinggi. Dengan batasan tersebut, hasil penelitian diarahkan untuk melihat kemiripan karakteristik antarprovinsi berdasarkan indikator yang digunakan, bukan untuk memberikan penilaian menyeluruh mengenai tingkat transformasi digital pendidikan tinggi di Indonesia.

Data penelitian berbentuk multivariat, dengan setiap provinsi direpresentasikan oleh lima variabel numerik. Karena kelima variabel memiliki satuan dan rentang nilai yang berbeda, perbandingan antarprovinsi perlu dilakukan setelah variabel berada pada skala yang sama. Selanjutnya, kedekatan antarprovinsi dapat dihitung berdasarkan seluruh variabel secara bersamaan. Jarak *Euclidean* digunakan untuk mengukur kedekatan tersebut karena dapat menggambarkan perbedaan karakteristik antarobjek berdasarkan beberapa variabel numerik [12]. Hasil pengukuran jarak kemudian digunakan sebagai dasar dalam proses pengelompokan menggunakan metode *K-Means Clustering*.

K-Means Clustering merupakan metode klasterisasi yang mengelompokkan objek berdasarkan kemiripan karakteristik, sehingga objek yang memiliki karakteristik relatif dekat ditempatkan dalam kelompok yang sama [13]. Metode ini telah digunakan dalam beberapa penelitian pendidikan untuk mengelompokkan wilayah berdasarkan sejumlah indikator pendidikan [14],[15],[16]. Penelitian-penelitian tersebut umumnya berfokus pada indikator pendidikan, seperti capaian atau kondisi pendidikan. Sementara itu, penelitian yang secara khusus menggabungkan indikator penggunaan

teknologi, kondisi pendidikan tinggi, ekonomi wilayah, dan pendidikan masyarakat dalam melihat kemiripan karakteristik antarprovinsi masih terbatas.

Atas dasar tersebut, penelitian ini mengelompokkan provinsi di Indonesia berdasarkan lima indikator yang mencakup penggunaan internet, penggunaan komputer, jumlah dosen, rasio PDRB sektor informasi dan komunikasi, serta rata-rata lama sekolah. Pengelompokan dilakukan untuk mengidentifikasi provinsi yang memiliki karakteristik relatif serupa berdasarkan indikator-indikator tersebut. Hasil klasterisasi diharapkan dapat memberikan gambaran mengenai pola karakteristik antarprovinsi dan menunjukkan kelompok wilayah yang memiliki kondisi pendukung pemanfaatan teknologi pada pendidikan tinggi yang relatif sama. Dengan demikian, hasil penelitian dapat menjadi informasi awal untuk memahami variasi karakteristik wilayah tanpa menggeneralisasikan hasil klaster sebagai ukuran tingkat transformasi digital pendidikan tinggi secara keseluruhan.

2. Landasan Teori

2.1. Analisis korelasi

Analisis korelasi digunakan untuk melihat hubungan antara dua variabel. Nilai korelasi berada pada rentang -1 sampai 1. Nilai positif menunjukkan hubungan searah, sedangkan nilai negatif menunjukkan hubungan berlawanan arah. Semakin mendekati 1 atau -1, semakin kuat hubungan antara kedua variabel. Koefisien korelasi *Pearson* dapat dihitung dengan menggunakan Persamaan 1 [17].

$$r_{jk} = \frac{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2 \sum_{i=1}^{n}(x_{ik} - \bar{x}_k)^2}} \quad (1)$$

di mana x_{ij} merupakan nilai variabel ke-j pada provinsi ke-i, x_{ik} merupakan nilai variabel ke-k pada provinsi ke-i, dan n merupakan jumlah provinsi.

2.2. Uji KMO dan Bartlett

Uji *Kaiser-Meyer-Olkin* (KMO) dan Bartlett digunakan untuk melihat hubungan antarvariabel sebelum analisis dilakukan secara bersama-sama. Nilai KMO yang lebih besar dari 0,50 menunjukkan bahwa hubungan antarvariabel cukup memadai. Uji Bartlett digunakan untuk melihat apakah terdapat hubungan yang signifikan antarvariabel. Jika $p - value < 0{,}05$, berarti terdapat hubungan yang signifikan antarvariabel [18].

2.3. Standardisasi Data

Standardisasi dilakukan agar variabel yang memiliki satuan dan rentang nilai berbeda dapat digunakan secara seimbang dalam perhitungan jarak. Pada penelitian ini digunakan standardisasi *z-score* dengan rumus pada Persamaan 2 [19].

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (2)$$

di mana z_{ij} merupakan nilai standar provinsi ke-i pada variabel ke-j, x_{ij} adalah nilai awal, $\bar{x}_j$ adalah rata-rata variabel ke-j, dan s_j adalah simpangan baku.

2.4. Jarak Euclidean

Jarak Euclidean digunakan untuk mengukur tingkat kedekatan antarprovinsi berdasarkan keenam variabel yang telah distandarisasi. Jarak antara provinsi ke-i dan provinsi ke-j dihitung dengan menggunakan Persamaan 3 [12].

$$d_{ij} = \sqrt{\left(z_{i1} - z_{j1}\right)^2 + \left(z_{i2} - z_{j2}\right)^2 + \cdots + \left(z_{i6} - z_{j6}\right)^2} \quad (3)$$

di mana d_{ij} adalah jarak antara provinsi ke-i dan provinsi ke-j, sedangkan z_{ik} dan z_{jk} merupakan nilai *z-score* variabel ke-k pada masing-masing provinsi.

2.5. K-Means Clustering

K-Means Clustering digunakan untuk mengelompokkan provinsi berdasarkan kemiripan karakteristiknya. Pada penelitian ini digunakan $k = 3$, sehingga 34 provinsi dikelompokkan menjadi tiga *cluster*. Pusat setiap *cluster* atau *centroid* dihitung berdasarkan rata-rata nilai standar dari seluruh anggota *cluster*. *Centroid* untuk variabel ke-j pada *cluster* ke-c dihitung dengan Persamaan 4 [20].

$$\bar{z}_{cj} = \frac{1}{n_c} \sum_{i \in C_c} z_{ij} \tag{4}$$

di mana $\bar{z}_{cj}$ merupakan *centroid cluster* ke-c pada variabel ke-j, C_c merupakan kumpulan provinsi yang berada pada *cluster* ke-c, n_c merupakan provinsi dalam *cluster* ke-c, dan z_{ij} merupakan nilai standar variabel ke-j pada provinsi ke-i. Setiap provinsi kemudian ditempatkan ke dalam *cluster* dengan *centroid* yang memiliki jarak paling dekat. Proses tersebut dilakukan berulang sampai pembagian *cluster* tidak mengalami perubahan yang berarti.

3. Metode Penelitian

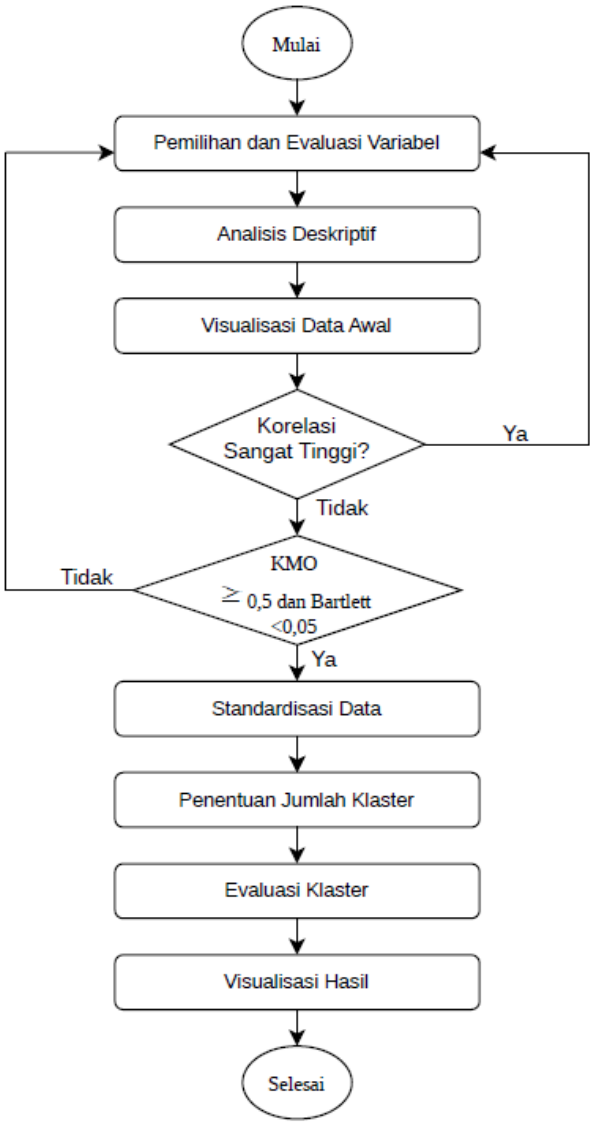
3.1 Data Penelitian

Penelitian ini menggunakan data sekunder tahun 2024 dengan unit analisis berupa 34 provinsi di Indonesia. Data yang digunakan terdiri atas lima variabel, yaitu persentase penduduk yang menggunakan internet, persentase penduduk yang menggunakan komputer, rasio PDRB sektor informasi dan komunikasi terhadap total PDRB, jumlah dosen, dan rata-rata lama sekolah.

Data diperoleh dari *website* Badap Pusat Statistik (BPS) Indonesia yang menyediakan informasi terkait kondisi sosial, ekonomi, pendidikan, dan teknologi di setiap provinsi. Dengan demikian, data penelitian terdiri atas 34 observasi dan 5 variabel. Sebelum dilakukan analisis, data diperiksa terlebih dahulu untuk memastikan kelengkapan setiap variabel. Hasil pemeriksaan menunjukkan bahwa tidak terdapat *missing value*, sehingga seluruh data 34 provinsi dapat digunakan dalam tahap analisis klasterisasi menggunakan *K-Means Clustering*.

3.2 Langkah Analisis

Langkah analisis dalam penelitian ini ditampilkan melalui *flowchart* pada Gambar 1.

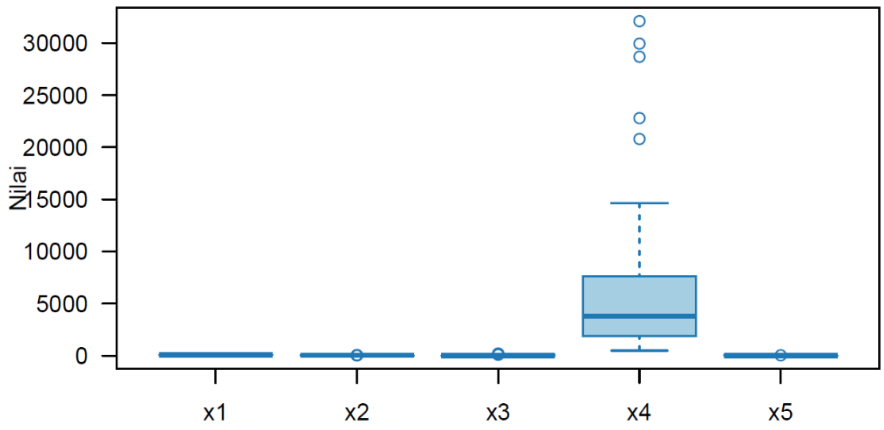


Gambar 1. Langkah Analisis Data

4. Hasil dan Pembahasan

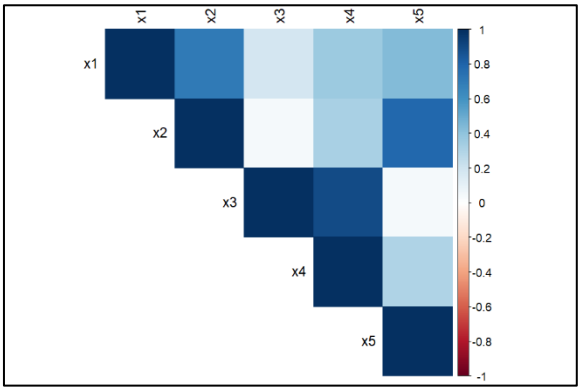
Sebelum dilakukan proses klasterisasi, selanjutnya dilakukan analisis deskriptif untuk melihat gambaran kondisi setiap variabel pada 34 provinsi. Hasil analisis menunjukkan bahwa kelima variabel memiliki tingkat variasi yang berbeda antarprovinsi. Penggunaan internet memiliki rata-rata sebesar 71,94%, dengan nilai berkisar antara 54,84% hingga 90,10%. Sementara itu, penggunaan komputer memiliki rata-rata 20,32% dengan rentang nilai 10,89% hingga 51,55%. Perbedaan ini menunjukkan bahwa pemanfaatan teknologi digital masyarakat masih belum sama pada setiap provinsi. Variasi yang sangat jelas terlihat pada jumlah dosen yang berada pada rentang 480 hingga 32.099 orang.

Perbedaan karakteristik tersebut menunjukkan bahwa setiap provinsi memiliki kondisi pendukung transformasi digital pendidikan tinggi yang berbeda. Oleh karena itu, kelima variabel tersebut dianalisis secara bersama-sama untuk mengidentifikasi provinsi yang memiliki karakteristik relatif sama melalui metode *K-Means Clustering*. Perbedaan rentang nilai pada masing-masing variabel juga terlihat melalui visualisasi awal data penelitian yang ditampilkan pada Gambar 2 berikut.



Gambar 2. Distribusi Awal Variabel Penelitian

Berdasarkan Gambar 2, terlihat bahwa variabel jumlah dosen (x_4) memiliki skala nilai yang jauh lebih besar dibandingkan variabel lainnya. Sementara itu, variabel penggunaan internet (x_1), penggunaan komputer (x_2), rasio PDRB sektor informasi dan komunikasi (x_3), dan rata-rata lama sekolah (x_5) memiliki rentang nilai yang lebih kecil. Selain melihat perbedaan skala antarvariabel, dilakukan juga analisis korelasi untuk melihat hubungan antarvariabel yang digunakan. Analisis ini dilakukan untuk mengetahui apakah terdapat hubungan antara satu variabel dengan variabel lainnya sebelum proses klasterisasi dilakukan.



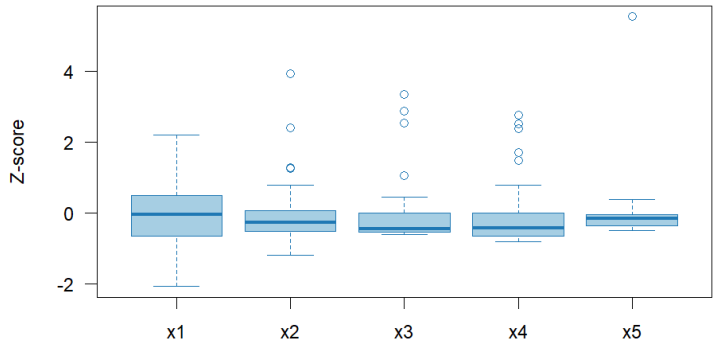
Gambar 3. Matriks Korelasi Variabel Penelitian

Berdasarkan Gambar 3, terdapat beberapa variabel yang memiliki hubungan cukup kuat. Korelasi tertinggi terlihat antara rasio PDRB sektor informasi dan komunikasi (x_3) dengan jumlah dosen (x_4) sebesar 0,897. Selain itu, variabel penggunaan internet (x_1) dan penggunaan komputer (x_2) memiliki korelasi sebesar 0,7, sedangkan variabel penggunaan komputer (x_2) dan rata-rata lama sekolah (x_5) memiliki korelasi sebesar 0,787. Korelasi yang cukup tinggi menunjukkan bahwa beberapa variabel memiliki informasi yang saling berkaitan. Kondisi ini perlu diperhatikan dalam analisis *K-Means Clustering* karena pembentukan klaster didasarkan pada jarak antarobjek. Variabel yang memiliki korelasi tinggi dapat memberikan informasi yang serupa sehingga berpotensi memberikan pengaruh yang lebih besar terhadap pembentukan klaster. Sebelumnya, juga ditemukan korelasi yang sangat tinggi antara jumlah dosen dan jumlah perguruan tinggi sebesar 0,931. Kondisi tersebut telah ditinjau melalui analisis sensitivitas dan menjadi salah satu pertimbangan dalam pemilihan variabel yang digunakan pada analisis akhir. Dengan demikian, variabel yang digunakan dalam penelitian ini dipilih setelah mempertimbangkan kemungkinan adanya informasi yang tumpang tindih antarvariabel.

Selanjutnya, dilakukan uji KMO dan Bartlett untuk menilai kecukupan hubungan antarvariabel dan melihat apakah terdapat keterkaitan yang cukup dalam matriks korelasi. Hasil uji KMO menunjukkan nilai *Overall MSA* sebesar 0,56. Nilai tersebut menunjukkan bahwa kecukupan hubungan antarvariabel secara keseluruhan berada pada tingkat yang cukup. Nilai MSA masing-masing variabel berkisar antara

0,46 hingga 0,65. Nilai terendah terdapat pada rasio PDRB sektor informasi dan komunikasi (x_3), sedangkan nilai tertinggi terdapat pada penggunaan internet (x_1). Nilai MSA x_3 yang berada di bawah 0,5 menunjukkan bahwa kecukupan hubungan variabel tersebut dengan variabel lainnya relatif lebih rendah dibandingkan variabel lainnya. Meskipun demikian, nilai *Overall MSA* sebesar 0,65 menunjukkan bahwa secara keseluruhan terdapat kecukupan hubungan antarvariabel yang dapat dianalisis lebih lanjut. Hasil uji Bartlett menghasilkan nilai *chi-square* sebesar 121,4 dengan derajat bebas 10 dan *p-value* sebesar $2{,}672 \times 10^{-21}$. Nilai *p-value* yang jauh lebih kecil dari 0,05 menunjukkan bahwa hipotesis nol yang menyatakan matriks korelasi merupakan matriks identitas ditolak. Dengan demikian, terdapat hubungan yang signifikan antarvariabel dalam matriks korelasi. Secara keseluruhan, hasil uji KMO dan Bartlett menunjukkan bahwa matriks korelasi memiliki keterkaitan antarvariabel yang dapat dipertimbangkan dalam analisis lebih lanjut.

Kelima variabel kemudian distandarisasi agar perbedaan satuan dan rentang nilainya tidak memengaruhi hasil pengelompokan. Hal ini penting karena beberapa variabel memiliki satuan yang berbeda. Setelah distandarisasi, seluruh variabel berada pada skala yang sama sehingga dapat dibandingkan secara lebih seimbang dalam proses *K-Means Clustering*. Hasil standardisasi kemudian ditampilkan dalam bentuk *boxplot* pada Gambar 4.



Gambar 4. Distribusi Variabel Setelah Standarisasi

Berdasarkan Gambar 4, setelah standardisasi seluruh variabel berada pada skala yang sama dalam bentuk *z-score*, sehingga penyebaran nilai antarvariabel dapat dibandingkan dengan lebih mudah. Pusat data berada di sekitar nilai nol, meskipun terdapat beberapa nilai yang berada cukup jauh dari sebaran utama. Nilai pencilan positif paling tinggi terdapat pada variabel x_5. Kondisi tersebut menunjukkan adanya beberapa provinsi yang memiliki nilai indikator jauh lebih tinggi dibandingkan sebagian besar provinsi lainnya. Nilai-nilai tersebut tetap dipertahankan karena merupakan bagian dari karakteristik data dan dapat memberikan informasi dalam proses pengelompokan.

Setelah kelima variabel distandarisasi, langkah selanjutnya adalah menghitung jarak *Euclidean* antarprovinsi. Perhitungan ini digunakan untuk melihat tingkat kedekatan atau perbedaan karakteristik antarprovinsi berdasarkan kelima variabel yang digunakan, dengan menggunakan Persamaan (3). Salah satu hasil perhitungan ditunjukkan pada jarak antara Provinsi Aceh dan Sumatera Utara. Dengan menggunakan nilai *z-score* dari kelima variabel diperoleh:

$$d = \sqrt{(-1{,}08-0{,}27)^2+(-0{,}60-0{,}63)^2+(-0{,}32-0{,}04)^2+(-0{,}11-0{,}79)^2+(-0{,}032-0{,}030)^2}$$
$$= 2{,}0685.$$

Hasil tersebut menunjukkan bahwa jarak *Euclidean* antara Aceh dan Sumatera Utara adalah 2,0685. Nilai jarak ini selanjutnya digunakan dalam proses pengelompokan provinsi berdasarkan tingkat kemiripannya. Semakin kecil nilai jarak, semakin dekat karakteristik kedua provinsi, sedangkan semakin besar nilai jarak menunjukkan bahwa karakteristik kedua provinsi semakin berbeda.

Hasil perhitungan jarak antarprovinsi selanjutnya digunakan dalam proses *K-Means Clustering*. Sebelum menentukan jumlah *cluster*, dilakukan pengujian terhadap beberapa nilai (k), yaitu ($k = 2$) sampai ($k = 5$). Pemilihan jumlah *cluster* dilakukan dengan mempertimbangkan nilai *Silhouette* dan *Elbow Method*. Hasil pengujian disajikan pada Tabel 1.

Tabel 1. Evaluasi Jumlah *Claster*

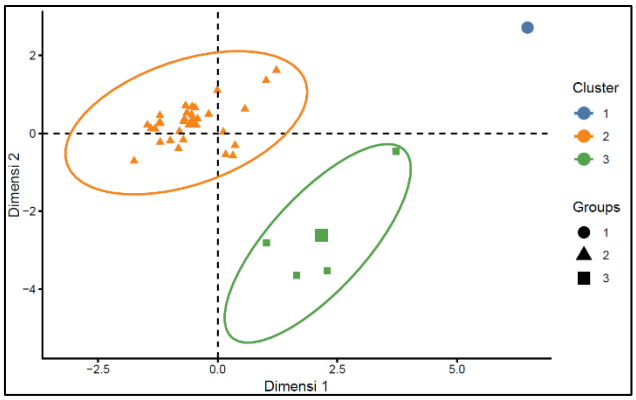
Jumlah Klaster (k)	Nilai *Silhouette*	Nilai WSS
2	0.6254466	108.36711
3	0.6253707	62.43768
4	0.3135851	44.62267
5	0.3203234	32.73212

Berdasarkan Tabel 1, nilai *Silhouette* tertinggi diperoleh pada 2 *cluster*, yaitu sebesar 0,6254466. Sementara itu, pada 3 *cluster* diperoleh nilai *Silhouette* sebesar 0,6253707. Selisih kedua nilai tersebut sangat kecil yaitu sekitar 0,00008, sehingga kualitas pengelompokan pada 2 dan 3 *cluster* relatif sebanding. Hasil *Elbow Method* menunjukkan bahwa nilai WSS mengalami penurunan yang cukup besar dari 2 *cluster* ke 3 *cluster*, yaitu dari 108,36711 menjadi 62,43768. Setelah jumlah *cluster* ditingkatkan menjadi 4 dan 5, penurunan WSS tetap terjadi, tetapi tambahan penurunannya semakin kecil. Kondisi ini menunjukkan bahwa penambahan *cluster* setelah 3 kelompok memberikan peningkatan yang semakin terbatas.

Berdasarkan hasil kedua metode tersebut, pilihan antara 2 dan 3 *cluster* tidak ditentukan hanya berdasarkan nilai *Silhouette*, tetapi juga mempertimbangkan hasil *Elbow Method* dan tujuan analisis untuk memperoleh kelompok provinsi yang dapat dibedakan secara lebih informatif. Oleh karena itu, penelitian ini menggunakan 3 *cluster*. Selanjutnya, masing-masing *cluster* memiliki sebuah pusat (*centroid*) yang menggambarkan nilai rata-rata variabel dari provinsi yang termasuk di dalamnya. Hasil estimasi *centroid* untuk masing-masing *cluster* disajikan pada Tabel 2.

Tabel 2. *Estimasi Centroid K-Means*

Variabel	***Cluster* 1**	***Cluster* 2**	***Cluster* 3**
x_1	2,19	-0,155	0,580
x_2	3,93	-0,171	0,257
x_3	0,459	-0,354	2,45
x_4	1,71	-0,373	2,28
x_5	5,54	-0,168	-0,164



Gambar 5. Hasil Klasterisasi *K-Means*

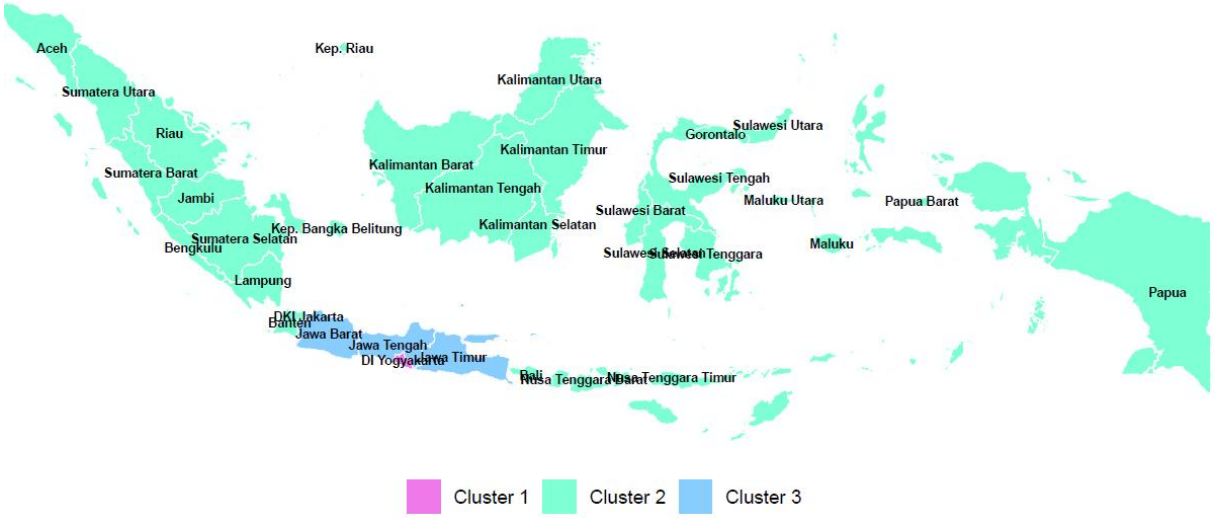
Berdasarkan Tabel 2 dan Gambar 5, terlihat adanya perbedaan karakteristik antar-*cluster* berdasarkan kombinasi lima variabel yang digunakan. *Cluster* 1 memiliki nilai *centroid* positif pada seluruh variabel, dengan nilai yang relatif tinggi pada penggunaan internet (x_1), penggunaan komputer (x_2), jumlah dosen (x_4), dan rata-rata lama sekolah(x_5). Hal ini menunjukkan bahwa *Cluster* 1 memiliki kondisi yang relatif lebih tinggi secara menyeluruh dibandingkan rata-rata keseluruhan.

Cluster 2 memiliki nilai *centroid* negatif pada seluruh variabel. Pola tersebut menunjukkan bahwa penggunaan internet (x_1) dan komputer (x_2), rasio PDRB sektor informasi dan komunikasi (x_3), jumlah dosen (x_4), serta rata-rata lama sekolah (x_5) cenderung berada di bawah rata-rata keseluruhan. Dengan demikian, karakteristik *Cluster* 2 tidak hanya ditandai oleh rendahnya satu indikator, tetapi oleh

kondisi yang relatif rendah pada beberapa aspek yang secara bersama-sama mendukung transformasi digital pendidikan tinggi.

Berbeda dengan *Cluster* 2, *Cluster* 3 memiliki nilai yang relatif tinggi pada rasio PDRB sektor informasi dan komunikasi (x_3) sebesar 2,45 dan jumlah dosen (x_4) sebesar 2,28. Penggunaan internet (x_1) dan komputer (x_2) juga berada di atas rata-rata, masing-masing sebesar 0,580 dan 0,257. Namun, rata-rata lama sekolah (x_5) sedikit berada di bawah rata-rata dengan nilai $-0{,}164$. Pola ini menunjukkan bahwa *Cluster* 3 lebih menonjol pada dukungan ekonomi sektor informasi dan komunikasi serta kapasitas sumber daya manusia perguruan tinggi, yang disertai dengan tingkat penggunaan teknologi yang relatif lebih tinggi.

Dengan demikian, perbedaan antar-*cluster* terbentuk dari kombinasi beberapa indikator. *Cluster* 1 menunjukkan kondisi yang relatif tinggi secara menyeluruh, *Cluster* 2 menunjukkan kondisi yang relatif rendah pada seluruh indikator, sedangkan *Cluster* 3 memiliki keunggulan terutama pada sektor informasi dan komunikasi serta jumlah dosen. Hasil ini menunjukkan bahwa karakteristik pendukung transformasi digital pendidikan tinggi berbeda antarprovinsi dan tidak hanya ditentukan oleh satu indikator.



Gambar 6. Pemetaan Provinsi Berdasarkan Hasil *K-Means Clustering*

Secara geografis, *Cluster* 1 terdiri atas DI Yogyakarta, *Cluster* 2 terdiri atas 29 provinsi, dan *Cluster* 3 terdiri atas DKI Jakarta, Jawa Barat, Jawa Tengah, dan Jawa Timur. Sebaran pada Gambar 6 menunjukkan bahwa *Cluster* 3 terutama terdapat di Pulau Jawa, sedangkan *Cluster* 2 mencakup sebagian besar provinsi lainnya. Meskipun demikian, pengelompokan tersebut tidak semata-mata menunjukkan pengaruh lokasi geografis, tetapi lebih mencerminkan perbedaan karakteristik indikator. Provinsi dalam *Cluster* 3 memiliki kecenderungan lebih tinggi pada rasio PDRB sektor informasi dan komunikasi, jumlah dosen, serta penggunaan teknologi. Sebaliknya, *Cluster* 2 memiliki nilai yang relatif rendah pada seluruh indikator. DI Yogyakarta yang berada dalam *Cluster* 1 memiliki pola yang berbeda karena seluruh indikator berada di atas rata-rata.

Berdasarkan karakteristik masing-masing *cluster*, strategi pengembangan transformasi digital pendidikan tinggi perlu disesuaikan dengan kondisi kelompok. *Cluster* 2 memerlukan penguatan secara lebih menyeluruh karena seluruh indikator berada di bawah rata-rata. Prioritas dapat diarahkan pada peningkatan penggunaan internet dan komputer, penguatan kapasitas sumber daya manusia perguruan tinggi, serta pengembangan sektor informasi dan komunikasi. Pendekatan tersebut diperlukan karena keterbatasan terjadi pada beberapa aspek secara bersamaan.

Cluster 3 memiliki kapasitas yang relatif kuat pada sektor informasi dan komunikasi, jumlah dosen, serta penggunaan teknologi. Oleh karena itu, strategi dapat diarahkan pada pemanfaatan kapasitas tersebut untuk memperkuat penerapan teknologi digital dalam pembelajaran, meningkatkan inovasi digital, dan mengembangkan kapasitas perguruan tinggi. Sementara itu, rata-rata lama sekolah yang

masih sedikit di bawah rata-rata menunjukkan bahwa aspek pendidikan masyarakat tetap perlu diperhatikan.

Cluster 1 memiliki nilai yang relatif tinggi pada seluruh indikator. Karena kelompok ini hanya terdiri atas satu provinsi, strategi yang lebih tepat adalah mempertahankan kondisi yang telah terbentuk serta mengidentifikasi praktik yang dapat menjadi pembelajaran bagi provinsi lain. Dengan demikian, hasil *K-Means* dapat digunakan untuk membedakan arah strategi, yaitu penguatan menyeluruh untuk *Cluster* 2, optimalisasi kapasitas untuk *Cluster* 3, dan pemeliharaan serta pembelajaran praktik untuk *Cluster* 1

5. Kesimpulan dan Saran

Penelitian ini mengelompokkan 34 provinsi di Indonesia berdasarkan lima faktor pendukung transformasi digital pendidikan tinggi, yaitu penggunaan internet, penggunaan komputer, rasio PDRB sektor informasi dan komunikasi, jumlah dosen, dan rata-rata lama sekolah. Hasil *K-Means Clustering* membentuk tiga kelompok dengan karakteristik yang berbeda. *Cluster* 1 hanya terdiri atas DI Yogyakarta, *Cluster* 2 terdiri atas 29 provinsi, sedangkan *Cluster* 3 terdiri atas DKI Jakarta, Jawa Barat, Jawa Tengah, dan Jawa Timur.

Berdasarkan nilai *centroid*, *Cluster* 1 menunjukkan kondisi yang relatif tinggi pada seluruh indikator, terutama penggunaan komputer dan rata-rata lama sekolah. *Cluster* 2 memiliki nilai *centroid* negatif pada seluruh indikator sehingga menunjukkan kondisi yang relatif lebih rendah secara menyeluruh. Sementara itu, *Cluster* 3 memiliki karakteristik yang lebih menonjol pada rasio PDRB sektor informasi dan komunikasi serta jumlah dosen, disertai penggunaan internet dan komputer yang relatif lebih tinggi, tetapi rata-rata lama sekolah sedikit berada di bawah rata-rata. Hasil tersebut menunjukkan bahwa perbedaan antarprovinsi tidak hanya ditentukan oleh satu indikator, tetapi oleh kombinasi beberapa faktor yang mendukung transformasi digital pendidikan tinggi.

Evaluasi *Silhouette Coefficient* menunjukkan bahwa nilai rata-rata untuk tiga *cluster* sebesar 0,6254. Nilai tersebut menunjukkan bahwa hasil pengelompokan memiliki pemisahan karakteristik yang cukup baik. Hasil *clustering* juga menunjukkan bahwa kebutuhan pengembangan transformasi digital pendidikan tinggi berbeda antarprovinsi. Oleh karena itu, strategi pengembangan perlu disesuaikan dengan karakteristik masing-masing kelompok, terutama dengan memperhatikan aspek yang masih relatif rendah.

Saran untuk penelitian selanjutnya dapat menggunakan periode data yang lebih mutakhir dan menambahkan indikator lain yang relevan untuk menggambarkan kesiapan transformasi digital pendidikan tinggi, seperti infrastruktur digital, kualitas sumber daya manusia, atau akses terhadap layanan pendidikan tinggi. Selain itu, metode klasterisasi lain dapat dibandingkan dengan *K-Means* untuk mengetahui kestabilan dan kualitas hasil pengelompokan. Evaluasi dapat dilakukan menggunakan beberapa ukuran, seperti *Silhouette Coefficient*, sehingga hasil antar metode dapat dibandingkan secara lebih komprehensif.

6. Referensi

[1] S. Harahap and Z. Napitupulu, “Pengaruh Teknologi Terhadap Pendidikan Di Indonesia: Systematic Literatur Riview,” *J. Pendidik. dan Kependidikan*, vol. 8, no. 2, pp. 9–17, 2023.

[2] S. N. Andani, E. U. Hanik, R. Kusumastuti, and S. L. Ifana, “Pemanfaatan Digital Learning Dalam Pembelajaran,” *PENDASI J. Pendidik. Dasar Indones.*, vol. 9, no. 1, pp. 67–77, 2025.

[3] D. Sugari and H. Hilalludin, “Kesetaraan Akses Pendidikan Teknologi Tantangan dan Peluang di Indonesia dan Dunia,” *J. <ultidisciplinary Res.*, vol. 1, no. 1, pp. 44–56, 2025.

[4] L. M. C. Benavides, J. A. T. Arias, M. D. A. Serna, J. W. B. Bedoya, and D. Burgos, “Digital Transformation in Higher Education Institutions : A Systematic Literature Review,” *sensors*, vol. 20, no. 3291, pp. 1–22, 2020.

[5] R. Sira and T. Sihombing, “Pemerataan Pendidikan: Studi Kasus 34 Provinsi di Indonesia,” *Parahyangan Econ. Dev. Rev.*, vol. 1, no. 2, pp. 143–151, 2022.

[6] Badan Pusat Statistik (BPS), “Masyarakat Informasi.” [Online]. Available: https://www.bps.go.id/id/statistics-table?subject=565&sortBy=date%2Ctitle&sortOrder=desc%2Casc

[7] T. Gkrimpizi, V. Peristeras, and I. Magnisalis, “Defining the Meaning and Scope of Digital

Transformation in Higher Education Institutions," *Adm. Sci.*, vol. 14, no. 48, pp. 1–18, 2024.
[8] M. A. Selan and K. T. Wahyuni, "Analisis Pembangunan Teknologi Informasi dan Komunikasi di Indonesia Tahun 2015-2020," *Semin. Nas. Off. Stat. 2022 Pemanfaat.*, pp. 197–206, 2022.
[9] S. Bumbungan, "Peran dan Perkembangan Jaringan Internet dalam Mendukung Transformasi Digital Global," *Bull. Netw. Eng. Informatics*, vol. 3, no. 1, pp. 11–15, 2025, doi: 10.59688/bufnets.v3i1.66.
[10] M. A. I. Fikri and D. Handayani, "Pengaruh Teknologi Informasi dan Komunikasi Terhadap Produk Domestik Regional Bruto Indonesia," *Syntax Lit. J. Ilm. Indones.*, vol. 8, no. 11, pp. 6552–6564, 2023.
[11] A. T. Wulandari and J. Sumarah, "Kluster Rata-Rata Lama Sekolah (RLS) Menurut Jenis Kelamin di Provinsi Jawa Tengah dengan K-Means," *J. Media Inform. Budidarma*, vol. 5, no. 4, pp. 1548–1558, 2021, doi: 10.30865/mib.v5i4.3279.
[12] A. P. Kusuma and A. D. Oktavianto, "Analisis Metode Euclidean Distance dalam Menentukan Koordinat Peta pada Alamat Rumah," *J. Teknol. dan Manaj. Inform.*, vol. 8, no. 2, pp. 108–115, 2022.
[13] A. A. Arbeit, Ferdiansyah, M. B. Adna, M. Ridwan, and R. Vindua, "Pemanfaatan Klasterisasi K-Means untuk Pengelompokan Berdasarkan Indikator Ekonomi, Digitalisasi, dan Produksi Komoditas," *J. Artif. Intell. Digit. Bus.*, vol. 4, no. 2, pp. 5119–5127, 2025.
[14] C. Lumbantoruan and D. A. Megawaty, "Penerapan Metode K-Means Clustering dan Principal Component Analysis (PCA) untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan," *J. Artif. Intell. Technol. Inf.*, vol. 4, no. 2, pp. 120–131, 2026.
[15] Nurfajriyani, D. D. Amaliana, and S. P. Wulandari, "Pengelompokan Provinsi berdasarkan Aspek Pembangunan Pendidikan di Indonesia Tahun 2023 menggunakan Analisis Cluster," *Pentagon J. Mat. dan Ilmu Pengetah. Alam*, vol. 3062, no. 8652, pp. 134–153, 2024.
[16] C. E. Widiantoro, S. P. Putri, Silvi, and T. Purwandari, "Pengelompokkan Provinsi di Indonesia Berdasarkan Indikator Pendidikan Jenjang SMA/Sederajat dengan Analisis Klaster Non Hierarki," *BIAStatistics J. Stat. Theory Apl.*, vol. 1907, no. 6274, pp. 1–9, 2024.
[17] Patimah, Zulpan, D. U. Alfansuri, E. Munawaroh, and M. Ilyas, "Memahami Dan Menerapkan Uji Korelasi Dalam Analisis Data Penelitian Pendidikan," *J. Educ. Innov.*, vol. 3, no. 4, pp. 740–752, 2025.
[18] H. Ihsan, M. S. Wahyuni, and A. I. Kurnadipare, "Penerapan Analisis Faktor Eksplanatori pada Pengambilan Keputusan Mahasiswa Membeli Produk Online di Kota Makassar," *J. Math. Comput. Stat.*, vol. 6, no. 2, pp. 169–176, 2023.
[19] R. G. Whendasmoro and Joseph, "Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score Pada," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, pp. 872–876, 2022, doi: 10.30865/jurikom.v9i4.4526.
[20] A. Septianingsih, "Analisis K-Means Clustering Pada Pemetaan Provinsi Indonesia Berdasarkan Indikator Rumah Layak Huni," *Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 3, no. 1, pp. 224–241, 2022.