

Clustering Provinsi di Indonesia Berdasarkan Indeks Pembangunan Teknologi Informasi dan Komunikasi Menggunakan *K-Medoids*

Mardhani Dwi Novianto^{1*}, Galang Sumantri¹, Pusparani Puan Prihastuti¹

¹Universitas Negeri Yogyakarta, Jl. Colombo No.1 Karangmalang Yogyakarta 55281, Indonesia

*Corresponding author: mardhani9497fmipa.2019@student.uny.ac.id



P-ISSN: 2986-4178
E-ISSN: 2988-4004

Riwayat Artikel

Dikirim: 30 Maret 2023
Direvisi: 06 September 2023
Diterima: 7 Desember 2023

ABSTRAK

Teknologi, Informasi, dan Komunikasi (TIK), saat ini berkembang pesat, namun dapat menyebabkan kesenjangan perkembangan teknologi antarwilayah. Penelitian ini bertujuan sebagai langkah awal untuk mengatasi permasalahan tersebut dengan mengelompokkan provinsi berdasarkan Indeks Pembangunan TIK menggunakan metode *K-Medoids*. Data yang digunakan adalah data Indeks Pembangunan TIK tahun 2021 dengan 3 variabel subindeks Pembangunan TIK. *Clustering* dilakukan pada rentang 2 s.d. 5 *cluster* dengan validasi menggunakan *Davies Bouldin Index*. Berdasarkan hasil penelitian, diperoleh 5 *cluster* optimal yaitu, *cluster* pertama (1 provinsi), *cluster* kedua (6 provinsi), *cluster* ketiga (13 provinsi), *cluster* keempat (8 provinsi), dan *cluster* kelima (6 provinsi). Beberapa provinsi di Indonesia tergolong sebagai *cluster* dengan Indeks Pembangunan TIK rendah dan sangat rendah. Oleh karena itu, pemerintah dapat memaksimalkan pembangunan TIK di seluruh provinsi untuk mengurangi kesenjangan perkembangan teknologi.

Kata Kunci: Indeks Pembangunan TIK, *Clustering*, *K-Medoids*, *Davies Bouldin Index*

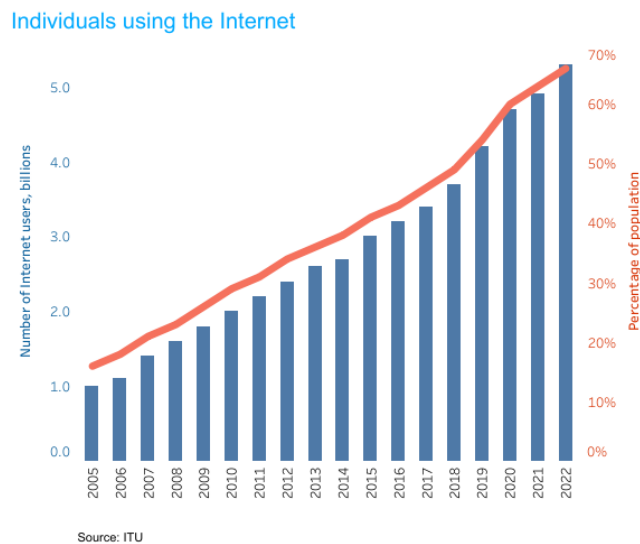
ABSTRACT

Technology, Information, and Communication (ICT) is currently growing rapidly, but it can cause a gap in technological development between regions. This research to be a first step to overcome this problem by clustering provinces based on the ICT Development Index using the K-Medoids method. The data used is the 2021 ICT Development Index data with 3 ICT Development sub-index variables. Clustering is carried out in the range of 2 to 5 clusters, with validation using the Davies Bouldin Index. Based on the research results, 5 optimal clusters were obtained, namely, the first cluster (1 province), the second cluster (6 provinces), the third cluster (13 provinces), the fourth cluster (8 provinces), and the fifth cluster (6 provinces). Some provinces in Indonesia are classified as clusters with a low or very low ICT Development Index. Therefore, the government can maximize ICT development in all provinces to reduce the gap in technological development.

Keywords: *ICT Development Index, Clustering, K-Medoids, Davies Bouldin Index.*

1. Pendahuluan

Teknologi, Informasi, dan Komunikasi atau yang disebut dengan TIK, saat ini semakin berkembang pesat di seluruh dunia dan semakin banyak orang yang memiliki akses untuk berbagi informasi [1]. Menurut Parsaorantua *et al.* (2017), sekarang ini, perkembangan teknologi informasi dan komunikasi dalam kehidupan sehari-hari sangat pesat. Kehadiran internet dalam teknologi informasi dan komunikasi membantu aktivitas dalam memperoleh dan menyampaikan informasi secara cepat [2].



Gambar 1. Perkembangan jumlah pengguna internet

Berdasarkan Gambar 1, terlihat bahwa jumlah pengguna internet dari tahun ke tahun semakin meningkat. *International Telecommunication Union* (ITU) memperkirakan bahwa sekitar 5,3 miliar orang atau 66 persen dari populasi dunia menggunakan internet pada tahun 2022. Sedangkan di Indonesia, menurut BPS (2022) dalam publikasinya yang berjudul Indeks Pembangunan Teknologi Informasi dan Komunikasi 2021, pada tahun 2021 penetrasi internet berkembang dengan pesat yaitu dari 53,73 persen pada tahun 2020 menjadi 62,10 persen pada tahun 2021 [3]. Internet telah menjadi kebutuhan yang penting dalam kehidupan sehari-hari. Internet dapat digunakan untuk bekerja, belajar, mengakses berbagai platform digital, dan bersosialisasi satu sama lain terutama karena adanya pandemi Covid-19. Hal ini menunjukkan bahwa teknologi informasi dan komunikasi sangat penting.

Menurut Jayanthi & Dinaseviani (2022), pandemi Covid-19 memberikan dampak buruk, tetapi di sisi lain, terdapat sisi positif yang muncul yaitu percepatan perkembangan teknologi digital [4]. Dengan berbagai kemudahan yang ditawarkan dengan berkegiatan secara daring, muncul masalah yang baru yaitu kesenjangan digital, baik antarwilayah maupun antarstrata-sosial. Jika hal ini terus terjadi, dapat menyebabkan kesenjangan perkembangan kemajuan teknologi.

Tingkat kemajuan teknologi informasi dan komunikasi suatu daerah dapat diukur melalui Indeks Pembangunan Teknologi, Informasi, dan Komunikasi (IP-TIK). Menurut BPS (2022), Indeks Pembangunan TIK sangat penting sebagai ukuran standar tingkat pembangunan TIK pada suatu wilayah [3]. Selain itu, indeks ini juga mampu mengukur pertumbuhan TIK, mengukur gap digital atau kesenjangan digital antarwilayah, dan mengukur potensi pembangunan TIK. Semakin tinggi nilai indeks menunjukkan bahwa pembangunan TIK di suatu wilayah semakin pesat, sebaliknya semakin rendah nilai indeks menunjukkan pembangunan TIK di suatu wilayah relatif masih lambat.

Langkah awal yang dapat dilakukan untuk mengatasi kesenjangan perkembangan kemajuan teknologi antarwilayah adalah dengan mengelompokkan provinsi di Indonesia ke dalam beberapa *cluster*. Pembagian provinsi ke dalam beberapa *cluster* akan memudahkan pemerintah dalam membuat kebijakan. *Cluster* adalah kumpulan objek yang mirip satu sama lain dan berbeda dengan objek di *cluster* lain [5]. Untuk memperoleh *cluster*, diperlukan suatu proses *clustering*. *Clustering* adalah proses pengelompokan sekumpulan objek menjadi beberapa kelompok atau *cluster* sehingga objek dalam suatu *cluster* memiliki kesamaan yang tinggi, tetapi sangat berbeda dengan objek dalam *cluster* lain [6].

Penelitian terkait IP-TIK menggunakan teknik *clustering* sebelumnya telah dilakukan oleh [1] dengan metode *K-Means* yang diuji validitas hasil *cluster*-nya menggunakan *Davies Bouldin Index* (DBI). Data yang digunakan adalah data IP-TIK tahun 2019 yang terdiri dari 34 provinsi dengan 3 variabel subindeks pembangunan TIK yaitu subindeks akses dan infrastruktur TIK, subindeks penggunaan TIK, dan subindeks keahlian TIK. Penelitian tersebut menghasilkan 3 *cluster* yaitu *cluster* dengan tingkat IP-TIK rendah terdiri dari 4 provinsi, *cluster* dengan tingkat IP-TIK sedang terdiri dari 22 provinsi, dan *cluster* dengan tingkat IP-TIK tinggi terdiri dari 8 provinsi.

Penelitian ini akan menggunakan metode *K-Medoids* untuk mengelompokkan *cluster*. *K-Medoids* adalah suatu algoritma untuk menemukan *medoids* di dalam sebuah kelompok (*cluster*) [7]. Adapun *medoids* sendiri dapat diartikan sebagai suatu objek pusat *cluster* dan objek yang mewakili *cluster*. Menurut Septiani *et al.* (2022), algoritma *K-Medoids* ini memiliki kinerja yang lebih baik dibandingkan dengan algoritma *K-Means* baik jika diterapkan pada dataset berukuran kecil maupun besar [8]. Hal ini dikarenakan dalam algoritma *K-Medoids*, *k* didefinisikan sebagai objek yang representatif untuk meminimalkan jumlah ketidaksamaan objek data, sedangkan pada *K-Means* yang menggunakan jumlah jarak *euclidean distances* untuk objek data [7]. Selain itu, Gunawan *et al.* (2020) juga menambahkan bahwa algoritma *K-Medoids* dapat digunakan untuk mengatasi kelemahan algoritma *K-Means* yang sangat sensitif terhadap *outlier* karena objek-objek ini sangat jauh karakteristiknya dari mayoritas data lain, sehingga jika dimasukkan ke dalam suatu *cluster* akan mendistorsi nilai rata-rata (*mean*) dari *cluster* tersebut [9].

Berdasarkan uraian permasalahan di atas, maka diperlukan suatu pengelompokan provinsi di Indonesia untuk meninjau lebih lanjut terkait Indeks Pembangunan TIK pada tiap provinsi di Indonesia. Data yang digunakan adalah data Indeks Pembangunan Teknologi, Informasi, dan Komunikasi tahun 2021. Sedangkan metode yang digunakan adalah *K-Medoids* karena dapat mengatasi kelemahan metode *K-Means*. Hasil *cluster K-Medoids* ini diuji validitasnya dengan *Davies Bouldin Index*. Dengan hasil *clustering* provinsi ini, diharapkan dapat ditemukan *cluster* atau kelompok provinsi berdasarkan kemiripan karakteristik terkait IP-TIK sehingga pemerintah dapat mengambil tindakan yang tepat dalam pembangunan TIK sehingga tercapai kesetaraan dan mengurangi adanya kesenjangan perkembangan kemajuan teknologi antarwilayah.

2. Metodologi Penelitian

Penelitian ini menggunakan data sekunder yang diperoleh dari [3] dalam publikasinya yang berjudul Indeks Pembangunan Teknologi Informasi dan Komunikasi 2021. Data yang digunakan merupakan data Indeks Pembangunan Teknologi, Informasi, dan Komunikasi yang terdiri dari 34 provinsi dan 3 variabel subindeks pembangunan TIK dengan periode data yang digunakan adalah tahun 2021. Ketiga variabel tersebut merupakan ringkasan dari 11 indikator penyusun IP-TIK.

2.1. Variabel Penelitian

2.1.1. SubIndeks Akses dan Infrastruktur TIK (X1)

Subindeks ini menggambarkan kesiapan TIK yang diukur dari sisi akses dan infrastruktur. Subindeks ini terdiri atas lima indikator penyusun subindeks yaitu pelanggan telepon tetap per 100 penduduk, pelanggan telepon seluler per 100 penduduk, *bandwidth* internet internasional (bit/s) per pengguna, persentase rumah tangga dengan komputer, dan persentase rumah tangga dengan akses internet.

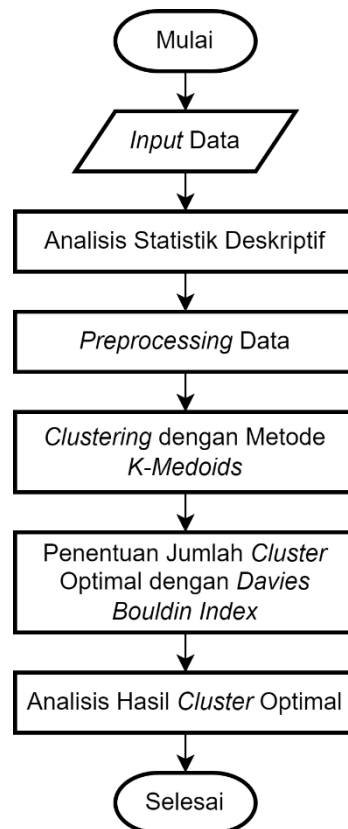
2.1.2. Subindeks Penggunaan TIK (X2)

Subindeks ini menggambarkan intensitas TIK yang diukur dari penggunaan TIK tingkat penggunaan TIK oleh masyarakat. Subindeks ini disusun oleh tiga indikator yaitu persentase individu yang menggunakan internet, pelanggan *fixed broadband* internet per 100 penduduk, dan pelanggan *mobile broadband* internet aktif per 100 penduduk.

2.1.3. Subindeks Keahlian TIK (X3)

Subindeks ini menggambarkan kemampuan atau keahlian yang diperlukan dalam TIK. Subindeks ini menggunakan tiga indikator *proxy* sebagai pendekatan keahlian TIK, yaitu rata-rata lama sekolah, Angka Partisipasi Kasar (APK) sekunder, dan APK tersier.

2.2. Tahapan Penelitian



Gambar 2. Flowchart Penelitian

Gambar 2 menunjukkan mengenai tahapan penelitian yang dilakukan. Secara garis besar, tahapan penelitian dalam penelitian ini meliputi *input data*, analisis statistik deskriptif, *preprocessing data*, *clustering* dengan metode *K-Medoids*, penentuan jumlah *cluster* optimal dengan *Davies Bouldin Index*, dan analisis hasil *cluster* optimal.

2.2.1. Analisis Statistik Deskriptif

Statistik deskriptif adalah statistik yang tingkat pekerjaannya mencakup cara-cara menghimpun, menyusun, atau mengatur, mengolah, menyajikan, dan menganalisis data angka, agar dapat memberikan gambaran yang teratur, ringkas, dan jelas mengenai suatu gejala, peristiwa, atau keadaan [11]. Statistik deskriptif yang digunakan pada penelitian ini berupa nilai minimal, rata-rata, dan maksimal. Analisis ini bertujuan untuk menggambarkan data agar karakteristiknya lebih mudah diamati.

2.2.2. Preprocessing Data

Preprocessing data merupakan salah satu langkah data *mining* yang berhubungan dengan persiapan dan transformasi data. *Preprocessing* data bertujuan untuk mengurangi ukuran data, menemukan relasi antar data, normalisasi data, menghapus pencilan, dan mengekstrak fitur dari data. Langkah ini meliputi beberapa teknik seperti data *cleaning*, *integration*, *transformation*, dan *reduction* [12]. *Preprocessing* data yang akan digunakan dalam penelitian ini adalah *data cleaning* dan *transformation*. Data *cleaning* meliputi pengecekan nilai *null* dan *outlier*/pencilan, sedangkan *transformation* dilakukan dengan menormalisasi data.

2.2.3. Clustering dengan Metode K-Medoids

Clustering merupakan pengelompokan suatu kelompok data ke dalam beberapa kelompok yang lebih kecil yang biasa disebut dengan *cluster*. Data-data dalam *cluster* yang sama dianggap memiliki karakteristik yang sama. Salah satu metode yang dapat digunakan untuk melakukan *clustering* adalah metode *K-Medoids*. *K-Medoids* adalah suatu algoritma untuk menemukan *medoids* di dalam sebuah kelompok (*cluster*) [7]. *Medoids* sendiri dapat diartikan sebagai suatu objek pusat *cluster* dan objek yang mewakili *cluster*. Adapun langkah-langkah dari algoritma *K-Medoids* adalah sebagai berikut [13] [14].

- Inisialisasi pusat *cluster* sebanyak k (jumlah *cluster*).
- Alokasikan *setiap* data (objek) ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidean Distance* dengan persamaan:

$$d_{euc}(x_{ij}, c_{kj}) = \sqrt{\sum_{j=1}^p \sum_{i=1}^n (x_{ij} - c_{kj})^2} \quad (1)$$

Jarak *Euclidean Distance* (d_{euc}) mengukur seberapa jauh suatu pengamatan ke- i pada variabel ke- j (x_{ij}) dari pusat *cluster* ke- k pada variabel ke- j (c_{kj}). Ini merupakan metrik yang digunakan untuk menilai seberapa dekat atau jauh suatu objek dari pusat *cluster* pada setiap variabel yang diamati, dengan p variabel dan n pengamatan.

- Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat *medoids* baru.
- Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoids* baru.
- Hitung total simpangan (S) dengan menghitung nilai total *distance* baru – total *distance* lama. Jika $S < 0$, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai *medoid*.
- Ulangi langkah c sampai e hingga tidak terjadi perubahan *medoids*, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

2.2.4. Penentuan Jumlah Cluster Optimal

Cluster yang dihasilkan dari proses *clustering* ini perlu divalidasi untuk memperoleh hasil *cluster* yang optimal. Salah satu indeks validitas yang dapat digunakan dalam algoritma *K-Medoids* ini adalah Indeks *Davies-Bouldin*. Indeks validitas *Davies-Bouldin* (DB) diperoleh dengan menghitung rata-rata nilai setiap titik pada himpunan data,

dengan perhitungan nilai setiap titik dilakukan dengan menjumlahkan nilai *compactness* yang dibagi dengan jarak antara kedua titik pusat *cluster* [15]. Adapun nilai dari indeks DB ini berkisar antara 0 hingga 1. Semakin kecil nilai yang didapat, maka kemiripan data dalam satu kelompok akan semakin tinggi atau dengan kata lain, jumlah *cluster* terbaik ditunjukkan oleh nilai DB yang terkecil [16].

Ada empat tahapan dalam menghitung *Davies Bouldin Index*. Adapun tahapannya adalah sebagai berikut [8].

- a. Menghitung *Sum of square within cluster* (SSW), untuk mengetahui matrik/kohesi/homogenitas. Kohesi merupakan keterikatan anggota *cluster* dalam satu *cluster*. Adapun persamaannya sebagai berikut.

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \quad (2)$$

Jumlah data dalam *cluster* ke- i (m_i) dihitung dengan menjumlahkan total data yang termasuk dalam *cluster* tersebut. Dalam konteks ini, setiap data dalam *cluster* (x) memiliki jarak ($d(x, c)$) dari *centroid cluster* ke- i (c_i), diukur sebagai jarak antara data pada *cluster* tersebut (x_j) dan pusat *cluster* (c_i).

- b. Menghitung *Sum of square between cluster* (SSB), untuk mengetahui separasi/heterogenitas. Separasi merupakan perbedaan antara satu *cluster* dengan *cluster* lainnya. Adapun persamaannya sebagai berikut.

$$SSB_{i,j} = d(c_i, c_j) \quad (3)$$

Jarak antara *centroid cluster* ke- i (c_i) dan *centroid cluster* ke- j (c_j), disimbolkan sebagai $d(c_i, c_j)$, menggambarkan seberapa jauh pusat dua *cluster* berbeda satu sama lain. Ini memberikan informasi tentang tingkat kesamaan atau perbedaan antara karakteristik rata-rata dari dua *cluster* yang berbeda dalam analisis *cluster*.

- c. Menghitung Rasio (R_{ij}), untuk mengetahui seberapa bagus nilai perbandingan *cluster* satu dengan *cluster* yang lainnya. Adapun persamaannya sebagai berikut.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (4)$$

Rasio antar *cluster* (R_{ij}), dalam konteks ini, diukur dengan membandingkan kohesi dalam suatu *cluster* dengan kohesi antara dua *cluster*. Kohesi dari *cluster* ke- i (SSW_i) dan *cluster* ke- j (SSW_j) menggambarkan seberapa rapat data dalam masing-masing *cluster*. Sementara itu, separasi antara *cluster* ke- i dan ke- j ($SSB_{i,j}$) mengukur seberapa jauh atau terpisah kedua *cluster* tersebut.

- d. Menghitung DBI (*Davies Bouldin Index*). Nilai DBI dapat menunjukkan kualitas hasil *clustering*, hasil *clustering* akan semakin baik jika hasil dari perhitungan DBI yang diperoleh semakin kecil mendekati nol, tetapi tidak negatif (*non-negatif* ≥ 0). Adapun persamaannya sebagai berikut.

$$DBI = \frac{1}{K} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (5)$$

Jumlah *cluster* yang digunakan, disimbolkan sebagai k , mencerminkan banyaknya kelompok yang dihasilkan dalam suatu analisis *cluster*. Dalam konteks ini, k merujuk pada jumlah *cluster* yang telah ditentukan sebelumnya atau yang dihasilkan selama proses *clustering*.

2.2.5. Analisis Hasil Cluster Optimal

Analisis data hasil *cluster* optimal yang digunakan dalam penelitian ini juga menggunakan teknik statistik deskriptif. Statistik deskriptif yang digunakan adalah dengan membentuk suatu profilisasi *cluster*. Profilisasi *cluster* dibentuk dengan melihat rata-rata atau *mean* pada setiap variabelnya. Langkah ini dilakukan agar mempermudah dalam melihat karakteristik setiap *clusternya*.

3. Hasil dan Pembahasan

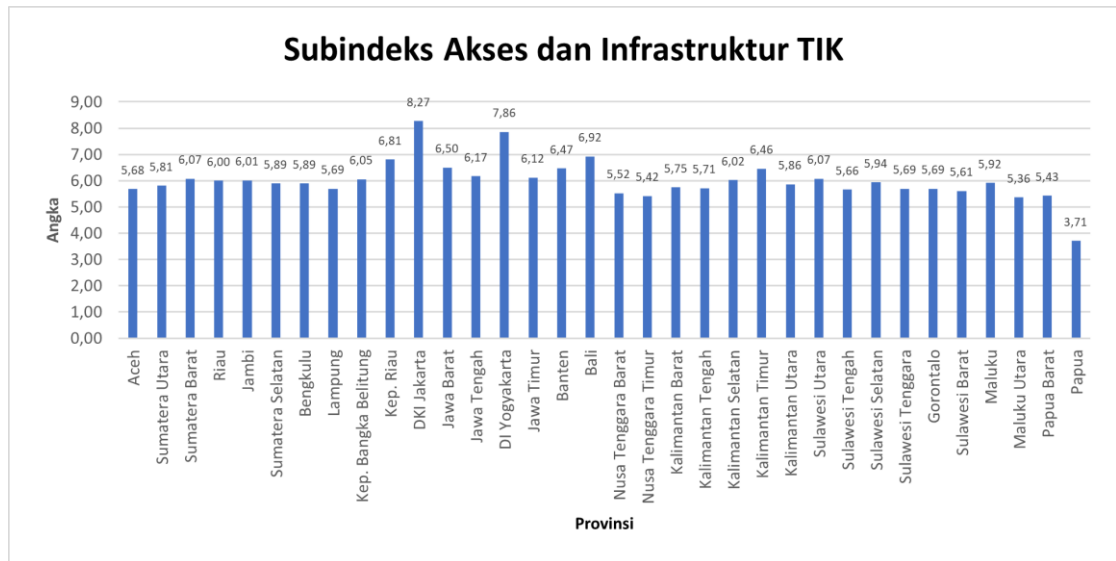
3.1. Analisis Statistik Deskriptif

Analisis statistik deskriptif dilakukan untuk mendapatkan gambaran distribusi data. Pada hasil analisis ini, ditampilkan nilai minimal, rata-rata, dan maksimal dari variabel subindeks akses dan infrastruktur TIK (X1), variabel subindeks penggunaan TIK (X2), dan variabel subindeks keahlian TIK (X3). Hasil statistika deskriptif yang menggambarkan Indeks Pembangunan Teknologi, Informasi, dan Komunikasi di Indonesia pada tahun 2021 ditampilkan dalam Tabel 1 berikut.

Tabel 1 Statistik Deskriptif

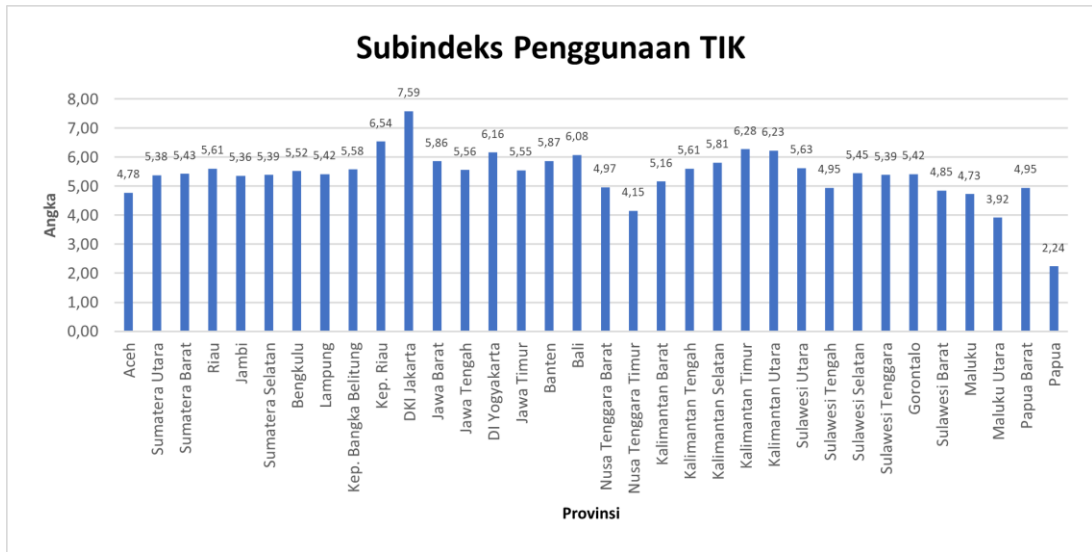
Variabel	Min	Mean	Max
X1	3,71	6	8,27
X2	2,24	5,39	7,59
X3	4,84	6,13	7,68

Statistik deskriptif juga dilakukan dengan menampilkan diagram atau grafik sehingga mempermudah dalam analisis.



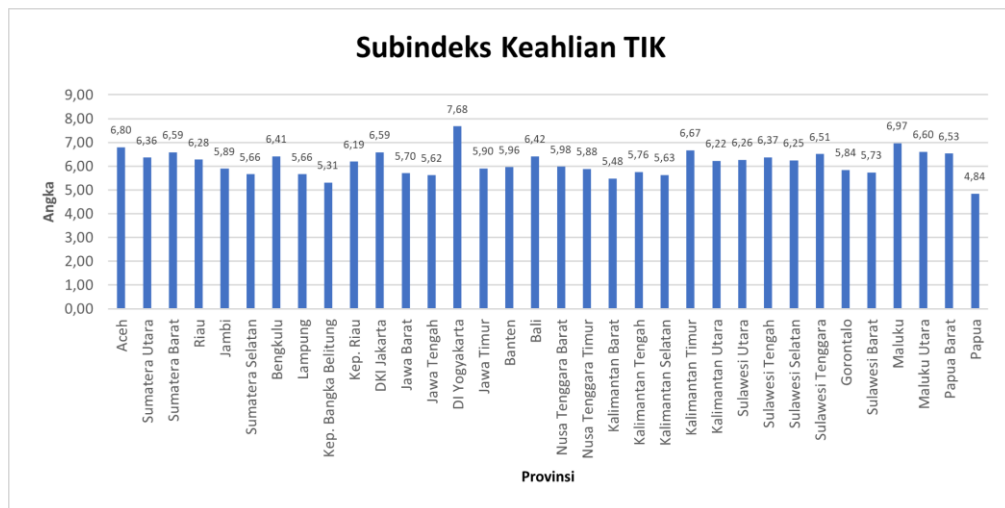
Gambar 3. Subindeks Akses dan Infrastruktur TIK berdasarkan Provinsi

Berdasarkan pada Tabel 1 dan Gambar 3, diketahui hasil analisis statistik deskriptif mengenai subindeks akses dan infrastruktur TIK dengan rata-rata 6 sedangkan subindeks tertinggi terdapat di DKI Jakarta dengan angka 8,27 dan subindeks terendah terdapat di Papua dengan angka 3,71.



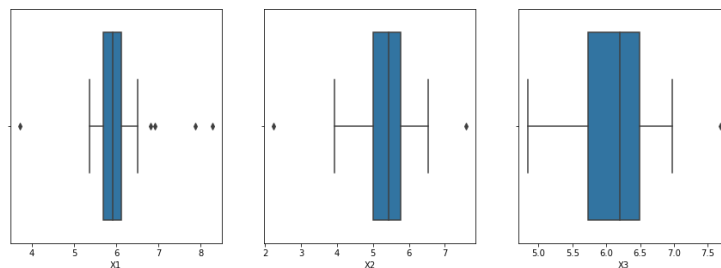
Gambar 4. Subindeks Penggunaan TIK berdasarkan Provinsi

Berdasarkan pada Tabel 1 dan Gambar 4, diketahui hasil analisis statistik deskriptif mengenai subindeks penggunaan TIK dengan rata-rata 5,39 sedangkan subindeks tertinggi terdapat di DKI Jakarta dengan angka 7,59 dan subindeks terendah terdapat di Papua dengan angka 2,24.



Gambar 5. Subindeks Keahlian TIK

Berdasarkan pada Tabel 1 dan Gambar 5, diketahui hasil analisis statistik deskriptif mengenai subindeks keahlian TIK dengan rata-rata 6,13 sedangkan subindeks tertinggi terdapat di DI Yogyakarta dengan angka 7,68 dan subindeks terendah terdapat di Papua dengan angka 4,84.



Gambar 6. Boxplot Subindeks Akses dan Infrastruktur, Penggunaan, serta Keahlian TIK

Berdasarkan pada Gambar 6, diketahui bahwa pada variabel subindeks akses dan infrastruktur TIK (X1) terdapat 5 pencilan, variabel subindeks penggunaan TIK (X2) terdapat 2 pencilan, dan variabel subindeks keahlian TIK (X3) terdapat 1 pencilan.

3.2. Preprocessing Data

Pada tahap ini, dilakukan *preprocessing* data. *Preprocessing* data yang dilakukan adalah *data cleaning* dan *transformation*. *Data cleaning* yang dilakukan pada penelitian ini adalah pengecekan nilai *null*, untuk data pencilan tidak dihilangkan (*drop*) karena metode *K-Medoids* cukup kuat atau tidak terlalu sensitif terhadap pencilan [9]. Kemudian setelah dilakukan pengecekan, dalam data ini juga tidak terdapat nilai *null*. Selanjutnya untuk transformasi pada *preprocessing* data ini dilakukan dengan menormalisasi data. Pada data ini sudah dilakukan normalisasi dengan mengubah ulang skala (*rescaled*) untuk rentang identik (rentang yang sama) dari 1 sampai 10 untuk setiap variabel. Nilai ini disesuaikan dengan kondisi negara masing-masing.

3.3. Clustering dengan Metode K-Medoids

Pada bagian ini, dilakukan *clustering* dengan metode *K-Medoids*. *Clustering* dilakukan dengan skenario jumlah *cluster* sebanyak 2 s.d. 5 *cluster*. Dalam hal ini, ditunjukkan *clustering* untuk jumlah *cluster* sebanyak 2 *cluster*. Adapun langkah-langkah yang dilakukan adalah sebagai berikut.

- a. Inisialisasi pusat *cluster* sebanyak 2

Tabel 2 Inisialisasi pusat *cluster*

Cluster	X1	X2	X3
1	5,94	5,45	6,25
2	6,01	5,36	5,89

- b. Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidian Distance* dengan persamaan (1).

$$d_1(x_{ij}, c_1) = \sqrt{(5,68 - 5,94)^2 + (4,78 - 5,45)^2 + (6,80 - 6,25)^2}$$

$$d_2(x_{ij}, c_2) = \sqrt{(5,68 - 6,01)^2 + (4,78 - 5,36)^2 + (6,80 - 5,89)^2}$$

Perhitungan di atas untuk sampel ke-1 dalam data atau data pertama. Perhitungan data yang lainnya dengan cara yang sama. Setelah diperoleh jarak *Euclidean Distance* dengan persamaan (1) untuk masing-masing data, langkah berikutnya adalah sebagai berikut.

- c. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat *medoids* baru.
- d. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoids* baru.
- e. Hitung total simpangan (*S*) dengan menghitung nilai total *distance* baru – total *distance* lama. Jika $S < 0$, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan *k* objek baru sebagai *medoid*.
- f. Ulangi langkah c sampai e hingga tidak terjadi perubahan *medoids*, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

Dengan melakukan perulangan langkah c sampai e, diperoleh total simpangan yaitu 26,88 dan *medoids* akhir untuk *cluster* pertama adalah sampel ke-27 dalam data yaitu data Provinsi Sulawesi Selatan dan untuk *cluster* kedua adalah sampel ke-17 dalam data yaitu data Provinsi Bali.

Tabel 3 *Medoid* akhir

Cluster	X1	X2	X3
1	5,94	5,45	6,25
2	6,92	6,08	6,42

Medoids ini digunakan untuk menghitung jarak *antara* anggota *cluster* dan medoidnya yang menjadi dasar untuk melakukan pengelompokan atau *clustering* sehingga dapat diperoleh *cluster* beserta anggota *cluster* masing-masing. Untuk *clustering* dengan jumlah *cluster* sebanyak 3, 4, dan 5 menggunakan cara yang sama.

3.4. Penentuan Jumlah Cluster Optimal

Pada bagian ini, dilakukan penentuan jumlah *cluster* yang optimal dengan melihat nilai *Davies-Bouldin Index*. Adapun hasil validasi menggunakan nilai *Davies-Bouldin Index* untuk data penelitian dapat dilihat pada Tabel 4.

Tabel 4 Hasil Validasi

Jumlah cluster	Davies-Bouldin Index
2	0,9295045874
3	1,026006037
4	0,7566314991
5	0,7271384511

Berdasarkan Tabel 4, dapat diketahui bahwa nilai *Davies-Bouldin Index* terkecil yaitu pada jumlah *cluster* 5. Oleh karena itu, dapat disimpulkan bahwa jumlah *cluster* 5 terpilih menjadi jumlah *cluster* yang optimal.

3.5. Analisis Hasil Cluster Optimal

Setelah mendapatkan jumlah *cluster* optimal untuk pengelompokan provinsi di Indonesia berdasarkan Indeks Pembangunan TIK, selanjutnya hasil dari *clustering* tersebut divisualisasikan pada Gambar 7 dan disajikan pada Tabel 5.



Gambar 7. Visualisasi *clustering* provinsi di Indonesia berdasarkan IP-TIK

Tabel 5 Hasil *clustering* provinsi di Indonesia berdasarkan IP-TIK

Cluster	Jumlah Anggota	Anggota Cluster
1	1	Papua
2	6	Aceh, Maluku, Maluku Utara, Nusa Tenggara Timur, Papua Barat, dan Sulawesi Tengah
3	13	Gorontalo, Jambi, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kepulauan Bangka Belitung, Lampung, Nusa Tenggara Barat, Sulawesi Barat, dan Sumatera Selatan
4	8	Bengkulu, Kalimantan Utara, Riau, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Utara, Sumatera Barat, dan Sumatera Utara
5	6	Bali, Banten, DI Yogyakarta, DKI Jakarta, Kalimantan Timur, dan Kepulauan Riau

Setelah divisualisasikan dan disajikan hasil *cluster*-nya, selanjutnya akan dihitung nilai rata-rata dari setiap *cluster* untuk mengetahui karakteristik di setiap *cluster*. Adapun hasil perhitungan tersebut disajikan pada Tabel 6.

Tabel 6 Profilisasi setiap *cluster*

<i>Cluster</i>	X1	X2	X3
1	3,710	2,240	4,840
2	5,578	4,580	6,525
3	5,902	5,426	5,705
4	5,916	5,580	6,360
5	7,132	6,420	6,585

Berdasarkan Tabel 6 dapat diketahui ciri-ciri dari setiap *cluster* sehingga dapat diinterpretasikan sebagai berikut:

1. *Cluster* 1 terdiri dari 1 provinsi yang merupakan *cluster* dengan Indeks Pembangunan TIK yang sangat rendah. Hal itu terjadi karena nilai rata-rata dari masing-masing variabel menjadi paling rendah daripada *cluster* yang lain.
2. *Cluster* 2 terdiri dari 6 provinsi yang merupakan *cluster* dengan Indeks Pembangunan TIK yang rendah. Hal itu terjadi karena nilai rata-rata dari variabel X1 dan X2 menjadi paling rendah kedua daripada *cluster* yang lain, meskipun variabel X3 cukup tinggi.
3. *Cluster* 3 terdiri dari 13 provinsi yang merupakan *cluster* dengan Indeks Pembangunan TIK yang sedang. Hal itu terjadi karena nilai rata-rata dari variabel X1 dan X2 menjadi paling rendah ketiga daripada *cluster* yang lain, meskipun variabel X3 lebih rendah daripada *cluster* 2.
4. *Cluster* 4 terdiri dari 8 provinsi yang merupakan *cluster* dengan Indeks Pembangunan TIK yang tinggi. Hal itu terjadi karena nilai rata-rata dari variabel X1 dan X2 menjadi paling tinggi kedua daripada *cluster* lain, meskipun variabel X3 lebih rendah daripada *cluster* 2.
5. *Cluster* 5 terdiri dari 6 provinsi yang merupakan *cluster* dengan Indeks Pembangunan TIK yang sangat tinggi. Hal itu terjadi karena nilai rata-rata dari masing-masing variabel menjadi paling tinggi daripada *cluster* yang lain.

4. Kesimpulan

Berdasarkan hasil dan pembahasan dalam penelitian ini, dapat disimpulkan bahwa metode *K-Medoids* dapat diimplementasikan dalam *clustering* provinsi berdasarkan Indeks Pembangunan TIK. *Cluster* yang dihasilkan divalidasi dengan *Davies-Bouldin Index*. Dari berbagai skenario jumlah *cluster* yang diterapkan (2 s.d. 5 *cluster*), diperoleh hasil bahwa 5 *cluster* merupakan jumlah *cluster* yang optimal untuk mengelompokkan provinsi di Indonesia berdasarkan Indeks Pembangunan TIK. *Cluster* pertama terdiri dari 1 provinsi yang dapat digolongkan sebagai *cluster* dengan Indeks Pembangunan TIK yang sangat rendah, sementara *cluster* kedua terdiri dari 6 provinsi yang dapat digolongkan sebagai *cluster* dengan Indeks Pembangunan TIK yang rendah. Selanjutnya, *cluster* ketiga terdiri dari 13 provinsi yang dapat digolongkan sebagai *cluster* dengan Indeks Pembangunan TIK yang sedang. Adapun *cluster* keempat terdiri dari 8 provinsi yang dapat digolongkan sebagai *cluster* dengan Indeks Pembangunan TIK yang tinggi, sedangkan *cluster* kelima terdiri dari 6 provinsi yang dapat digolongkan sebagai *cluster* dengan Indeks Pembangunan TIK yang sangat tinggi.

Mempertimbangkan hasil tersebut, masih terdapat beberapa provinsi di Indonesia yang tergolong sebagai *cluster* dengan Indeks Pembangunan TIK yang rendah dan sangat rendah. Hal ini dapat menjadi bahan evaluasi bagi pemerintah agar dapat memperhatikan provinsi-provinsi tersebut dan dapat memaksimalkan pembangunan TIK di Indonesia sehingga pembangunan TIK di setiap provinsi di Indonesia dapat merata dan dapat meminimalkan kesenjangan perkembangan kemajuan teknologi. Untuk penelitian yang

lebih lanjut, disarankan dapat menambah atau menggunakan variabel atau subindeks pembangunan TIK yang lainnya, yang belum digunakan dalam penelitian ini. Selain itu, juga dapat menggunakan metode yang lain selain *K-Medoids* untuk mengelompokkan provinsi di Indonesia berdasarkan Indeks Pembangunan TIK.

5. Daftar Pustaka

- [1] Suharni, "Data Mining Analisis *Cluster K-Means* pada Indeks Pembangunan Teknologi, Informasi dan Telekomunikasi," *Journal of Technopreneurship and Information System (JTIS)*, vol. 3, no. 3, pp. 92-96, 2020.
- [2] P. H. Parsaorantua, Y. Pasoreh and S. A. Rondonuwu, "Implementasi Teknologi Informasi dan Komunikasi (Studi Tentang Web E-Government di Kominfo Kota Manado)," *e-Jurnal Acta Diurna*, vol. 6, no. 3, pp. 1-14, 2017.
- [3] B. P. Statistik, *Indeks Pembangunan Teknologi Informasi dan Komunikasi 2021*, Jakarta: Badan Pusat Statistik, 2022.
- [4] R. Jayanthi and A. Dinaseviani, "Kesenjangan Digital dan Solusi yang Diterapkan di Indonesia Selama Pandemi Covid-19," *Jurnal IPTEK-KOM (Jurnal Ilmu Pengetahuan dan Teknologi Komunikasi)*, vol. 24, no. 2, pp. 187-200, 2022.
- [5] D. T. Larose, *Discovering Knowledge In Data: An Introduction to Data Mining*, Hoboken: John Wiley & Sons, 2014.
- [6] J. Han, M. Kamber and J. Pei, *Data Mining: Concepts and Techniques*, Waltham: Morgan Kaufmann, 2014.
- [7] T. Juninda, Mustakim and E. Andri, "Penerapan Algoritma *K-Medoids* untuk Pengelompokan Penyakit di Pekanbaru Riau," *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI)*, vol. 11, no. 1, pp. 42-49, 2019.
- [8] I. W. Septiani, A. C. Fauzan and M. M. Huda, "Implementasi Algoritma *K-Medoids* Denan Evaluasi *Davies-Bouldin-Index* untuk Klasterisasi Harapan Hidup Pasca Operasi pada Pasien Penderita Kanker Paru-Paru," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 3, no. 4, pp. 556-566, 2022.
- [9] I. Gunawan, G. Anggraeni, E. S. Rini, Y. M. Putri and Y. K. Zikri, "Klasterisasi Provinsi di Indonesia Berbasis Perkembangan Kasus Covid-19 Menggunakan Metode *K-Medoids*," *Seminar Nasional Matematika dan Pendidikan Matematika (5th SENATIK)*, pp. 301-306, 2020.
- [10] L. M. Nasution, "Statistika Deskriptif," *Jurnal Hikmah*, vol. 14, no. 1, pp. 49-55, 2017.
- [11] A. Sholikah, "Statistik Deskriptif dalam Penelitian Kualitatif," *Komunika*, vol. 10, no. 2, pp. 342-362, 2016.
- [12] S. A. Alasadi and W. S. Bhaya, "Review of Data Preprocessing Techniques in Data Mining," *Journal of Engineering and Applied Sciences*, vol. 12, no. 16, pp. 4102-4107, 2017.
- [13] D. Marlina, N. F. Putri, F. Andri and A. Ramadhan, "Implementasi Algoritma *K-Medoids* dan *K-Means* untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," *Jurnal CoreIT*, vol. 4, no. 2, pp. 64-71, 2018.
- [14] F. Hardiyanti, H. S. Tambunan and I. S. Saragih, "Penerapan Metode *K-Medoids Clustering* pada Penanganan Kasus Diare di Indonesia," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komunikasi)*, vol. 3, no. 1, pp. 598-603, 2019.
- [15] E. Luthfi and A. W. Wijayanto, "Analisis Perbandingan Metode *Hierarchical*, *K-Means*, dan *K-Medoids Clustering* dalam Pengelompokan Indeks Pembangunan Manusia Indonesia," *Inovasi*, vol. 17, no. 4, pp. 761-773, 2021.
- [16] M. N. P. Pamulang, M. N. Aini and U. Enri, "Komparasi Distance Measure pada *K-Medoids Clustering* untuk Pengelompokan Penyakit Ispa," *Edumatic: Jurnal Pendidikan Informatika*, vol. 5, no. 1, pp. 99-107, 2021.