

ANALISIS *CLUSTER K-MEANS* PADA INDIKATOR INDEKS PEMBANGUNAN TEKNOLOGI, INFORMASI, DAN KOMUNIKASI

Zuhana Realita Alfy¹, Ardhi Dinullah Baihaqie², Zakiah Fithah A'ini³
Program Studi Teknik Informatika^{1,2}, Program Studi Pendidikan Biologi³
Universitas Indraprasta PGRI^{1,2,3}
zuhanarealita28@gmail.com

Submitted February 27, 2023; Revised August 15, 2023; Accepted October 31, 2023

Abstrak

Indeks Pembangunan Teknologi, Informasi dan Komunikasi (IP-TIK) digunakan untuk mendeskripsikan adanya tingkat pembangunan TIK, ketimpangan dan peluang perkembangan suatu wilayah dari adanya penggunaan TIK. IP-TIK disusun dari 11 indikator dan penghitungannya dilakukan oleh Badan Pusat Statistik (BPS). Analisisnya menggunakan analisis *Cluster K-Means* yang berfungsi untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan kemiripan indikator penyusun IP-TIK. Pengelompokkan provinsi berdasarkan indikator IP-TIK akan membantu pemerintah menetapkan target yang realistis, melacak, dan mengevaluasi perkembangan dari waktu ke waktu untuk mendorong pembangunan dan pertumbuhan berdasarkan kemampuan setiap provinsi. Sehingga pemerintah mampu menentukan kebijakan yang tepat untuk membantu pembangunan TIK di setiap *cluster*. Hasil dari penelitian ini didapatkan 4 *cluster* dan masing-masing *cluster* memiliki karakteristik yang berbeda. *Cluster 1* dan *Cluster 2* memerlukan perbaikan untuk pembangunan IP-TIK.

Kata Kunci : *Cluster*, Indeks Pembangunan Teknologi Informasi dan Komunikasi (IP-TIK), *K-Means*

Abstract

The information and Communication Technology Development Index (IP-ICT) is used to describe the level of ICT Development, the imbalance and development opportunities of a region from the use of ICT. IP-TIK is composed of 11 indicators, with the calculation conducted by BPS. The analysis is a K-Means Cluster analysis used to group provinces in Indonesia based on the similarity of IP-ICT constituent indicators. Grouping provinces based on IP-TIK indicators will help the government set realistic targets, track and evaluate progress over time to promote development and growth based on the capabilities of each province, so that the government is able to determine the right policies to assist in ICT development in each cluster. The results of this study obtained 4 clusters with different characteristics each. Cluster 1 and Cluster 2 require improvement in IP-TIK development.

Keywords : *Cluster, Information and Communication Technology Development Index (IP-TIK), K-Means.*

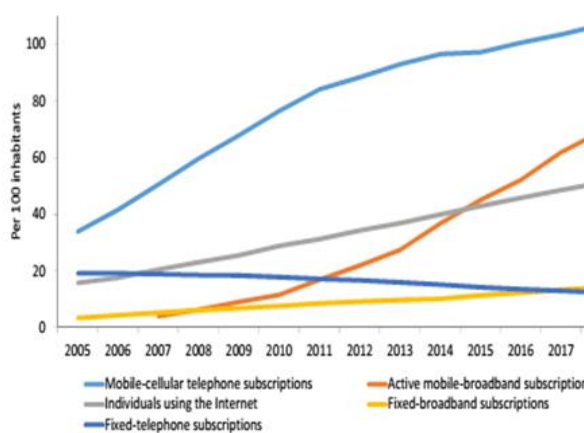
1. PENDAHULUAN

Teknologi, informasi, dan komunikasi (TIK) mengacu pada penggunaan teknologi digital untuk membuat, memproses, menyimpan, dan bertukar informasi [1]. Beberapa tahun belakangan ini, TIK terus menyebar ke seluruh negara-negara yang ada di dunia. Oleh sebab itu tingkat akses internet, kekayaan informasi, dan aplikasinya semakin meningkat pesat. Akses internet melalui jaringan seluler

telah berkembang pesat karena didukung oleh berbagai penemuan di bidang TIK, baik pada perangkat keras maupun perangkat lunak. Kecepatan akses internet juga meningkat seiring dengan tergantikannya internet *dial up* oleh internet *broadband* disebagian besar negara maju serta didukung pula oleh penurunan tarif internet. Sementara itu di negara berkembang terjadi revolusi telekomunikasi yang disebabkan

oleh berkembangnya teknologi telepon seluler.

Rata-rata tingkat penetrasi telepon seluler mencapai 49,5 persen pada akhir tahun 2008, meningkat dari mendekati nol pada sepuluh tahun yang lalu [2]. Jumlah pengguna telepon seluler mengalami peningkatan yang pesat, hingga pada akhir 2018 diperkirakan mencapai angka 107 per 100 penduduk. Perkembangan ini lebih cepat daripada perkembangan teknologi lainnya yang sudah ada sebelumnya.



Sumber : Badan Pusat Statistik

Gambar 1. Perkembangan TIK Global Tahun 2005-2018

Penggunaan TIK memiliki dampak yang potensial terhadap pembangunan sosial dan ekonomi, sehingga pemantauan berkelanjutan terhadap pembangunan TIK sangat penting bagi pemerintah dan pihak lain seperti pelaku usaha dalam mengambil kebijakan. IP-TIK digunakan untuk memenuhi kebutuhan data dan indikator TIK di Indonesia [3]. Indeks ini dihitung oleh Badan Pusat Statistik (BPS) sejak tahun 2016 dengan mengacu pada metodologi dari *International Telecommunication Union* (ITU).

Mengacu pada penelitian yang telah dilakukan oleh Khoirunnisa & W. Budiarti yang berjudul “Pengaruh Teknologi Informasi dan Komunikasi terhadap Tingkat Kemiskinan di Indonesia Tahun 2012-2017” menyatakan bahwa IP-TIK

berpengaruh terhadap derajat ukuran kemiskinan suatu wilayah, sehingga perlu adanya kemajuan dari infrastruktur, pemanfaatan TIK yang menyeluruh serta pelatihan TIK terhadap warga yang masuk kategori miskin untuk dapat memanfaatkan adanya TIK agar masyarakat tersebut dapat meningkatkan produktivitasnya [4].

Berdasarkan permasalahan yang ada, untuk selanjutnya peneliti bertujuan untuk melakukan *clusterisasi* provinsi yang ada di Indonesia dengan melihat adanya kemiripan dari indikator penyusun IP-TIK. Pengelompokan provinsi berdasarkan indikator IP-TIK akan membantu pemerintah menetapkan target yang realistis, melacak, dan mengevaluasi perkembangan dari waktu ke waktu untuk mendorong pembangunan dan pertumbuhan berdasarkan kemampuan setiap provinsi. Sehingga pemerintah mampu menentukan kebijakan yang tepat untuk membantu pembangunan TIK di setiap *cluster*.

2. METODE PENELITIAN

Analisis statistik yang digunakan adalah analisis *cluster*, yaitu suatu teknik dalam menganalisis data yang memiliki tujuan untuk mengelompokkan data kedalam beberapa kelompok individu atau objek yang memiliki karakteristik yang berbeda, sehingga individu atau objek tersebut berada dalam satu kelompok yang memiliki sifat relatif yang sama [5]. Sifat kemiripan tersebut diukur berdasarkan jarak [6].

Metode *cluster* dapat dibedakan menjadi dua, yaitu metode hirarki dan tidak berhirarki. Metode *cluster* termasuk kedalam metode yang tidak berhirarki, sehingga bekerja dengan menggunakan matriks jarak yang lebih kecil dan iterasi yang lebih singkat [7]. Hal ini menyebabkan metode ini lebih efisien karena proses analisis lebih cepat. Salah

satu metode *cluster* tidak berhirarki adalah analisis *cluster K-Means*.

Berikut merupakan tahapan dalam melakukan analisis *cluster K-Means* yaitu [8]:

1. Tentukan banyaknya *cluster*
Penentuan banyaknya *cluster* pada *K-means* dengan menggunakan metode *rule of thumb* [9].

$$K = \sqrt{\frac{n}{2}} \quad (1)$$

K = banyaknya cluster
n = banyaknya data

2. Membagi objek penelitian pada tahap *K-cluster* awal.
3. Lalu tiap objek dimasukkan ke suatu *cluster* yang memiliki rataan terdekat (jarak). Untuk menentukan jaraknya menggunakan *Euclidean*. Hitung ulang jarak untuk kelompok yang menerima objek baru dan yang mendapat pengurangan objek. Persamaan jarak *euclidean* adalah sebagai berikut [10] :

$$d(x_i, x_j) = \sqrt{\sum_{i=1}^n (x_i - x_j)^2} \quad (2)$$

4. Ulangi tahapan ke-2 sampai tidak terjadi perpindahan objek antar *cluster*.

Hasil akhir dari penentuan objek berpindah ke *cluster* tertentu tidak berdasarkan dari penentuan K pertama. Hal ini menunjukkan bahwa perubahan terbesar terjadi pada perpindahan yang pertama saja.

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) yaitu data IP-TIK dan indikator penyusun IP-TIK tahun 2018. Hanya 10 dari 11 indikator

penyusun IP-TIK yang dapat digunakan. Hal ini dikarenakan peubah *bandiwidth* untuk seluruh provinsi di Indonesia memiliki nilai yang sama. Indikator ini selanjutnya dianalisis menggunakan analisis *Cluster K-Means*.

Satuan pengamatan penelitian ini yaitu seluruh provinsi yang ada di Indonesia. Perangkat lunak yang digunakan dalam penelitian ini adalah Program R i386 4.0.1. *Package* yang digunakan antara lain *cluster*, *ggplot*, dan *factorextra*. *Package factorextra* digunakan untuk visualisasi *cluster* dan *ggplot* untuk menampilkan grafik. Sedangkan *Package Cluster* digunakan untuk pembentukan *cluster*.

Tabel 1. Variabel Penelitian

Kode Peubah	Peubah	Jenis Peubah
X ₁	IP-TIK	Numerik
X ₂	Pelanggan telepon tetap per 100 penduduk	Numerik
X ₃	Pelanggan telepon seluler per 100 penduduk	Numerik
X ₄	Persentase rumah tangga dengan komputer	Numerik
X ₅	Persentase rumah tangga dengan akses internet	Numerik
X ₆	Persentase individu yang menggunakan internet	Numerik
X ₇	Pelanggan <i>fixed broadband</i> per 100 penduduk	Numerik
X ₈	Pelanggan <i>mobile broadband</i> per 100 penduduk	Numerik
X ₉	Rata-rata lama sekolah	Numerik
X ₁₀	Angka partisipasi kasar sekunder	Numerik
X ₁₁	Angka partisipasi kasar tersier	Numerik

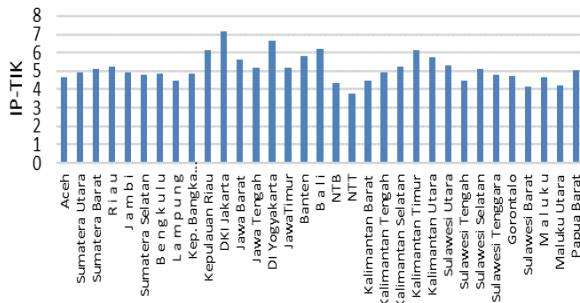
Sumber: Pribadi

3. HASIL DAN PEMBAHASAN

Deskripsi Data

Sebaran nilai IP-TIK Provinsi di Indonesia pada tahun 2018 berkisar antara 3,30 dan 7,14. DKI Jakarta memiliki IP-TIK

tertinggi yaitu sebesar 7,14. IP-TIK terendah dimiliki oleh Papua. Secara nasional, IP-TIK Indonesia pada tahun 2018 adalah sebesar 5,07. Terlihat bahwa sebagian besar provinsi di Pulau Sumatera memiliki nilai IP-TIK yang hampir homogen, sedangkan di Pulau Jawa sangat beragam.



Sumber : Badan Pusat Statistik

Gambar 2. IP-TIK Menurut Provinsi di Indonesia Tahun 2018

Jumlah Cluster

Untuk menentukan jumlah dari pengelompokkan ini adalah dengan menggunakan metode *rule of thomb*. Metode ini dapat diterapkan pada berbagai jenis data. Banyaknya objek dalam penelitian ini adalah 34, sehingga jumlah cluster yang didapatkan yaitu :

$$K = \sqrt{\frac{n}{2}} = \sqrt{\frac{34}{2}} = 4.12 = 4$$

Jadi jumlah cluster yang didapatkan yaitu sebanyak 4 (empat) *cluster*.

Hasil Analisis Cluster K-Means

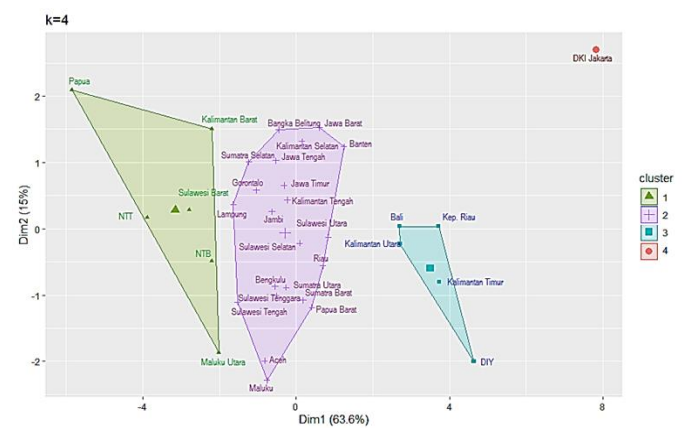
Analisis *cluster K-Means* pada penelitian ini menghasilkan 4 (empat) *cluster* sesuai dengan jumlah *cluster* yang sudah dihitung dengan metode *rule of thomb*, sehingga menghasilkan anggota *cluster 1* sebanyak 6 provinsi, *cluster 2* sebanyak 22 provinsi, *cluster 3* sebanyak 5 provinsi, dan *cluster ke-empat* sebanyak 1 provinsi. Daftar anggota *cluster* dapat dilihat pada Tabel 2.

Tabel 2. Hasil Analisis Cluster K-Means

Cluster	Jumlah Provinsi	Provinsi
1	6	Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Sulawesi Barat, Maluku Utara, Papua
2	22	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Maluku, Papua Barat
3	5	Kepulauan Riau, DI Yogyakarta, Bali, Kalimantan Timur, Kalimantan Utara
4	1	DKI Jakarta

Sumber : Pribadi

Secara umum, terlihat bahwa *cluster 1* dan *cluster 2* tidak memiliki perbedaan jarak yang terlalu jauh. Namun, perbedaan jarak semakin besar dari *cluster 2* menuju ke *cluster 3*, dan selanjutnya dari *cluster 3* menuju ke *cluster 4*. Terlihat bahwa *cluster 4* memiliki jarak yang sangat jauh dibandingkan dengan *cluster* lainnya dapat dilihat pada Gambar 3.



Sumber: Pribadi

Gambar 3. Visualisasi Hasil Analisis Cluster K-Means

Tabel 3. Rata-rata Peubah Tiap Cluster

Rata-rata Peubah	Cluster				Rataan umum
	1	2	3	4	
X ₁	4,05	5,01	6,19	7,14	5,08
X ₂	1,10	2,28	6,33	19,52	3,18
X ₃	103,42	122,24	132,61	134,50	120,81
X ₄	15,88	20,18	31,46	32,24	21,43
X ₅	46,54	63,21	77,27	89,04	63,10
X ₆	25,31	36,42	50,55	65,89	37,40
X ₇	1,02	2,38	6,51	21,52	3,31
X ₈	54,84	79,88	111,53	145,11	82,03
X ₉	7,79	8,79	9,51	11,06	8,79
X ₁₀	84,42	86,32	93,19	84,24	86,94
X ₁₁	28,51	32,56	36,82	36,04	32,57

Sumber : Pribadi

Cluster 1 terdiri dari provinsi-provinsi dengan nilai indikator IP-TIK yang paling rendah. Terlihat bahwa rata-rata dari 10 peubah (kecuali X₁₀) pada *cluster 1* paling rendah dibandingkan *cluster* lainnya. Hal ini menunjukkan bahwa provinsi-provinsi pada *cluster 1* memerlukan perbaikan pada dimensi akses dan infrastruktur, penggunaan, dan keahlian. Provinsi-provinsi pada *cluster 1* memerlukan perbaikan akses, infrastruktur, dan peningkatan keahlian agar dapat menggunakan TIK dengan sebaik mungkin.

Cluster 2 terdiri dari provinsi-provinsi dengan nilai indikator yang cukup lebih baik dibandingkan *cluster 1* terutama pada jumlah pelanggan telepon seluler (X₃), persentase rumah tangga dengan internet (X₅), dan rata-rata lama sekolah (X₉). Terlihat bahwa akses dan infrastruktur cukup memadai pada *cluster* ini. Namun masih memiliki cukup keterbatasan pada jumlah pelanggan telepon tetap (X₂) dan pelanggan *fixed broadband* (X₇).

Cluster 3 terdiri dari provinsi-provinsi yang sudah tercukupi baik dari dimensi akses dan infrastruktur, penggunaan, maupun keahlian. Sebagian besar provinsi di Indonesia termasuk di dalam *cluster 3*.

Cluster 4 hanya berisikan provinsi DKI Jakarta, hampir seluruh indikator memiliki nilai yang paling tinggi dibandingkan *cluster* lainnya. Hanya indikator angka partisipasi kasar sekunder (X₁₀) yang lebih kecil dibandingkan *cluster* lainnya. Nilai rata-rata pada X₁₀ bahkan lebih kecil dari nilai rata-rata pada *cluster 1*. Hal ini menunjukkan bahwa meskipun angka partisipasi kasar sekunder di DKI Jakarta kecil, namun masyarakat di DKI Jakarta tetap mampu untuk memaksimalkan penggunaan TIK. Indikator yang sangat berbeda jauh dibandingkan dengan *cluster* lain yaitu jumlah pelanggan telepon tetap (X₂) dan pelanggan *fixed broadband* (X₇) DKI Jakarta sebagai ibukota negara memiliki *provider* penyedia jasa internet *fixed broadband* yang cukup banyak dibandingkan provinsi lain.

4. SIMPULAN

Kesimpulan yang dihasilkan adalah dengan menggunakan analisis *cluster K-Means* provinsi-provinsi di Indonesia dapat dikelompokkan menjadi 4 *cluster*. Masing-masing *cluster* memiliki karakteristik yang berbeda. Pemerintah perlu memfokuskan pembangunan IP-TIK pada *cluster 1* dan 2. *Cluster 1* memerlukan perbaikan di seluruh indikator. Sangat penting bagi pemerintah untuk menyediakan infrastruktur TIK yang memadai, meningkatkan kualitas sumber daya manusia (SDM), dan kemudahan untuk mengakses TIK pada provinsi-provinsi di *Cluster 1*. Pada *cluster 2* pemerintah perlu meningkatkan akses internet.

DAFTAR PUSTAKA

- [1] J. Kimani and J. Scott, Information Communication Technology Diploma Level, USA: Finstock Evarsity Publisher, 2023.
- [2] International Technology Union (ITU), Measuring The Information Society The ICT Development Index.

- Geneva (CH): International Technology Union, 2018.
- [3] Badan Pusat Statistik, Indeks Pembangunan Teknologi Informasi dan Komunikasi (ICT Development Index) 2018. Jakarta (ID): Badan Pusat Statistik, 2019.
- [4] Khoirunnisa dan W. Budiarti, "Pengaruh TIK terhadap Tingkat Kemiskinan di Indonesia Tahun 2012-2017". Jakarta : Politeknik Statistika Jakarta, 2019.
- [5] M.W Talakua, Z.A. Leleury, dan A.W. Talluta, "Analisis Cluster dengan Menggunakan Metode K-Means untuk Pengelompokkan Kabupaten/Kota di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014", Jurnal Ilmu Matematika dan Terapan, vol.11, No.2, pp. 119-128, Ambon : Universitas Pattimura, 2017.
- [6] R. Scitovski., et all, Cluster Analysis and Applications. Switzerland : Springer Cham, 2021.
- [7] B. Sartono, D. K. Bodro dan G. A. Ditto, Teknik Eksplorasi Data yang Harus dikuasai Data Scientist. Bogor : PT. Penerbit IPB Press, 2021.
- [8] M. Nishom dan M.Y. Fathoni, "Implementasi Pendekatan Rule-Of-Thumb untuk Optimasi Algoritma K-Means Clustering", Jurnal Informatika: Jurnal Pengembangan IT, vol.3, No.2, pp. 237-241, 2018.
- [9] T.M. Kodinariya dan P.R. Makwana, "Review on Determining number of Cluster in K-means Clustering", *International Journal of Advance Research in Computer Science and Management Studies*, vol. 1, no. 6, pp. 90-95, 2013.
- [10] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-means Clustering Berbasis Chi-square", Jurnal Informatika: Jurnal Pengembangan IT,
- vol.4, no.1, pp. 20-24, Tegal : Politeknik Harapan Bersama, 2019.