

Data Mining Analisis Cluster K-Means Pada Indeks Pembangunan Teknologi, Informasi dan Telekomunikasi

Suharni¹

¹Program Studi Teknik Informatika, Fakultas Teknik, Universitas Islam Makassar
Jl. Perintis Kemerdekaan KM.9 No. 29, Tamalanrea Makassar, Sulawesi Selatan 90245, Indonesia
Email: suharni.dty@uim-makassar.ac.id

Abstract— Currently, Technology, Information and Communication (ICT) are an important factor in the progress of a region. To measure the potential for ICT development and the digital divide between region, it is necessary to do grouping. This study applies data mining using the clustering method with the k-means algorithm. The data source was obtained from Badan Pusat Statistik using the Technology, Information and communication Development Index data based on provinces in Indonesia in 2019. This study aims to classify areas based on the ICT Development Index in to 3 clusters. Based on the results of the k-means analysis, it is found that the low level (C1) consists of 4 provinces, East Nusa Tenggara, West Sulawesi, North Maluku, and Papua. The medium cluster (C2) consists of 22 provinces and the high cluster (C3) consists of 8 Provinces, Riau Island, DKI Jakarta, West Java, DI Yogyakarta, Banten, Bali, East Kalimantan and North Kalimantan. By conducting a cluster, the ICT Development Index can provide on overview of areas that still need improvement so that there is an even distribution of ICT development in every province in Indonesia.

Keywords—Data Mining, ICT Development Index, K-Means

Abstrak— Saat ini Teknologi, Informasi, dan Komunikasi (TIK) menjadi faktor penting kemajuan suatu wilayah. Untuk mengukur potensi pembangunan TIK dan kesenjangan digital antarwilayah, perlu dilakukan pengelompokan. Penelitian ini menerapkan datamining menggunakan metode clustering dengan algoritma k-means. Sumber data diperoleh dari Badan Pusat Statistik menggunakan data Indeks Pembangunan Teknologi, Informasi, dan Komunikasi (IP-TIK) berdasarkan provinsi di Indonesia tahun 2019. Penelitian ini bertujuan untuk mengklasifikasikan wilayah berdasarkan Indeks Pembangunan TIK ke dalam 3 cluster. Berdasarkan hasil analisis k-means diperoleh tingkat IP-TIK rendah (C1) terdiri atas 4 provinsi yaitu Nusa Tenggara Timur, Sulawesi Barat, Maluku Utara, dan Papua, cluster sedang (C2) terdiri atas 22 Provinsi dan cluster tinggi (C3) terdiri atas 8 Provinsi yaitu Kepulauan Riau, DKI Jakarta, Jawa Barat, DI Yogyakarta, Banten, Bali, Kalimantan Timur dan Kalimantan Utara. Dengan melakukan cluster Indeks Pembangunan TIK dapat memberikan gambaran wilayah-wilayah yang masih memerlukan perbaikan sehingga terjadi pemerataan pembangunan TIK di setiap provinsi di Indonesia.

Keywords— Data Mining, Indeks Pembangunan Teknologi, Informasi dan Komunikasi (IP-TIK), K-Means

I. Pendahuluan

Saat ini Teknologi, Informasi, dan Komunikasi (TIK) semakin berkembang pesat di seluruh dunia dan semakin banyak orang memiliki akses ke internet untuk berbagi informasi. Menurut data Badan Pusat Statistik, perkembangan penetrasi internet di Indonesia mengalami perkembangan positif yaitu dari 25,37 pada tahun 2016 menjadi 47,69 pada tahun 2019. Teknologi informasi dan komunikasi ini memungkinkan terjadinya pertukaran informasi secara cepat dan efisien. Di samping itu, perkembangan TIK juga mempengaruhi ekonomi, pembangunan manusia [1], memberikan peluang usaha dan penghasilan serta meningkatkan keefektifan pelayanan jasa masyarakat sehingga meningkatkan kualitas kehidupan [2]. Ahmad Helmy Fuady (2018) telah menguraikan hubungan antara perkembangan TIK dengan ketimpangan ekonomi di Indonesia. Analisis dalam tulisan ini menunjukkan bahwa perkembangan TIK di

Indonesia masih rendah, dan bahkan masih tertinggal jika dibandingkan negara-negara lain di kawasan Asia Tenggara. Sebagaimana pembangunan infrastruktur, pembangunan TIK antarwilayah provinsi juga masih sangat timpang. Hal ini mengindikasikan bahwa keuntungan kemajuan TIK belum dapat dinikmati secara merata di negara kepulauan yang sangat luas ini [3].

Untuk mengukur tingkat kemajuan Teknologi Informasi dan Komunikasi suatu daerah dapat ditunjukkan melalui Indeks Pembangunan Teknologi, Informasi, dan Komunikasi (IP-TIK). IP-TIK sangat penting sebagai ukuran standar tingkat pembangunan TIK di suatu wilayah. Selain itu juga dapat mengukur gap digital atau kesenjangan digital antarwilayah. Indeks pembangunan TIK ini disusun berdasarkan 11 indikator yang meliputi 3 subindeks yaitu akses dan infrastruktur, penggunaan dan keahlian. Semakin tinggi nilai indeks menunjukkan pembangunan TIK pada suatu wilayah semakin pesat, sebaliknya

semakin rendah nilai indeks menunjukkan pembangunan TIK di suatu wilayah relatif masih lambat [4].

Berdasarkan permasalahan tersebut, peneliti ingin menganalisis daerah berdasarkan pengelompokan tingkat IP-TIK tiap provinsi dengan menggunakan teknik data mining. Data mining adalah metode yang digunakan untuk memproses data, untuk menemukan gambar tersembunyi dari data yang diproses. Data yang diolah dengan metode data mining kemudian menghasilkan suatu pengetahuan baru yang bersumber dari data lama, hasil pengolahan data tersebut dapat digunakan sebagai informasi untuk menentukan keputusan di masa mendatang [5]. Salah satu metode data mining untuk pengelompokan adalah metode *clustering*. Metode *Clustering* disebut juga klasifikasi *unsupervised* merupakan proses mengkategorikan sekumpulan data menjadi kelompok yang homogen (*cluster*). Banyak metode yang bisa digunakan untuk melakukan *clustering* salah satunya adalah metode *K-Means*. *K-Means* adalah salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok. Metode ini mempartisi data ke dalam cluster/ kelompok sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu cluster yang sama [6]. Untuk menguji validitas menggunakan *Index Davies-Bouldin (IDB)* yang bertujuan untuk memvalidasi data pada masing-masing *cluster* untuk melihat apakah kelompok yang terbentuk sudah baik atau tidak [7].

Dengan melakukan cluster Indeks Pembangunan Teknologi Informasi dan Komunikasi dapat memberikan gambaran wilayah-wilayah yang masih memerlukan perbaikan di Indonesia.

II. Metode Penelitian

A. Data Set

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh berdasarkan survei dan pendataan yang dilakukan oleh Badan Pusat Statistik (BPS). Data yang digunakan adalah data Indeks Pembangunan Teknologi, Informasi dan Komunikasi (IP-TIK) tahun 2019 terdiri dari 34 Provinsi di Indonesia.

IP-TIK merupakan suatu indeks komposit yang disusun oleh tiga subindeks, dan masing-masing subindeks terdiri dari indikator penyusun subindeks. Subindeks penyusun IP-TIK, yaitu:

1. Subindeks akses dan infrastruktur (X1), menggambarkan kesiapan TIK (*ICT readiness*) yang diukur dari sisi akses dan infrastruktur TIK.
2. Subindeks penggunaan (X2), menggambarkan intensitas TIK (*ICT intensity*) yang diukur dari penggunaan TIK.

3. Subindeks keahlian (X3), menggambarkan kemampuan atau keahlian yang diperlukan dalam TIK (*ICT Skill*).

TABEL I
DATA IP-TIK MENURUT PROVINSI BERDASARKAN SUB INDEKS TAHUN 2019

NO.	Provinsi	X1	X2	X3
1	Aceh	5.33	3.54	6.71
2	Sumatra Utara	5.64	4.21	6.24
3	Sumatra Barat	5.78	4.11	6.41
4	Riau	5.79	4.45	6.15
5	Jambi	5.71	4.3	5.8
6	Sumatra Selatan	5.64	3.84	5.55
7	Bengkulu	5.73	4.13	6.29
8	Lampung	5.28	4	5.52
9	Kep. Bangka Belitung	5.85	4.65	5.19
10	Kep. Riau	7.03	5.91	6.08
11	DKI Jakarta	8.03	6.99	6.53
12	Jawa Barat	6.52	5.34	5.57
13	Jawa Tengah	6.18	4.83	5.46
14	DI Yogyakarta	7.86	5.67	7.49
15	Jawa Timur	6	4.86	5.73
16	Banten	6.38	5.43	5.85
17	Bali	6.94	5.52	6.26
18	Nusa Tenggara Barat	5.28	3.88	5.84
19	Nusa Tenggara Timur	4.89	2.67	5.65
20	Kalimantan Barat	5.47	3.71	5.31
21	Kalimantan Tengah	5.64	4.55	5.65
22	Kalimantan Selatan	5.77	4.91	5.51
23	Kalimantan Timur	6.87	5.52	6.53
24	Kalimantan Utara	6.06	5.63	6.1
25	Sulawesi Utara	5.69	4.58	6.2
26	Sulawesi Tengah	5.38	3.45	6.22
27	Sulawesi Selatan	5.68	4.28	6.17
28	Sulawesi Tenggara	5.64	4.04	6.43
29	Gorontalo	5.45	4.07	5.78
30	Sulawesi Barat	5.08	3.02	5.61
31	Maluku	5.25	3.23	6.89
32	Maluku Utara	4.79	2.79	5.51
33	Papua Barat	5.6	4.17	6.41
34	Papua	3.82	2.1	4.79

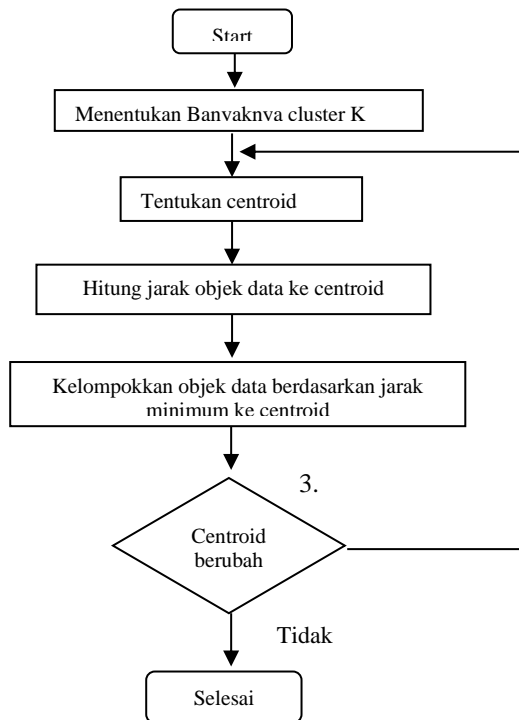
B. Pengolahan Data

Pada tahap pengolahan data, aplikasi tools yang digunakan adalah RapidMiner 9.8. RapidMiner merupakan *software* untuk pengolahan data. Dengan menggunakan prinsip dan algoritma data mining, RapidMiner mengekstrak pola-pola dari data set yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan dan database. RapidMiner memudahkan penggunaanya dalam melakukan perhitungan data yang sangat banyak dengan menggunakan operator-operator yang berfungsi untuk memodifikasi data. Data dihubungkan dengan node-node pada operator kemudian kita hanya tinggal menghubungkannya ke node hasil untuk melihat hasilnya. Hasil yang diperlihatkan RapidMiner pun dapat ditampilkan secara visual dengan grafik [8].

Data dibagi menjadi 3 cluster, yaitu cluster provinsi dengan IP-TIK tertinggi, sedang dan rendah yang hasilnya akan dianalisis

C. Algoritma K-Means

Proses algoritma K-Means tertuang dalam flowchart dibawah ini:



Gbr.1 Proses K-Means

Berdasarkan gbr.1 tahapan algoritma K-Means dapat dijabarkan sebagai berikut:

1. Mulai
2. Menentukan nilai k sebagai jumlah klaster yang ingin dibentuk.
3. Menentukan cluster (titik centroid)
4. Menghitung jarak data ke centroid

$$D(ij) = \sqrt{(X_{ki} - X_{kj})^2} \quad (1)$$
5. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terkecil)
6. Hitung kembali centroid C berdasarkan data yang mengikuti cluster masing-masing

$$R_k = \frac{1}{N_k} (X_{1k} + X_{2k} + \dots + X_{nk}) \quad (2)$$
 Di mana:
 R_k = Rata-rata baru
 N_k = Jumlah *training pattern* pada *cluster* (k)
 X_{nk} = Pola ke- n yang menjadi bagian *cluster* ke- k
7. Mengulangi perulangan dari langkah 3-6 hingga kondisi konvergen tercapai.

III. Hasil dan Diskusi

Adapun langkah dalam cluster dengan algoritma K-Means yaitu :

1. Menentukan Jumlah Cluster
 Jumlah cluster yang digunakan pada pengelompokan Indeks Pembangunan TIK sebanyak 3 cluster diantaranya tinggi, sedang dan rendah berdasarkan data IP-TIK tahun 2019.

2. Menentukan centroid
 Dalam penerapan algoritma *K-means* dihasilkan nilai titik tengah atau *centroid* dari data. Penentuan titik awal cluster ini dilakukan dengan mengambil nilai tertinggi di wilayah yang memiliki Indeks Pembangunan TIK tertinggi ($C1$), nilai rata-rata IP-TIK untuk cluster tingkat sedang ($C2$) dan nilai rata-rata IP-TIK terendah untuk cluster tingkat rendah ($C3$). Berikut adalah centroid awal pada iterasi pertama dapat diketahui pada Tabel 2 berikut:

TABEL 2
CENTROID DATA AWAL

Atribut	Akses dan Infrastruktur	Penggunaan	Keahlian
Max (c1)	8,03	6,99	7,49
Average (c2)	5,825	4,36	5,98
Min (c3)	3,82	2,1	4,79

3. Menghitung Jarak Dari Centroid
 Untuk menghitung jarak antara data dengan pusat *cluster* menggunakan persamaan (1). Berikut adalah hasil perhitungan jarak antara data dengan pusat *cluster*:

TABEL 3
PERHITUNGAN ITERASI CLUSTER PUSAT PERTAMA

NO.	Provinsi	C1	C2	C3
1	Aceh	4.450	1.204	2.836
2	Sumatra Utara	3.873	0.353	3.141
3	Sumatra Barat	3.811	0.499	3.241
4	Riau	3.642	0.196	3.355
5	Jambi	3.934	0.222	3.071
6	Sumatra Selatan	4.404	0.700	2.630
7	Bengkulu	3.861	0.398	3.165
8	Lampung	4.515	0.799	2.505
9	Kep. Bangka Belitung	3.939	0.842	3.284
10	Kep. Riau	2.038	1.966	5.146
11	DKI Jakarta	0.960	3.476	6.683
12	Jawa Barat	2.948	1.269	4.289
13	Jawa Tengah	3.494	0.786	3.670
14	DI Yogyakarta	1.331	2.853	6.030
15	Jawa Timur	3.429	0.586	3.641
16	Banten	2.801	1.212	4.332
17	Bali	2.205	1.633	4.857
18	Nusa Tenggara Barat	4.467	0.740	2.530
19	Nusa Tenggara Timur	5.649	1.959	1.486
20	Kalimantan Barat	4.697	0.999	2.363
21	Kalimantan Tengah	3.880	0.423	3.171
22	Kalimantan Selatan	3.654	0.726	3.495
23	Kalimantan Timur	2.104	1.655	4.902
24	Kalimantan Utara	2.768	1.297	4.381
25	Sulawesi Utara	3.598	0.339	3.411
26	Sulawesi Tengah	4.601	1.041	2.510
27	Sulawesi Selatan	3.822	0.252	3.181
28	Sulawesi Tenggara	3.942	0.582	3.125
29	Gorontalo	4.255	0.515	2.742
30	Sulawesi Barat	5.291	1.577	1.762
31	Maluku	4.714	1.561	2.781
32	Maluku Utara	5.662	1.938	1.391
33	Papua Barat	3.876	0.521	3.175
34	Papua	6.995	3.247	0.000

4. *Clustering* Data
 Proses iterasi cluster pertama dilakukan dengan mengambil jarak terdekat masing-masing cluster data yang diproses. Proses clustering dengan menggunakan nilai centroid awal yang terdapat pada tabel 3, akan memperoleh hasil

pengelompokan pada iterasi 1 yang dapat dilihat pada tabel 4 berikut:

Tabel 4
Pengelompokan Data Iterasi Pertama

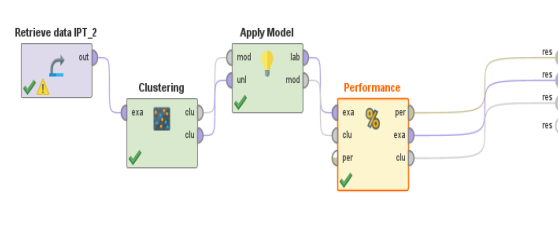
NO.	Provinsi	C1	C2	C3
1	Aceh		1	
2	Sumatra Utara		1	
3	Sumatra Barat		1	
4	Riau		1	
5	Jambi		1	
6	Sumatra Selatan		1	
7	Bengkulu		1	
8	Lampung		1	
9	Kep. Bangka Belitung		1	
10	Kep. Riau		1	
11	DKI Jakarta	1		
12	Jawa Barat		1	
13	Jawa Tengah		1	
14	DI Yogyakarta	1		
15	Jawa Timur		1	
16	Banten		1	
17	Bali		1	
18	Nusa Tenggara Barat		1	
19	Nusa Tenggara Timur			1
20	Kalimantan Barat		1	
21	Kalimantan Tengah		1	
22	Kalimantan Selatan		1	
23	Kalimantan Timur		1	
24	Kalimantan Utara		1	
25	Sulawesi Utara		1	
26	Sulawesi Tengah		1	
27	Sulawesi Selatan		1	
28	Sulawesi Tenggara		1	
29	Gorontalo		1	
30	Sulawesi Barat		1	
31	Maluku		1	
32	Maluku Utara			1
33	Papua Barat		1	
34	Papua			1
		2	29	3

Dari data Indeks Pembangunan TIK menurut Provinsi, ditemukan pengelompokan pada iterasi pertama untuk 3 cluster. Kelompok regional indeks Pembangunan TIK tertinggi (C1) terdiri atas 2 provinsi. Regional Indeks Pembangunan TIK sedang (C2) terdiri atas 29 provinsi dan regional Indeks Pembangunan TIK terendah (C3) terdiri atas 3 provinsi.

Proses iterasi akan terus berlangsung sampai hasil iterasi terakhir sama dengan hasil iterasi sebelumnya. Proses nilai titik tengah atau *centroid* akan menyesuaikan sesuai dengan iterasi yang ada. Dalam hal ini proses akan dilanjutkan dengan menggunakan aplikasi *RapidMiner*.

5. Penerapan aplikasi *RapidMiner*

Penggunaan algoritma K-Means untuk pengelompokan Indeks Pembangunan Teknologi, Informasi dan Komunikasi (IP-TIK) berdasarkan 3 atribut subindeks dengan menggunakan software *RapidMiner 9.8* terbagi menjadi 3 cluster, yaitu tinggi, sedang, dan rendah. Berikut adalah hasilnya:



Gbr.2 Proses K-Means dengan RapidMiner

Cluster Model

Cluster 0: 8 items
Cluster 1: 4 items
Cluster 2: 22 items
Total number of items: 34



Gbr.3 Hasil Clustering Menggunakan K-Means

Berdasarkan Gbr.3 dapat dijelaskan bahwa dari 3 cluster yang dihasilkan terdapat 8 Provinsi pada cluster_0 yaitu Kep. Riau, DKI Jakarta, Jawa Barat, DI Yogyakarta, Banten, Bali, Kalimantan Timur dan Kalimantan Utara, 4 provinsi pada cluster_1 Nusa Tenggara Timur, Sulawesi Barat, Maluku Utara, dan Papua. dan 22 Provinsi untuk cluster_2 yaitu Aceh, Sumatra Utara, Sumatra Barat, Riau, Jambi, Sumatra selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Jawa Tengah, Jawa Timur, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Barat, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Tengah, Nusa Tenggara Barat, Papua Barat, dan Maluku

Untuk menentukan klasifikasi cluster yang ditentukan oleh hasil centroid pada Gbr.4

Attribute	cluster_0	cluster_1	cluster_2
Akses dan Infrastruk...	6.961	4.645	5.626
Penggunaan	5.751	2.645	4.172
Keahlian	6.301	5.390	5.975

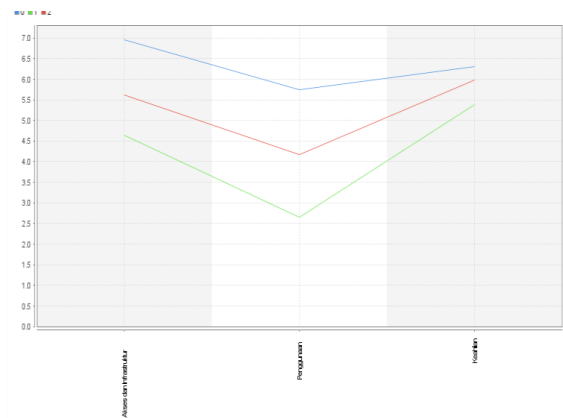
Gbr. 4 Tabel Centroid

Berdasarkan Gbr.4 diperoleh bahwa dari ketiga atribut, cluster_0 merupakan cluster IP-TIK tinggi, cluster_1 merupakan cluster IP-TIK rendah dan cluster_2 merupakan cluster IP-TIK sedang.

Berikut hasil tampilan plot cluster IP-TIK berdasarkan Provinsi di Indonesia disajikan pada Gbr.5 dan Gbr.6 menunjukkan grafik hubungan centroid dari tiap cluster.



Gbr 5. Plot IP-TIK Berdasarkan Cluster



Gbr 6. Hubungan Titik Centroid dari Tiap Cluster

Melihat performa K-Means dengan menggunakan tools RapidMiner adalah menambahkan operator % Performance yang dapat mengevaluasi performa algoritma clustering berbasis centroid. Diperoleh hasil *Davies_Bouldin* sebesar -0,668

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -0.575
Avg. within centroid distance_cluster_0: -0.952
Avg. within centroid distance_cluster_1: -0.475
Avg. within centroid distance_cluster_2: -0.457
Davies Bouldin: -0.668
```

Gbr.7 Performance Vector K-Means

IV. Kesimpulan

Berdasarkan hasil penelitian dapat dilihat bahwa data mining dengan metode k-means dapat diterapkan dalam pengelompokan data Indeks Pembangunan TIK di Indonesia. Diperoleh hasil terdapat beberapa provinsi yang memiliki tingkat Indeks Pembangunan TIK yang tinggi terdiri atas 8 Provinsi atau sebesar 24% yaitu provinsi Kep. Riau, DKI Jakarta, Jawa Barat, DI Yogyakarta, Banten, Bali, Kalimantan Timur dan Kalimantan Utara, Indeks Pembangunan TIK cluster sedang sebanyak 22 Provinsi atau sebesar 64%, sedangkan Indeks Pembangunan TIK yang rendah sebanyak 4 Provinsi atau sebesar 12% dari total Provinsi di Indonesia yang terdiri atas provinsi Nusa Tenggara Timur, Sulawesi Barat, Maluku Utara, dan Papua. Berdasarkan hasil yang diperoleh Provinsi di Indonesia didominasi oleh cluster Indeks Pembangunan TIK kategori sedang dan masih ada Provinsi yang memiliki IP-TIK rendah. Hal Ini merupakan masukan bagi pemerintah yang berwenang menjadi bahan informasi untuk memperhatikan wilayah-wilayah tersebut sehingga terjadi pemerataan pembangunan TIK di setiap provinsi di Indonesia.

Referensi

- [1] B. D. E. L. A. Hoz-rosales, J. Antonio, C. Ballesta, I. Tamayo-torres, and K. Buelvas-ferreira, "Effects of Information and Communication Technology Usage by Individuals, Businesses, and Government on Human Development: An International Analysis," *IEEEAccess*, vol. 7, pp. 129225–129243, 2019.
- [2] W. Iniasari, "Analisis Runtun Waktu dalam Pengujian Pengaruh TIK terhadap Penurunan Laju Kemiskinan di Indonesia," *IPTEKKOM*, vol. 17, no. 1, pp. 19–30, 2015.
- [3] A. H. Fuady, "Teknologi Digital dan Ketimpangan Ekonomi di Indonesia," *Masy. Indones.*, vol. 41(1), pp. 75–88, 2018.
- [4] BPS RI, *Indeks Pembangunan Teknologi, Informasi dan Komunikasi (ICT Development Index) 2018*. 2019.
- [5] D. Nofriansyah, *Konsep Data Mining Vs Sistem Pendukung Keputusan*. Yogyakarta: Deepublish, 2014.
- [6] Eko Prasetyo, *Data Mining: Konsep Dan Aplikasi Menggunakan MATLAB*. Yogyakarta: PENERBIT ANDI, 2012.
- [7] H. Rehioui, A. Idrissi, M. Abourezq, and F. Zegrari, "DENCLUE-IM: A New Approach for Big Data Clustering," *Procedia - Procedia Comput. Sci.*, vol. 83, no. Ant 2016, pp. 560–567, 2016.
- [8] B. R. C. T. I et al., "Implementasi k-means clustering pada rapidminer untuk analisis daerah rawan kecelakaan," 2017, no. April, pp. 58–62.