

Analisis *Cluster* dengan *K-Means* untuk Pengelompokan Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Pembangunan TIK Tahun 2021-2022

Rahmatina Hidayati ^{1*}, Luthfi Indana ², Mochammad Daffa Putra Karyudi ³, Redoti Zulfikar Sasongko ⁴

^{1,2,3,4} Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Merdeka Malang, Kota Malang, Provinsi Jawa Timur, Indonesia.

article info

Article history:

Received 16 November 2023

Received in revised form

10 January 2024

Accepted 20 March 2024

Available online April 2024.

DOI:

<https://doi.org/10.35870/jti.k.v8i2.1815>.

Keywords:

K-Means; Regency/City Grouping; ICT Development; East Java Province; Cluster Analysis.

Kata Kunci:

K-Means; Pengelompokan Kabupaten/Kota; Pembangunan TIK; Provinsi Jawa Timur; Analisis Cluster.

abstract

The National Statistics Agency has grouped the IP-TIK by province into 4 categories (high, medium, low and very low). However, the details of IP-TIK in each province have not been explained in detail by city or district. The province chosen was East Java. In this province, the Human Development Index (HDI) is at a high level, but there is inequality in the distribution. This research aims to explore information on whether there are gaps in ICT development using a data mining approach, namely grouping. This is so that the East Java Provincial government focuses more on ICT development at the lower level. The method used is k-Means with Euclidean distance calculations. The highest group results are in *cluster* 2, which indicates that the IP-ICT level in East Java province is on average moderate. Group inequality occurred in *cluster* 4 for the household having a computer. Those included in *cluster* 4 (low category) include Pacitan Regency, Lumajang Regency, Probolinggo Regency, Bangkalan Regency, Sampang Regency and Pamekasan Regency.

abstrak

Badan Pusat Statistik Nasional sudah mengelompokkan IP-TIK berdasarkan provinsi dalam 4 kategori (tinggi, sedang, rendah, dan sangat rendah). Namun rincian IP-TIK di masing-masing provinsi belum dijelaskan detail berdasarkan Kota atau Kabupaten. Provinsi yang dipilih adalah Jawa Timur. Pada provinsi tersebut, Indeks Pembangunan Manusia (IPM)-nya berada pada tingkat tinggi, namun sebarannya terjadi ketimpangan. Penelitian ini bertujuan untuk menggali informasi apakah ada kesenjangan pembangunan TIK dengan pendekatan data mining yakni pengelompokan. Hal ini supaya pemerintah Provinsi Jawa Timur lebih menfokuskan pembangunan TIK pada tingkatan rendah. Metode yang digunakan adalah k-Means dengan perhitungan jarak Euclidean. Hasil kelompok terbanyak berada pada *cluster* 2 yang menandakan tingkat IP-TIK di provinsi Jawa Timur rata-rata sedang. Ketimpangan kelompok terjadi pada *cluster* 4 untuk variabel RT memiliki komputer yang mana nilai minimumnya 2,26 dan maksimumnya 11,75. Yang termasuk pada *cluster* 4 (kategori rendah) antara lain Kab. Pacitan, Kab. Lumajang, Kab. Probolinggo, Kab. Bangkalan, Kab. Sampang, dan Kab. Pamekasan.

Corresponding Author. Email: rahmatina.hidayati@unmer.ac.id ^{1}.

© E-ISSN: 2580-1643.

Copyright © 2024 by the authors of this article. Published by Lembaga Otonom Lembaga Informasi dan Riset Indonesia (KITA INFO dan Riset). This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. 



ACM Computing Classification System (CCS)

EBSCOhost

Communication and Mass Media Complete (CMMC)

1. Latar Belakang

Indeks Pembangunan Teknologi Informasi dan Komunikasi (IP-TIK) merupakan alat ukur yang menggambarkan tingkat pembangunan TIK di suatu wilayah. Selain itu, IP-TIK dapat mengukur kesenjangan digital serta menginformasikan potensi pengembangan TIK [1]. Meningkatkan IP-TIK merupakan salah satu upaya meningkatkan Indeks Pembangunan Manusia (IPM). Berdasarkan penelitian [2] TIK memiliki pengaruh yang positif bagi pembangunan manusia di kawasan Indonesia Timur. Hal ini juga sejalan dengan [3] yang melakukan analisis pengaruh TIK terhadap IPM negara ASEAN tahun 2014-2019. Badan Pusat Statistik Nasional sudah mengelompokkan IP-TIK berdasarkan provinsi dalam 4 kategori yakni tinggi, sedang, rendah, dan sangat rendah. Namun rincian IP-TIK di masing-masing provinsi belum dijelaskan secara detail berdasarkan Kota atau Kabupaten. Maka dari itu, perlu dilakukan analisis IP-TIK di tingkat provinsi.

Pada penelitian ini penulis akan mengelompokkan tingkat pembangunan teknologi informasi dan komunikasi di Jawa Timur untuk mengetahui sebarannya. Bila mengacu pada Indeks Pembangunan Manusia (IPM) pada tahun 2021 Provinsi Jawa Timur berada pada kategori tinggi [4], namun sebarannya terjadi ketimpangan. Beberapa Kota/Kabupaten mengalami pertumbuhan pembangunan sangat cepat, sedangkan Kota/Kabupaten lain mengalami pertumbuhan yang lambat [5]. Untuk itu, penelitian ini bertujuan untuk menggali informasi apakah ada kesenjangan pembangunan TIK, supaya pemerintah Provinsi Jawa Timur lebih menfokuskan pembangunan TIK pada tingkatan yang masih rendah.

Indikator yang akan digunakan dalam pengelompokan pembangunan TIK antara lain akses dan infrastruktur TIK, serta penggunaan TIK [1]. Data yang digunakan berupa persentase dari indikator tersebut pada Kabupaten dan Kota di Provinsi Jawa Timur tahun 2021-2022. Sedangkan metode yang akan digunakan adalah *K-Means clustering*. Pemilihan metode ini karena membutuhkan

waktu yang lebih cepat dibandingkan dengan *K-Medoid* [6]. Langkah awal proses pengelompokan adalah menentukan jumlah *cluster*. Hartanti (2020) menerapkan *Indeks Davies Bouldin* (DBI) untuk optimasi jumlah *cluster* pada metode K-Means untuk mengelompokkan Provinsi Berdasarkan Potensi Desa [7]. Sedangkan Dewi & Pramita (2019) menggunakan Elbow untuk penentuan jumlah *cluster* [8]. Dalam penelitian Anggara, Sujaini, & Nasution (2016), penggunaan *Sillhouette* untuk menentukan jumlah *cluster* dan menghasilkan nilai DBI yang lebih baik dibandingkan metode Elbow pada metode *K-Medoid* [9]. Pada penelitian ini, penulis akan menerapkan metode Elbow untuk optimasi jumlah *cluster* pada metode *K-Means*, *Euclidean Distance* untuk perhitungan jarak minimum dan menentukan kelompok [10], sedangkan pengujiannya menggunakan *Sillhouette Coefficient*.

2. Metode Penelitian

Tahapan penelitian untuk mengelompokkan Kota/Kabupaten di Provinsi Jawa Timur berdasarkan Pembangunan TIK adalah sebagai berikut:

1) Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari Badan Pusat Statistik Provinsi Jawa Timur tahun 2021-2022 [4][11][12]. Datanya berupa:

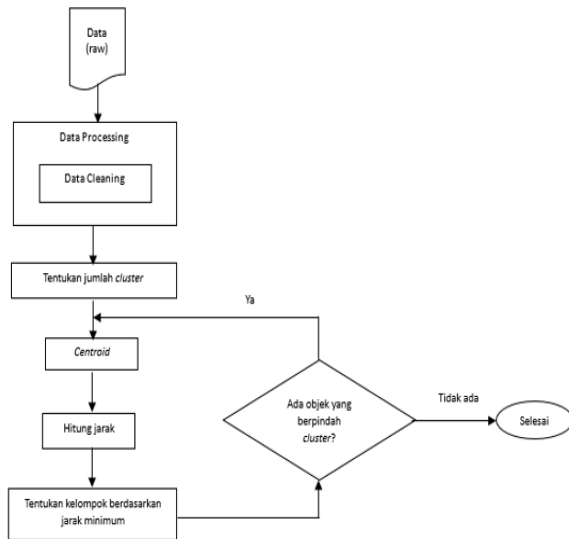
- Persentase rumah tangga yang memiliki komputer: komputer tersebut tersedia dan digunakan oleh semua anggota rumah tangga
- Persentase rumah tangga yang memiliki telepon seluler dalam 3 bulan terakhir
- Persentase rumah tangga dengan akses internet dalam 3 bulan terakhir
- Persentase individu yang menggunakan internet: individu yang dimaksud merupakan penduduk 5 tahun ke atas yang mengakses internet dalam 3 bulan terakhir
- Indeks Pembangunan Manusia (IPM)

Tabel 1 menampilkan Kota/Kabupaten Jawa Timur dengan sebagian variabel yang sudah disebutkan di atas untuk tahun 2021 yang terdiri 38 baris.

Tabel 1. Persertanse Variabel Tahun 2021 [11]

Kota/Kab	RT Komputer	5+ th akses internet	IPM
Pacitan	16,45	46,88	68,57
Ponorogo	25,73	63,09	71,06
Trenggalek	16,27	55,75	70,06
Tulungagung	24,64	62,51	73,15
Blitar	24,42	64,71	71,05
Kediri	23,34	58,2	72,56
Malang	19,67	61,36	70,6
Lumajang	17,99	45,1	66,07
Jember	18,86	49,46	67,32
Banyuwangi	23,43	57,78	71,38
Bondowoso	20,13	43,88	66,59
Situbondo	14,13	42,61	67,78
Probolinggo	14,16	43,2	66,26
Pasuruan	16,99	54,29	68,93
Sidoarjo	44,28	77,97	80,65
Mojokerto	25,94	65,02	74,15
Jombang	25,9	63,05	73,45
Nganjuk	19,94	55,56	71,97
Madiun	25,48	59,03	71,88
Magetan	27,17	59,89	74,15
Ngawi	17,95	56,24	71,04
Bojonegoro	21,83	53,84	69,59
Tuban	21,24	56,9	68,91
Lamongan	22,98	56,88	73,12
Gresik	30,38	72,51	76,5
Bangkalan	7,83	45,74	64,36
Sampang	1,59	38	62,8
Pamekasan	15,32	39,87	66,4
Sumenep	15,36	40,44	67,04
Kota Kediri	45,54	79,05	78,6
Kota Blitar	50,8	77,94	78,98
Kota Malang	50,64	82,96	82,04
Kota Probolinggo	37,47	67,26	73,66
Kota Pasuruan	34,86	73,99	75,62
Kota Mojokerto	44,01	77,1	78,43
Kota Madiun	53,6	80,9	81,25
Kota Surabaya	38,27	81,05	82,31
Kota Batu	35,79	73,45	76,28

Data yang telah diperoleh kemudian diolah menggunakan algoritma K-Means untuk melakukan proses pengelompokan. Gambar 1 mengilustrasikan tahapan proses *clustering* menggunakan metode K-Means.



Gambar 1. Flowchart Metode K-Means

2) Data Preprocessing

Pada penelitian ini, proses persiapan data yang dilakukan adalah data *cleaning* yakni menghilangkan data yang tidak memiliki nilai (null), tidak relevan, dan duplikat [13].

3) Penentuan jumlah *cluster* (k)

Penentuan jumlah *cluster* (k) menggunakan Elbow. Algoritma Elbow dalam menentukan nilai k pada *K-Means* [9]:

- Mulai
- Inisialisasi awal nilai k
- Naikkan nilai k
- Hitung hasil SSE seperti persamaan 1 dari tiap nilai k.

$$SSE = \sum_{K=1}^K \sum_{X_i} |x_i - c_k|^2$$

Keterangan:

K : *Cluster* ke-c

x_i : Jarak data obyek ke-i

c_k : Pusat *cluster* ke-i

Selain rekomendasi dari Elbow, dalam penelitian ini penentuan jumlah *cluster* juga dilakukan berdasarkan kategori dari pemerintah yakni 4.

4) Tentukan nilai centroid awal

5) Clustering

Hitung jarak data dengan centroid awal dan lakukan pengelompokan berdasarkan jarak terdekat. Lakukan proses ini sampai mencapai

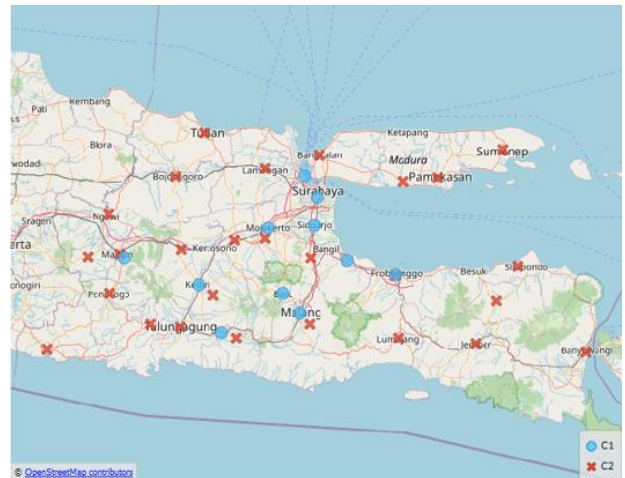
konvergensi=0 yang menandakan hasil *clustering* saat ini sama dengan *clustering* sebelumnya. Pada penelitian ini perhitungan jarak yang akan digunakan adalah *Euclidean Distance* seperti pada persamaan 3 [10].

$$d_{euc}(x, y) = \sqrt{\sum_{j=1}^d (x_j - y_j)^2}$$

- Hitung nilai *Silhouette Coefficient*
- Analisis karakteristik *cluster*

3. Hasil dan Pembahasan

Berdasarkan perhitungan algoritma Elbow pada RStudio, jumlah *cluster* yang direkomendasikan adalah 2. Selanjutnya dilakukan pengelompokan 3 skenario yakni menggunakan data tahun 2021, tahun 2022 secara terpisah dan pengelompokan dengan menggabungkan keduanya. Skenario pengelompokan tersebut menunjukkan hasil yang sama, 11 kota/kabupaten berada di *cluster* 1 (kategori tinggi) dan 27 berada di *cluster* 2 (kategori rendah). Gambar 2 menampilkan peta *cluster*. Tanda bulat biru untuk *cluster* 1, sedangkan X merah untuk *cluster* 2.

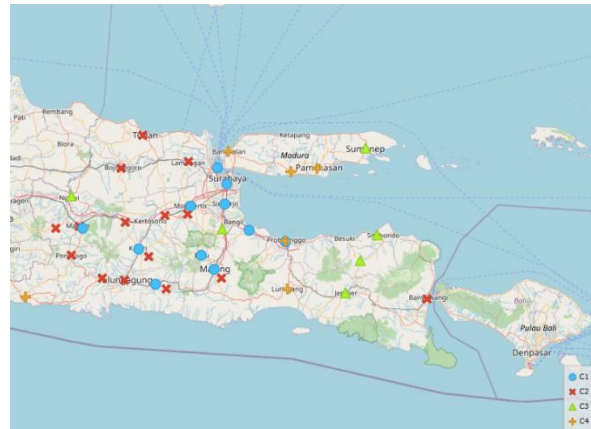
Gambar 2. Hasil Pengelompokan Tingkat IP-TIK dengan 2 *Cluster*

- C1 : Kab. Sidoarjo, Kab. Gresik, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, Kota Batu.
- C2 : Kab. Pacitan, Kab. Ponorogo, Kab.

Trenggalek, Kab. Tulungagung, Kab. Blitar, Kab. Kediri, Kab. Malang, Kab. Lumajang, Kab. Jember, Kab. Banyuwangi, Kab. Bondowoso, Kab. Situbondo, Kab. Probolinggo, Kab. Pasuruan, Kab. Mojokerto, Kab. Jombang, Kab. Nganjuk, Kab. Madiun, Kab. Magetan, Kabupaten Ngawi, Kab. Bojonegoro, Kabupaten Tuban, Kab. Lamongan, Kab. Bangkalan, Kab. Sampang, Kab. Pamekasan, Kab. Sumenep.

Jika mengacu pada ketentuan Badan Pusat Statistik Nasional, tingkat IP-TIK dikelompokkan ke dalam 4 kategori yakni tinggi, sedang, rendah, dan sangat rendah. Untuk itu, dalam penelitian ini juga dilakukan pengelompokan dengan 4 kategori. Skenario datanya menggunakan tahun 2021, tahun 2022 secara terpisah dan menggabungkan keduanya. Hasil dari pengelompokan tersebut adalah Kabupaten Pasuruan dan Kabupaten Ngawi berada di *cluster* 2 (sedang) pada tahun 2021, sedangkan pada pengelompokan tahun 2022 dan gabungan dari 2021 dan 2022 berada di *cluster* 3 (rendah). Untuk kota dan kabupaten lainnya berada di tingkat yang sama pada ketiga skenario data. Gambar 3 menampilkan peta penelompokan data gabungan tahun. Tanda bulat biru untuk *cluster* 1, X merah untuk *cluster* 2, segitiga hijau *cluster* 3, dan + orange *cluster* 4.

Dengan menggunakan pengelompokan yang telah ditetapkan oleh Badan Pusat Statistik Nasional, Penulis dapat melihat bagaimana kabupaten-kabupaten tertentu mengalami pergeseran dalam klasifikasi IP-TIK mereka dari tahun ke tahun. Analisis visual yang disajikan melalui Gambar 3 memberikan pemahaman tentang pola-pola yang dapat menjadi dasar bagi pengambilan keputusan yang lebih baik dalam perencanaan dan pengembangan teknologi informasi dan komunikasi di tingkat daerah.



Gambar 3. Hasil Pengelompokan Tingkat IP-TIK dengan 4 *Cluster* Untuk Data Tahun 2021 dan 2022

- C1 : Kab. Sidoarjo, Kab. Gresik, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, Kota Batu
- C2 : Kab. Ponorogo, Kab. Trenggalek, Kab. Tulungagung, Kab. Blitar, Kab. Kediri, Kab. Malang, Kab. Banyuwangi, Kab. Mojokerto, Kab. Jombang, Kab. Nganjuk, Kab. Madiun, Kab. Magetan, Kab. Bojonegoro, Kabupaten Tuban, Kab. Lamongan
- C3 : Kab. Jember, Kab. Bondowoso, Kab. Situbondo, Kab. Pasuruan, Kabupaten Ngawi, Kab. Sumenep
- C4 : Kab. Pacitan, Kab. Lumajang, Kab. Probolinggo, Kab. Bangkalan, Kab. Sampang, Kab. Pamekasan

Rata-rata pengujian Silhouette Coefficient pada masing-masing *cluster* berdasarkan urutan tertinggi adalah 0.661 *cluster* 1, 0.658 *cluster* 2, 0.611 *cluster* 4, dan 0.587 *cluster* 3. Nilai Silhouette Coefficient menunjukkan tingkat homogenitas dalam 1 kelompok/*cluster*.

Tabel 2. Perbandingan Min-Max di Masing-masing *Cluster*

Variabel	Tahun	<i>Cluster</i>	<i>Min</i>	<i>Max</i>
RT Komputer	2021	C1	30,38	53,6
		C2	16,27	27,17
		C3	14,13	20,13
		C4	1,59	17,99
	2022	C1	24,47	41,06
		C2	11,83	18,76

RT Telpon Seluler	2021	C3	8,1	11
		C4	2,26	11,75
		C1	91,95	97,16
		C2	82,72	90,41
		C3	77,94	83,59
		C4	85,17	92,73
		C1	91,85	95,69
		C2	83,13	89,73
	2022	C3	77,91	80,8
		C4	84,11	89,76
		C1	86,1	94,2
		C2	75,23	85,92
		C3	59,96	77,4
		C4	69,41	74,75
		C1	86,96	93,03
		C2	75,58	84,26
RT Akses Internet	2021	C3	58,15	74,52
		C4	66,39	75,93
	2022	C1	67,26	82,96
		C2	53,84	65,02
5+ th Akses Internet	2021	C3	40,44	56,24
		C4	38	46,88
	2022	C1	70,98	84,77
		C2	57,25	74,59
IPM	2021	C3	45,17	62,68
		C4	43,91	55,69
	2022	C1	73,66	82,31
		C2	68,91	74,15
	2021	C3	66,59	71,04
		C4	62,8	68,57
	2022	C1	74,56	82,74
		C2	69,67	74,89
	2022	C3	67,31	71,75
		C4	63,39	69,37

Berdasarkan Tabel 2 terlihat perbedaan yang sangat jauh pada nilai minimum dan maksimum *cluster* 4 untuk variabel rumah tangga yang memiliki komputer. Sedangkan pada *cluster* lain, perbedaan min-max untuk semua variabel relatif kecil.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa:

- Kelompok terbanyak pada pengelompokan 4 kategori berada pada *cluster* 2 yang menandakan tingkat IP-TIK di provinsi Jawa Timur rata-rata

sedang.

- Kelompok terbanyak pada pengelompokan 2 kategori berada di *cluster* 2 (rendah).
- Hasil Silhouette Coefficient tertinggi pada pengelompokan 4 kategori berada di *cluster* 1 yang menandakan kemiripan/homogenitas kelompok tersebut paling baik di antara yang lain
- Hasil Silhouette Coefficient tertinggi pada pengelompokan 2 kategori berada di *cluster* 1 (tinggi).
- Ketimpangan kelompok terjadi pada *cluster* 4 untuk variabel rumah tangga yang memiliki komputer.

- f) Kota atau kabupaten yang masuk kategori rendah (*cluster* 4) antara lain Kab. Pacitan, Kab. Lumajang, Kab. Probolinggo, Kab. Bangkalan, Kab. Sampang, dan Kab. Pamekasan.

5. Ucapan Terima kasih

Tim Penulis mengucapkan terima kasih kepada Lembaga Penelitian dan Pengabdian Kepada Masyarakat (LPPM) atas pendanaan yang diberikan pada penelitian ini untuk skema Penelitian Pemula tahun 2023.

6. Daftar Pustaka

- [1] Indeks Pembangunan Teknologi Informasi dan Komunikasi 2021, 2022. Jakarta: Badan Pusat Statistik Indonesia.
- [2] Al-Mursyid, A. R. (2020). Pengaruh pembangunan teknologi informasi dan komunikasi terhadap pembangunan kawasan timur Indonesia. *Jurnal Ekonomi Pembangunan STIE Muhammadiyah Palopo*, 5(2), 53-66. DOI: <http://dx.doi.org/10.35906/jep01.v5i2.372>.
- [3] Saputra, R. A., Nuryadin, D., & Winarti, A. S. (2022). PENGARUH TEKNOLOGI INFORMASI DAN KOMUNIKASI (TIK) DAN KEBEBASAN EKONOMI TERHADAP PEMBANGUNAN MANUSIA DI KAWASAN NEGARA ASEAN TAHUN 2014-2019. *SIBATIK JOURNAL: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, Dan Pendidikan*, 1(10), 2107-2120.
- [4] Jatipaningrum, M. T., Azhari, S. E., & Suryowati, K. (2022). Pengelompokan kabupaten dan kota di provinsi jawa timur berdasarkan tingkat kesejahteraan dengan metode k-means dan density-based spatial clustering of applications with noise. *Jurnal Derivat: Jurnal Matematika Dan Pendidikan Matematika*, 9(1), 70-81. DOI: <https://doi.org/10.31316/j.derivat.v9i1.2832>.
- [5] Kamila, I., Khairunnisa, U., & Mustakim, M. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, 5(1), 119-125. DOI: <http://dx.doi.org/10.24014/rmsi.v5i1.7381>.
- [6] Muningsih, E., & Kiswati, S. (2018). Sistem aplikasi berbasis optimasi metode elbow untuk penentuan clustering pelanggan. *Joutica: Journal of Informatic Unisla*, 3(1), 117-124. DOI: <https://doi.org/10.30736/jti.v3i1.196>.
- [7] Hartanti, N. T. (2020). Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional. *J. Nas. Teknol. dan Sist. Inf*, 6(2), 82-89.
- [8] Dewi, D. A. I. C., & Pramita, D. A. K. (2019). Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Matrix: Jurnal Manajemen Teknologi dan Informatika*, 9(3), 102-109. DOI: <http://dx.doi.org/10.31940/matrix.v9i3.1662>.
- [9] Anggara, M., Sujaini, H., & Nasution, H. (2016). Pemilihan distance measure Pada K-Means clustering Untuk Pengelompokkan member di Alvaro fitness. *JUSTIN (Jurnal Sistem Dan Teknologi Informasi)*, 4(1), 186-191.
- [10] Statistik, I. B. P. (1994). Statistik kesejahteraan rakyat. (*No Title*).
- [11] Nugraha, A., Nurdiawan, O., & Dwilestari, G. (2022). Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 849-855. DOI: <https://doi.org/10.36040/jati.v6i2.5755>.