

Penerapan Data Mining Clustering terhadap Perekonomian di Kelurahan Sawah Lebar Menggunakan Algoritma K-Means

Willi Novrian¹, Soni Ayi Purnama², Niska Ramadani³

^{1,2,3}Program Studi Sistem Informasi, Universitas Bengkulu, Bengkulu, Indonesia
willinovrian@unib.ac.id

Info Artikel

Sejarah artikel:

Diterima Desember 2024

Direvisi Desember 2024

Disetujui Desember 2024

Diterbitkan Desember 2024

ABSTRACT

This study aims to analyze the economic conditions of the community in Sawah Lebar Village, Ratu Agung District, Bengkulu City, Bengkulu Province, using the data mining clustering method based on the K-Means algorithm. The K-Means algorithm is applied to group economic data of the community based on several variables, such as Employment Status, Home Ownership, Dependents, Monthly Income, Monthly Expenditure and Age of Head of Family. Through this clustering process, several groups (clusters) are produced that identify different economic patterns in the region, namely Cluster_0 (C0), Cluster_1 (C1) and Cluster_2 (C2). In order for the results of this study to be accurate, the data processing in this study uses the Rapidminer Studio Application. From the results of the data processing that has been carried out, it was found that for C0 there are 17 data, C1 there are 16 Data and C2 there are 17 Data. Then, for the recommendation of the community that is recommended to receive government assistance, it is the community that is in the lowest economic class, namely Cluster_1 (C1), Cluster_2 (C2) is the community with the highest economy, while the community in Cluster_0 (C0) is the middle economy.

Keywords : Clustering; Data Mining; K-Means.

ABSTRAK

Penelitian ini bertujuan untuk menganalisis kondisi perekonomian masyarakat di Kelurahan Sawah Lebar, Kecamatan Ratu Agung, Kota Bengkulu, Provinsi Bengkulu, dengan menggunakan metode data *mining clustering* berbasis algoritma *K-Means*. Kelurahan Sawah Lebar memerlukan suatu informasi yang akurat sebagai acuan untuk rekomendasi bantuan pemerintah pusat, provinsi dan kota. Algoritma *K-Means* diterapkan untuk mengelompokkan data ekonomi masyarakat berdasarkan beberapa variabel, seperti Status Pekerjaan, Kepemilikan Rumah, Tanggungan, Pendapatan Bulanan, Pengeluaran Bulanan dan Usia Kepala Keluarga. Melalui proses *clustering* ini, dihasilkan beberapa kelompok (*cluster*) yang mengidentifikasi pola-pola perekonomian yang berbeda di wilayah tersebut yaitu *Cluster_0* (C0), *Cluster_1* (C1) dan *Cluster_2* (C2). Agar hasil penelitian ini akurat maka dalam pengolahan data pada penelitian ini menggunakan Aplikasi Rapidminer Studio. Dari hasil pengolahan data yang telah dilakukan didapatkan bahwa untuk C0 terdapat 17 data, C1 terdapat 16 Data dan C2 terdapat 17 Data. Kemudian untuk rekomendasi Masyarakat yang di rekomendasikan untuk mendapatkan bantuan pemerintah adalah Masyarakat yang berada pada perekonomian paling rendah yaitu *Cluster_1* (C1), *Cluster_2* (C2) adalah Masyarakat yang perekonomian paling tinggi sedangkan Masyarakat dalam *Cluster_0* (C0) adalah perekonomian sedang.

Kata Kunci : Clustering; Data Mining Data Mining; K-Means.

PENDAHULUAN

Perekonomian suatu wilayah sering kali menjadi indikator penting dalam menentukan kualitas hidup masyarakat yang tinggal di dalamnya [1]. Setiap kelurahan memiliki karakteristik dan potensi ekonomi yang berbeda-beda, yang dipengaruhi oleh faktor-faktor seperti tingkat pendidikan, akses terhadap infrastruktur, jenis pekerjaan mayoritas penduduk, serta kondisi demografi lainnya [2]. Untuk memahami dan menganalisis perekonomian suatu daerah, diperlukan pendekatan yang komprehensif agar dapat mengambil keputusan yang tepat, baik oleh pemerintah daerah maupun pihak terkait lainnya [3].

Kelurahan Sawah Lebar merupakan salah satu wilayah yang terletak di kota Bengkulu dengan dinamika ekonomi yang kompleks. Seperti halnya wilayah perkotaan pada umumnya, Sawah Lebar memiliki keragaman dalam hal pendapatan penduduk, sektor usaha, tingkat pengangguran, dan berbagai faktor lainnya yang berpengaruh terhadap kesejahteraan ekonomi. Oleh karena itu, diperlukan analisis yang mendalam untuk memahami bagaimana perekonomian di wilayah ini terbentuk dan berkembang.

Salah satu metode yang efektif untuk menganalisis data dalam jumlah besar dan memetakan pola ekonomi suatu wilayah adalah data *mining* [4]. Data *mining* dapat digunakan untuk menemukan pola-pola tersembunyi dalam data yang dapat membantu dalam pengambilan keputusan strategis [5], [6]. Di antara berbagai teknik data *mining*, algoritma *K-Means Clustering* merupakan salah satu metode yang populer untuk melakukan analisis klastering [7], yaitu membagi data kedalam kelompok-kelompok yang homogen berdasarkan karakteristik tertentu [8], [9]. Dengan *K-Means*, data ekonomi di Kelurahan Sawah Lebar dapat diidentifikasi dan dikelompokkan berdasarkan pola yang serupa, sehingga dapat memberikan gambaran yang lebih jelas mengenai perbedaan atau kesamaan antar kelompok ekonomi di wilayah tersebut

Penerapan *K-Means* dalam perekonomian Kelurahan Sawah Lebar diharapkan dapat membantu dalam mengklaster penduduk berdasarkan indikator ekonomi seperti pendapatan, jenis pekerjaan, dan pendidikan, pemerintah dapat lebih mudah merencanakan kebijakan yang sesuai dengan kebutuhan masing-masing kelompok. Selain itu analisis ini juga dapat membantu mengidentifikasi wilayah atau kelompok yang memerlukan intervensi ekonomi lebih lanjut, sehingga program-program yang dilaksanakan dapat lebih terfokus.

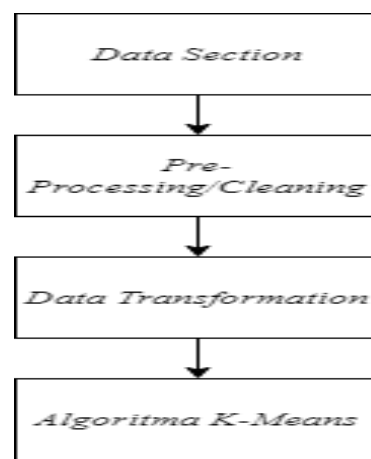
Beberapa penelitian terdahulu yang relevan sebagai landasan dalam penelitian ini yaitu algoritma *K-Means* untuk mengelompokkan data perekonomian di tingkat kecamatan di Kabupaten X. Mereka menganalisis berbagai variabel seperti tingkat pengangguran, pendapatan per kapita, dan sektor pekerjaan [10]. Hasil penelitian menunjukkan bahwa *K-Means* dapat secara efektif membagi data perekonomian menjadi beberapa klaster yang memiliki karakteristik ekonomi serupa, sehingga membantu pemerintah dalam merancang kebijakan sesuai dengan kondisi masing-masing klaster. Selanjutnya variabel yang digunakan meliputi pendapatan, tingkat pendidikan, dan akses terhadap layanan kesehatan. Penelitian tersebut menemukan bahwa teknik klastering ini dapat membantu mengidentifikasi wilayah dengan tingkat ekonomi rendah yang memerlukan

intervensi pemerintah [11]. Penelitian ini memberikan bukti bahwa metode K-Means dapat diadaptasi untuk berbagai wilayah dengan karakteristik ekonomi yang beragam. Selanjutnya adalah *K-Means* dalam mengelompokkan wilayah berdasarkan tingkat kemiskinan dan pengangguran di sebuah kabupaten di Indonesia. Hasil penelitian menunjukkan bahwa terdapat kluster-kluster yang memiliki ciri khas tertentu, seperti tingkat kemiskinan tinggi namun pengangguran rendah, atau sebaliknya. Hasil klustering ini membantu pemerintah daerah dalam mengidentifikasi prioritas intervensi ekonomi [12].

Penelitian-penelitian terdahulu menunjukkan bahwa algoritma *K-Means* secara luas telah digunakan dalam berbagai konteks ekonomi, baik di tingkat lokal maupun nasional. Dari klusterisasi berdasarkan pendapatan, sektor pekerjaan, hingga tingkat pendidikan dan pengangguran, metode ini terbukti dapat memberikan wawasan yang lebih mendalam terhadap kondisi ekonomi suatu wilayah. Oleh karena itu, penerapan K-Means pada perekonomian Kelurahan Sawah Lebar sangat relevan, dengan potensi untuk menghasilkan informasi yang bermanfaat dalam pengambilan keputusan terkait kebijakan ekonomi di tingkat kelurahan.

METODE

Penelitian ini akan menggunakan model standarisasi data *mining* yaitu *Knowledge Discovery Database* (KDD) untuk mendapatkan sebuah visualisasi pola pola yang terdapat dalam basis data. Adapun langkah-langkah yang ada didalamnya yaitu seperti gambar dibawah ini:



Gambar 1. Tahapan Penelitian

Data Section

Dalam langkah ini, data set yang akan digunakan untuk pengolahan selanjutnya berasal dari data primer yang dikumpulkan dari responden yaitu Masyarakat kelurahan sawah lebar, melalui metode kuesioner dan dipilih sesuai dengan persyaratan penelitian. Dalam penelitian ini data ini di dapat dari Masyarakat kelurahan sawah lebar Bengkulu.

Pre-Processing/Cleaning

Pada tahap ini, data dalam basis data yang sudah dikelola pada *data section* dibersihkan dengan cara-cara seperti menghapus data ganda, memeriksa data yang tidak konsisten, memperbaiki kesalahan, dan menggabungkan atau memperkaya data dengan data yang relevan dan diperlukan.

Data Transformation

Pada langkah ini, proses penggabungan atau perubahan data ke dalam format yang dibutuhkan dilakukan. Ini membuat proses data *mining* lebih mudah.

Algoritma K-Means

Pada langkah ini, peneliti mengolah dan mengidentifikasi data dari kumpulan yang telah ditentukan sebelumnya dengan menggunakan algoritma *K-Means* dengan memakai Aplikasi *Rapidminer Studio*. Berikut ini adalah tahapan tahapan dari algoritma *K-Means Clustering*:

1. Menentukan jumlah *clutser*
2. Menentukan *centroid* awal
3. Menghitung jarak antar *centroid*
4. Menentukan letak *cluster*
5. Melakukan nilai *centroid* baru
6. Ulangi Langkah point c sampai e sampai tidak ada lagi *cluster* yang berpindah

Evaluation

Fase ini mencakup memeriksa informasi yang diperoleh dari proses sebelumnya untuk membandingkannya dengan fakta dan hipotesa yang telah ada. Selain itu, hasil penelitian juga diuji keakuratannya.

HASIL DAN PEMBAHASAN**Data Section**

Berdasarkan penelitian yang telah dilakukan pada Masyarakat Kelurahan Sawah Lebar Kota Bengkulu maka, indikator atau variable yang dijadikan dalam penentuan perekonomian Masyarakat untuk menerima berbagai bantuan dari pemerintah yaitu seperti tabel dibawah ini:

Tabel 1. Variabel

Nomor	Variabel
1	Status Pekerjaan
2	Kepemilikan rumah
3	Tanggungan
4	Pendapatan Bulanan
5	Pengeluaran Bulanan
6	Usia Kepala Keluarga

Pre-Processing/Cleaning

Pada tahap ini, proses *cleaning* data untuk membersihkan data ganda, memperbaiki data yang tidak konsisten sehingga data yang didapatkan akurat. Terlihat pada tabel 2.

Data Transformation

Menyiapkan data set atau data yang akan nantinya diolah dalam penelitian ini, data set yang dibuat disini telah dirubah terlebih dahulu kedalam data yang bisa di mengerti oleh algoritma *K-Means Clustering*. Berikut ini adalah bentuk dataset dari penelitian ini.

Tabel 2. Data Set

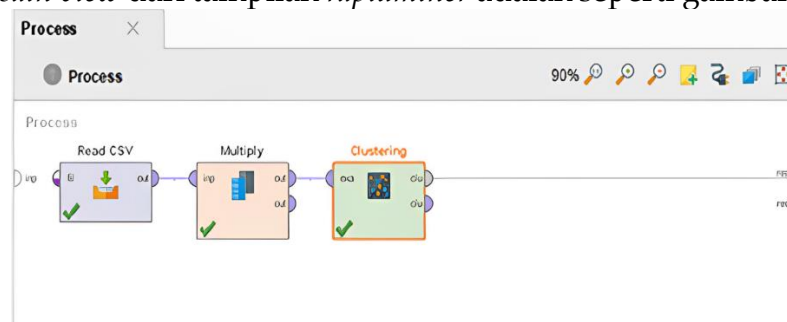
Inisial Nama	Status Pekerjaan	Kepemilikan Rumah	Tanggungan	Pendapatan Bulanan	Pengeluaran Bulanan	Usia Kepala Keluarga
AMS	2	2	6	2192743	798863	63
NS	1	0	4	4804572	667654	63
JK	2	0	6	2734489	2976188	24
AN	2	0	2	2070006	2927241	58
HM	1	1	2	1636074	3595990	23
SH	0	0	4	4544887	3888633	25
RH	0	0	8	1739911	2632514	64
HK	1	0	7	4972471	1792785	51
IL	0	0	9	2638242	3999608	49
MA	1	2	8	2266891	1097079	66
Dst..						

Pada penelitian ini, peneliti menentukan jumlah cluster yang akan di olah dalam data tersebut, Adapun jumlah cluster pada penelitian ini adalah seperti berikut ini:

1. Cluster C0
2. Cluster C1
3. Cluster C2

Algoritma K-Means

Pada tahapan ini setelah semua data sudah selesai, peneliti melakukan pengolahan data menggunakan aplikasi *rapidminer*. Data di siapkan menggunakan microsoft Excel. Selanjutnya pada pengolahan ini menggunakan beberapa fitur yang terdapat pada *rapidminer* yaitu *Read CSV*, *Multiply* dan *K-means*. Adapaun bentuk dari *desain view* dari tampilan *rapidminer* adalah seperti gambar dibawah ini.



Gambar 2. Desain View Rapidminer

Setelah data sudah di input dan memakai fitur yang terdapat pada gambar diatas, maka di lakukan iterasi sebanyak 10 kali dan membagi cluster ke 3 kelompok, sehingga didapatkan hasil dari pengolahan data Masyarakat Kelurahan Sawah Lebar Kota Bengkulu dalam penentuan perekonomian Masyarakat adalah seperti gambar dibawah ini.

```

Cluster Model

Cluster 0: 17 items
Cluster 1: 16 items
Cluster 2: 17 items
Total number of items: 50

```

Gambar 3. Hasil Klaster

Berdasarkan hasil pengolahan data diatas didapatkan hasil dari 3 cluster yaitu cluster_0 sebanyak 17 item data, cluster_1 sebanyak 16 item data dan cluster_2 terdapat 17 item data. Masing masing data dalam cluster tersebut dapat dilihat pada penjelasan berikut ini:

1. Cluster_0

Pada cluster_0 terdapat sebanyak 17 item data penduduk. Adapun data lengkapnya bisa dilihat pada gambar dibawah ini:

Row No.	Inisial Nama	cluster	Status Peke...	Kepemilikan...	Tanggungan	Pendapatan ...	Pengeluara...	Usia Kepala ...
1	JK	cluster_0	2	0	6	2734489	2976188	24
2	HM	cluster_0	1	1	2	1636074	3595990	23
3	SH	cluster_0	0	0	4	4544887	3888633	25
4	IL	cluster_0	0	0	9	2638242	3999608	49
5	LO	cluster_0	0	1	9	2753882	2999015	38
6	OP	cluster_0	2	0	7	4030409	3714527	36
7	PA	cluster_0	2	0	8	2454354	2803771	45
8	PL	cluster_0	0	1	3	2916182	3477825	42
9	HA	cluster_0	0	2	5	3086644	3538576	53
10	MM	cluster_0	1	2	7	3070406	3488649	51
11	NN	cluster_0	1	0	9	3916664	3102238	67
12	DR	cluster_0	1	0	5	3503758	3516041	48
13	UD	cluster_0	1	1	8	3169995	2581230	31
14	OD	cluster_0	2	1	6	4950812	3803157	45
15	ZA	cluster_0	0	2	3	3057489	3559360	41
16	ZU	cluster_0	2	0	5	2273415	3394791	48
17	ZI	cluster_0	1	2	3	2805413	2285588	33

Gambar 4. Cluster_0

2. Cluster_1

Pada cluster_1 terdapat sebanyak 16 item data penduduk. Adapun data lengkapnya bisa dilihat pada gambar dibawah ini.

Row No.	Inisial Nama	cluster	Status Peka...	Kepemilikan...	Tanggungan	Pendapatan ...	Pengeluara...	Usia Kepala ...
1	NS	cluster_1	1	0	4	4804572	667654	63
2	HK	cluster_1	1	0	7	4872471	1792785	51
3	HJ	cluster_1	1	2	5	3844769	964076	54
4	IJ	cluster_1	1	1	8	4760029	1832875	47
5	EP	cluster_1	1	1	1	3775709	748982	45
6	SW	cluster_1	0	2	1	3974675	1296161	28
7	SD	cluster_1	0	2	8	3152991	966326	31
8	QA	cluster_1	2	1	3	2928388	949150	20
9	KL	cluster_1	1	1	1	3163046	418834	66
10	PP	cluster_1	2	0	7	3885659	1999005	44
11	II	cluster_1	2	2	9	3885357	2054236	59
12	AA	cluster_1	2	2	8	3285513	584821	64
13	DD	cluster_1	1	1	7	3245683	1379600	24
14	DA	cluster_1	0	2	8	3897723	1325715	41
15	DI	cluster_1	1	2	3	456518	1738533	22
16	LZ	cluster_1	0	1	1	4655299	2358852	56

Gambar 5. Cluster_1

3. Cluster_2

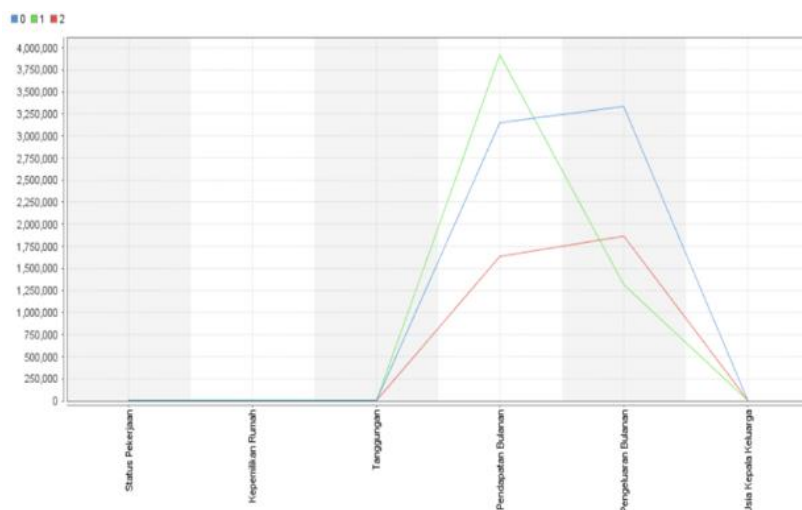
Pada cluster_2 terdapat sebanyak 17 item data penduduk. Adapun data lengkapnya bisa dilihat pada gambar dibawah ini.

Row No.	Inisial Nama	cluster	Status Peka...	Kepemilikan...	Tanggungan	Pendapatan ...	Pengeluara...	Usia Kepala ...
1	AMS	cluster_2	2	2	6	2182743	798863	63
2	AN	cluster_2	2	0	2	2070006	2927241	58
3	RH	cluster_2	0	0	8	1739911	2632514	64
4	MA	cluster_2	1	2	8	2266891	1097079	66
5	SL	cluster_2	2	0	2	1291743	2617115	59
6	JA	cluster_2	1	1	5	603355	2877913	35
7	IS	cluster_2	2	2	8	1762752	1597286	32
8	IN	cluster_2	0	0	9	1896025	2618136	69
9	NA	cluster_2	0	2	9	1807371	2205324	61
10	JK	cluster_2	0	2	1	1417040	2661664	49
11	TU	cluster_2	0	0	9	2000942	1588125	38
12	ET	cluster_2	1	0	8	584654	1472622	56
13	AE	cluster_2	1	2	3	1028178	3675957	20
14	AL	cluster_2	2	0	2	1965689	355609	20
15	SM	cluster_2	0	0	1	568148	527897	35
16	SS	cluster_2	2	0	7	2325665	1274165	58
17	PD	cluster_2	2	0	3	2370928	761243	35

Gambar 6. Cluster_2

Hasil

Berdasarkan pengolahan data yang telah dilakukan menggunakan algoritma k-means clustering dengan menggunakan aplikasi rapidminer maka didapatkan hasil dari pengeolahan data seperti Gambar dibawah ini.



Gambar 7. Grafik Hasil Penelitian

Berdasarkan Grafik tersebut dapat di lihat bawah kelompok atau cluster Masyarakat yang berekonomi paling rendah dan di rekomendasikan untuk mendapatkan bantuan dari pemerintah adalah Cluster_1. Kemudian untuk perekonomian Masyarakat sedang ditunjukan pada Cluster_0 dan Masyarakat perekonomian tinggi adalah pada Cluster_2.

PENUTUP

Berdasarkan hasil pengolahan data yang telah dilakukan menggunakan algoritma K-Means Clustering dalam mengetahui kelompok perekonomian Masyarakat Kelurahan Sawah Lebar Kota Bengkulu dengan Aplikasi Rapidminer di dapatkan Kesimpulan yaitu Pengelompokan data atau cluster dibagi menjadi 3 kelompok yaitu masyarakat perekonomian rendah adalah Cluster_1 (C1). Kemudian untuk perekonomian Masyarakat sedang ditunjukan pada Cluster_0 (C0) dan Masyarakat perekonomian tinggi adalah Cluster_2 (C2). Kemudian dari hasil pengolahan data yang telah dilakukan didapatkan bahwa untuk C0 terdapat 17 data, C1 terdapat 16 Data dan C2 terdapat 17 Data. Dari hasil yang di peroleh maka untuk rekomendasi Masyarakat yang di rekomendasikan untuk mendapatkan bantuan pemerintah adalah Masyarakat yang berada pada Cluster_1 (C1).

DAFTAR PUSTAKA

- [1] A. Setiawan and S. Huda, "Analisis Faktor yang Mempengaruhi Pertumbuhan Ekonomi Di Kabupaten Mojokerto," *J. Syntax Admiration*, vol. 2, no. 8, pp. 1384-1394, 2021, doi: 10.46799/jsa.v2i8.295.
- [2] P. Yuniarti, W. Wianti, and N. E. Nurgaheni, "Analisis Faktor-faktor yang Mempengaruhi Tingkat Pertumbuhan Ekonomi di Indonesia," *SERAMBI J. Ekon. Manaj. dan Bisnis Islam*, vol. 2, no. 3, pp. 169-176, 2020, doi: 10.36407/serambi.v2i3.207.
- [3] J. O. F. Management, "Simanungkalit / JOURNAL OF MANAGEMENT (SME's) Vol. 13, No.3, 2020, p327-340," vol. 13, no. 3, pp. 327-340, 2020.
- [4] C. J. Siti Mariam, Fitri Handayani, "Penerapan Algoritma Clustering K-Means

- Untuk Menentukan Prioritas Penerima Bantuan Rumah Akibat Bencana Alam," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 2, pp. 231–240, 2023.
- [5] F. Febriansyah and S. Muntari, "Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 8, no. 1, pp. 66–77, 2023, doi: 10.14421/jiska.2023.8.1.66-77.
- [6] ... Preddy, P. Marpaung, I. Pebrian, and W. Putri, "Penerapan Data Mining Untuk Pengelompokan Kepadatan Penduduk Kabupaten Deli Serdang Menggunakan Algoritma K-Means," *J. Ilmu Komput. dan Sist. Inf.*, vol. 6, no. 2, pp. 64–70, 2023.
- [7] I. Iin, R. Fadila, A. Rizki Rinaldi, and F. Fathurrohman, "Penerapan Data Mining Dalam Mengelompokan Jumlah Umkm Berdasarkan Kabupaten Kota Menggunakan K-Means Clustering," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 1446–1450, 2024, doi: 10.36040/jati.v8i2.8427.
- [8] G. Sonia and R. A. Putri, "Penerapan Metode K-Means Clustering Untuk Mengelompokkan Data Kelayakan Penerima Bantuan Renovasi Rumah," *Build. Informatics, Technol. Sci.*, vol. 5, no. 2, pp. 442–455, 2023, doi: 10.47065/bits.v5i2.4298.
- [9] D. Endrawati, S. Nujum, and A. Selong, "Pengaruh Pertumbuhan Ekonomi, Rasio Gini dan Indeks Pembangunan Manusia Terhadap Tingkat Kemiskinan Indonesia 2017-2022," *J. Pendidik. Tambusai*, vol. 7, no. 3, pp. 20144–20151, 2023.
- [10] Y. R. Sari, A. Sudewa, D. A. Lestari, and T. I. Jaya, "Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, p. 192, 2020, doi: 10.24114/cess.v5i2.18519.
- [11] R. Maulidiah, M. Muchtar, N. A. Fitri, I. Asriani, and M. P. Yasmine, "Pengelompokan Data Pertumbuhan dan Kontribusi Ekonomi Indonesia Menurut Provinsi Menggunakan Metode K-Means Clustering," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 6, no. 2, p. 436, 2023, doi: 10.53513/jsk.v6i2.7769.
- [12] N. Nurjanah, N. Suarna, W. Prihartono, T. Informatika, R. P. Lunak, and G. Tasikmalaya, "Implementasi K-Means Clustering Untuk Mengelompokan," vol. 8, no. 2, pp. 2462–2468, 2024.