



PENERAPAN K-MEANS CLUSTERING UNTUK PENGELOMPOKAN WILAYAH BERDASARKAN TINGKAT KEMISKINAN DI INDONESIA

Sagita Maesarah^{a*}, Zaehol Fatah^b,

^a Sains dan teknologi / Sistem Informasi, sagitamaisarah@gmail.com, Universitas Ibrahimy, Situbondo Jawa Timur

^b Sains dan Teknologi, / Sistem Informasi, zaeholfatah@gmail.com, Universitas Ibrahimy, Situbondo Jawa Timur

* Korespondensi

ABSTRACT

Poverty is one of the problems that hinder national and regional growth. Poverty is the inability to meet the minimum standards of basic needs including food and non-food needs. Poor people are people who are below a limit or called the poverty line. The resources used are sourced from www.kaggle.com. In the process of data processing with the k-means Clustering method. The K-means Clustering method is a method of grouping existing data into several groups where the data in one group has the same characteristics as each other and has different characteristics from the data in the group. The results of the study show that regions in Indonesia can be grouped into several clusters with different poverty level characteristics. These clusters reveal specific patterns, such as the concentration of areas with high poverty in certain areas and the factors that contribute to these conditions. With this approach, the government and policy makers can identify priority areas and design more effective programs to reduce poverty levels..

Keywords: Poverty, K-means Clustering area grouping

Abstrak

Kemiskinan merupakan salah satu masalah yang menghambat pertumbuhan nasional maupun regional. Kemiskinan adalah ketidakmampuan untuk memenuhi standar minimum kebutuhan dasar yang meliputi kebutuhan makanan maupun non-makanan. Penduduk miskin adalah penduduk yang berada dibawah suatu batas atau disebut sebagai garis kemiskinan. Daya yang digunakan bersumber dari www.kaggle.com. Dalam proses pengolahan data dengan metode k-means Clustering. Metode K-means Clustering merupakan suatu metode pengelompokkan data yang ada ke dalam beberapa kelompok dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan memiliki karakteristik yang berbeda dengan data yang ada dalam kelompok. Hasil penelitian menunjukkan bahwa wilayah di Indonesia dapat di kelompokkan kedalam beberapa klaster dengan karakteristik tingkat kemiskinan yang berbeda. Klaster ini mengungkapkan pola-pola spesifik, seperti konsentrasi wilayah dengan kemiskinan tinggi di daerah tertentu sert faktor-faktor yang berkontribusi terhadap kondisi tersebut. Dengan pendekatan ini, pemerintahan dan pemangku kebijakan dapat mengidentifikasi wilayah prioritas dan merancang program yang lebih efektif untuk mengurangi tingkat kemiskinan.

Kata Kunci: emiskinan, pengelompokkan wilayah K-means Clustering

1. PENDAHULUAN

Kemiskinan merupakan persoalan struktur dan kultural yang Multi dimensi, mencakup didalamnya masalah politik, sosial, ekonomi, asset dan sumberdaya alam, psigologi dan lain-lain.. secara umum masyarakat miskin adalah suatu kondisi masyarakat yang berada dalam situasi rentan, tak berdaya, terisolasi dan tidak mampu untuk menyampaikan aspirasinya. Masalah ini masih banyak dialami oleh kaum perempuan dan menyebabkan mereka tidak mampu untuk memenuhi kebutuhan minimal kehidupannya secara layak (prof. Dr. Ir. Keppi Sukesu, 2015).

Ada tiga ciri yang menonjol dari kemiskinan di Indonesia. Pertama, banyak rumah tangga yang berada di sekitar garis kemiskinan nasional, yang setara dengan US\$1,55 per hari, sehingga banyak penduduk yang meskipun tergolong tidak miskin tetapi rentan terhadap kemiskinan. Kedua, ukuran kemiskinan didasarkan pada pendapatan, sehingga tidak menggambarkan batas kemiskinan yang sebenarnya. Banyak orang yang mungkin tidak tergolong “miskin dari segi pendapatan” dapat dikategorikan sebagai miskin atas dasar kurangnya akses terhadap pelayanan dasar serta rendahnya indikator-indikator pembangunan manusia. Ketiga, mengingat sangat luas dan beragamnya wilayah Indonesia, perbedaan antar daerah merupakan ciri mendasar dari kemiskinan di Indonesia (Dkk, 2015)

Indonesia adalah Negara yang tergolong masih berkembang, dimana kemiskinan merupakan masalah yang cukup serius dan masih menjadi pusat perhatian. Maka dari itu, menjadi adanya upaya untuk memecahkan masalah tersebut, dimana upaya pengentasan kemiskinan menjadi salah satu aspek penting dalam mencapai kesejahteraan rakyat. Pada tingkat nasional maupun regional, kemiskinan telah menjadi masalah yang rumit dan berkepanjangan, sehingga masalah tersebut menjadi prioritas utama pembangunan yang memerlukan strategi penyelesaian yang tepat dan berkelanjutan untuk mengatasinya, maka dari itu cara penanganan kemiskinan pun membutuhkan analisis yang tepat untuk melibatkan semua komponen permasalahan (Mayasari & Nugraha, 2023)

Clustering adalah metode yang membagi data ke dalam grup-grup yang mempunyai obyek dengan karakteristik yang sama. Menyatakan clustering adalah mengelompokkan item data ke dalam sejumlah kecil grup sedemikian rupa sehingga masing-masing grup mempunyai sesuatu persamaan yang esensial. Menjelaskan bahwa clustering adalah proses membagi data tidak berlabel menjadi kelompok-kelompok data yang memiliki kemiripan. (Bahauddin et al., 2021) K-Means Clustering merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster. Tujuan algoritma K-Means Clustering yaitu untuk membagi data menjadi beberapa data menjadi beberapa kelompok (Arkham & Swanjaya, 2020)

2. METODOLOGI PENELITIAN

Penelitian Penerapan K-Means Clustering untuk Pengelompokan Wilayah Berdasarkan Tingkat Kemiskinan di Indonesia memiliki beberapa langkah yang mengacu kepada metode K-Means Clustering sebagai metode untuk menyelesaikan permasalahan dalam penelitian ini. Metode K-Means Clustering akan melakukan pengelompokan dari hasil data yang formatnya telah berubah dan sesuai dengan langkah yang ada. Prediksi data menentukan data yang dikelompokkan. Berikut langkah dari pengelompokan

2.1 pengumpulan data

Penelitian ini menggunakan data sekunder yang diperoleh <https://www.kaggle.com/datasets>. Data yang digunakan mencakup jumlah provinsi di Indonesia yang tercantum sebagai penduduk miskin. Pengambilan data pada penelitian dilakukan untuk mendapatkan informasi penting agar tujuan penelitian dapat tercapai. Data yang digunakan berdasarkan pada dokumen yang menggambarkan jumlah penduduk miskin di Indonesia. Data yang digunakan dapat diolah kemudian setelah dilakukan proses clustering (Sepriyanti et al., 2022)

2.2 Data Mining

Data mining merupakan proses yang mempekerjakan satu atau lebih teknik pembelajaran computer secara otomatis untuk menganalisa dan mengekstraksi pengetahuan. Data mining sering disebut dengan istilah Knowledge discovery in database (KDD). Knowledge discovery in database (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Hendrastuty, 2024)

2.3 Clustering

Clustering merupakan teknik dalam data mining yang berfungsi mengelompokkan sekumpulan objek ke dalam beberapa *clustering* memiliki kemiripan karakteristik sedemikian sehingga objek pada satu *clustering* namun tidak mirip dengan objek di cluster yang berbeda. Cluster adalah kumpulan objek yang serupa diantara kelompoknya dan berbeda dengan objek milik kelompok lain. Ada 2 metode clustering, yaitu hierarchical clustering dan partitional clustering. Data dikelompokkan melalui bagan hierarki pada metode hierarchical clustering, dimana dua grup terdekat digabung atau seluruh data dibagi ke dalam cluster. Pada partitional clustering data dikelompokkan tanpa ada struktur hierarki, setiap cluster mempunyai centroid dengan tujuan untuk meminimalkan jarak dari seluruh data ke centroid (Nabila et al., 2022)

2.4 K-Means

Metode K-Means melibatkan penentuan dan konfigurasi jumlah cluster yang akan dijadikan sebagai pusat atau acuan dalam proses pengelompokan data menggunakan library K_Means. Sebelum algoritma K-Means diterapkan, data perlu melalui tahap *preprocessing* terlebih dahulu. K-Means tergolong dalam metode *partitioning clustering* yang membagi data menjadi bagian-bagian yang terpisah. Algoritma ini sangat populer karena kemudahan penggunaannya dan kemampuannya dalam mengelompokkan data besar serta menangani *outlier* (Uddin & Fatah, 2024)

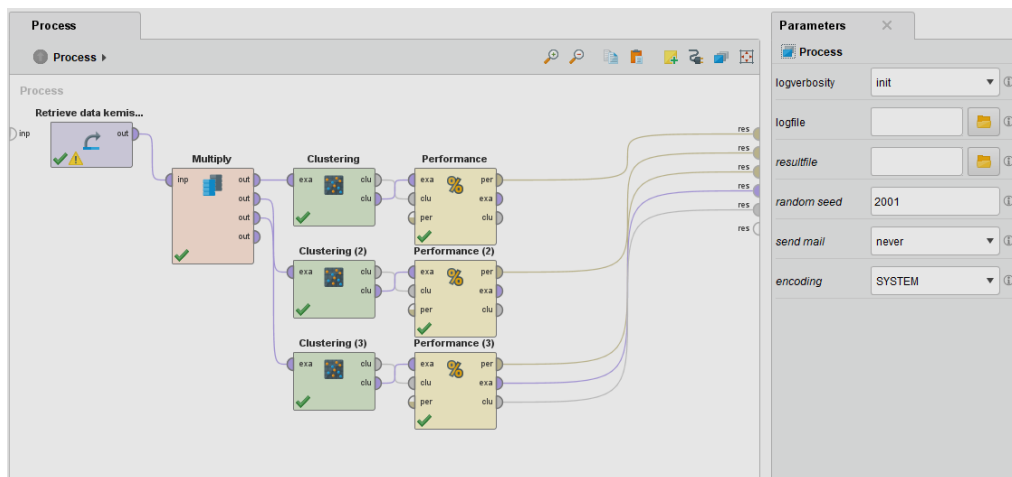
3. HASIL DAN PEMBAHASAN

3.1 Centroid Data

Sebelum melakukan proses clustering dengan menggunakan algoritma K-Means, maka yang perlu dilakukan adalah menentukan titik tengah atau centroid data. Dalam proses menentukan centroid data maka RapidMiner Studio akan otomatis menentukan nilai-nilai dari centroid data tersebut.

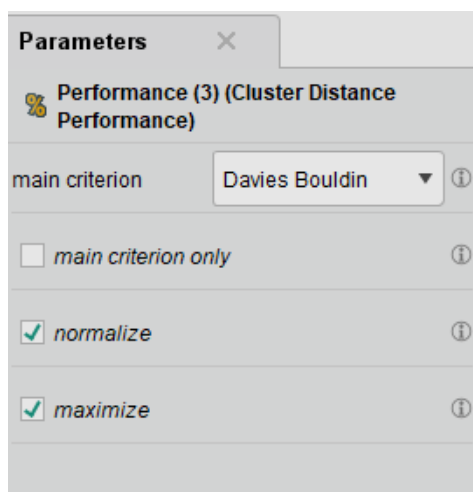
3.2 Model Data Mining

Dalam proses data klasterisasi menggunakan K-Means, pemilihan nilai K yang paling optimal dilakukan dengan menggunakan operator performance untuk menghitung nilai Davies-Bouldin Index (DBI). Pengujian dilakukan dengan mengatur nilai K mulai dari 2 hingga 10. Pengaturan parameter dalam algoritma K-Means merupakan langkah krusial. Pengaturan ini dirancang untuk memaksimalkan kinerja algoritma serta mengidentifikasi pola klaster yang paling optimal.



Gambar 1 Model Data Mining

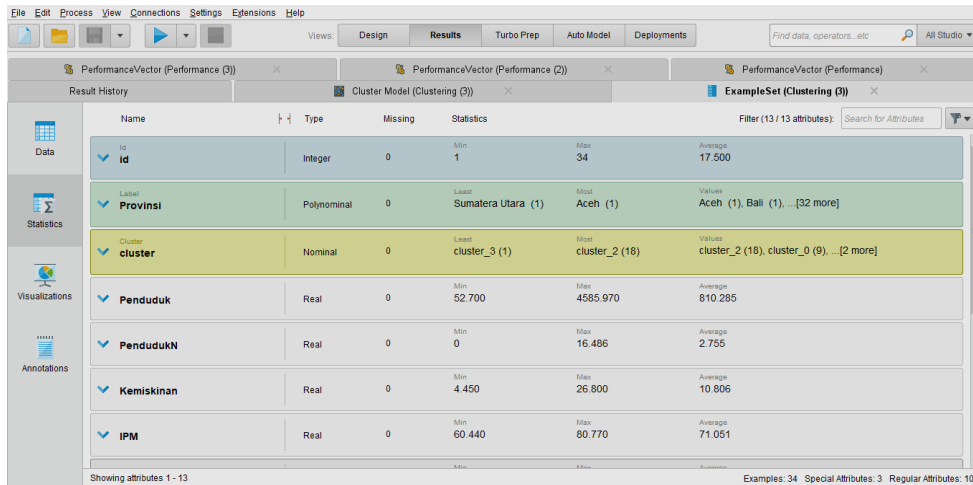
Ketika menggunakan operator clustering, penting untuk menentukan jumlah klaster yang optimal untuk dataset ini. Dalam penelitian ini, evaluasi dilakukan dengan menggunakan indeks Davies-Bouldin (DBI) pada operator Performance. Parameter yang diterapkan pada operator Performance dapat dilihat



Gambar 2 operator Performance

3.3 Data Proprocessing

Dalam tahap *Proprocessing*, dilakukan penghapusan atribut yang tidak relevan atau tidak diperlukan agar proses pengolahan data berjalan tanpa kendala.

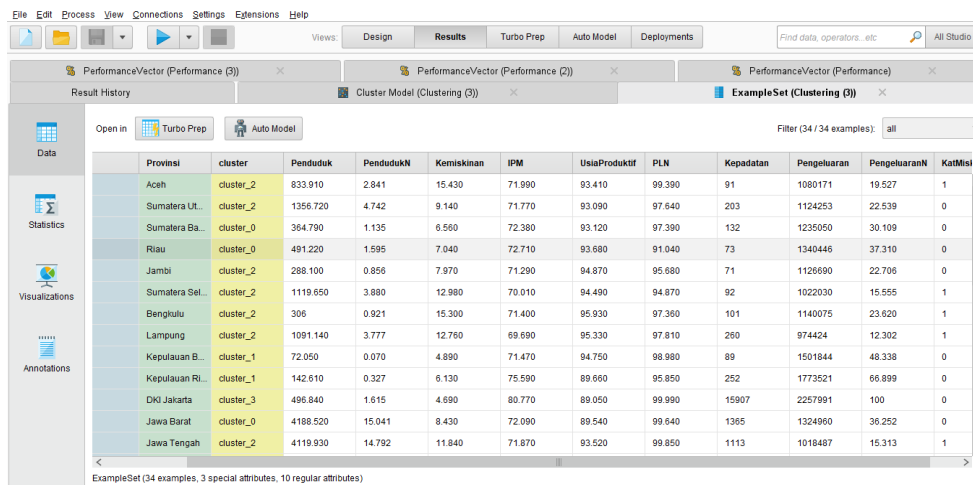


Gambar 3 3 Data Proprocessing

3.4. Clustering data dengan menggunakan RapidMIner Studio

Hasil dari clustering data terkait kabupaten/kota yang memiliki jumlah penduduk miskin terbanyak dengan rata-rata lama sekolah, dan pengeluaran per kapita di provinsi Banten menggunakan RapidMiner Studio :

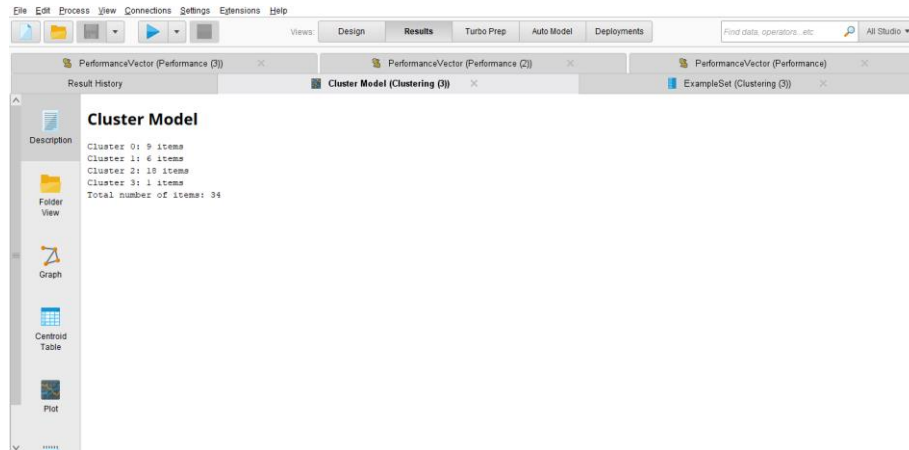
3.4.1. Visualisasi Data dengan menggunakan Data view. Dari data yang ada maka ditampilkan data awal:



Gambar 4 Clustering data dengan menggunakan RapidMIner Studio

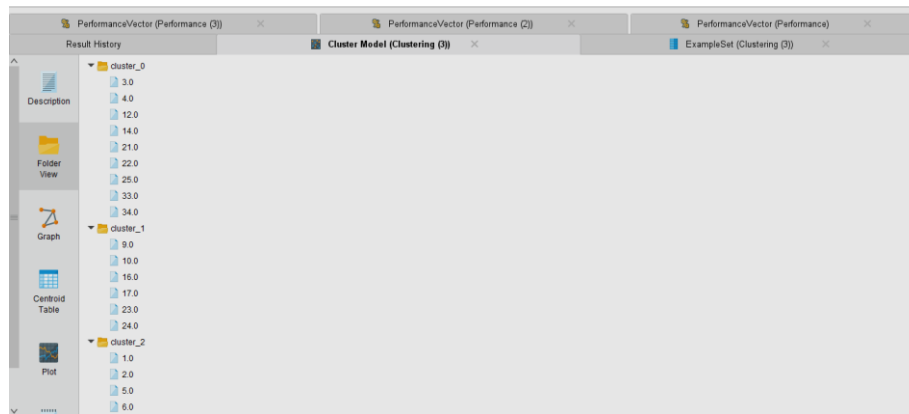
3.4.2. Menampilkan data cluster dengan menggunakan Cluster Model. Dari data awal maka terdapat 4 cluster antara lain :

- a. Cluster 0 : terdiri dari 9 items
- b. Cluster 1 : terdiri dari 6 items
- c. Cluster 2 : terdiri dari 18 items
- d. Cluster 3 : terdiri dari 1 items

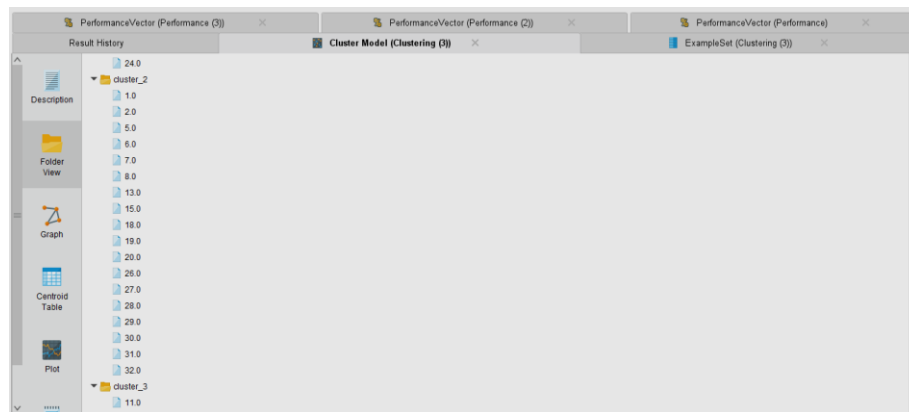


Gambar 5 data cluster

3.4.3. Mengelompokkan data berdasarkan clusternya menggunakan Folder View. Dari data awal maka dapat dikelompokkan menjadi 4 cluster antara lain.



Gambar 6 data cluster Folder View



Gambar 7 lanjutan data cluster Folder View

3.4.4. Menampilkan centroid data yang digunakan dengan menggunakan Centroid Table. Berikut centroid data yang digunakan :

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
Penduduk	802.186	260.985	1014.849	496.840
PendudukN	2.726	0.757	3.499	1.615
Kemiskinan	11.244	6.025	12.519	4.690
IPM	70.841	73.647	69.752	80.770
UsiaProduktif	93.778	92.717	94.751	89.050
PLN	88.564	97.132	93.986	99.990
Kepadatan	341.778	393	210.722	15907
Pengeluaran	1321974.444	1600326.500	1046107.111	2257991
PengeluaranN	36.048	55.066	17.200	100
KatMiskin	0.333	0	0.722	0

Gambar 8 Centroid Table

4. KESIMPULAN DAN SARAN

Penerapan algoritma K-Means Clustering dalam pengelompokan wilayah berdasarkan tingkat kemiskinan di Indonesia memberikan pendekatan yang efektif untuk menganalisis data dan mengidentifikasi pola-pola dalam distribusi kemiskinan. Secara keseluruhan, metode K-Means Clustering berpotensi menjadi alat yang berguna dalam mendukung upaya pengentasan kemiskinan, terutama bila didukung oleh data yang berkualitas dan analisis tambahan untuk validasi hasil. K-Means Clustering adalah metode yang efektif dalam memahami dan mengelompokkan wilayah berdasarkan tingkat kemiskinan. Dengan pengelolaan data yang tepat dan pemilihan parameter yang sesuai.

metode ini dapat menjadi alat yang kuat untuk mendukung pengembangan kebijakan yang lebih terarah dan berbasis bukti. Penelitian ini menunjukkan bahwa algoritma *K-Means Clustering* efektif dalam mengelompokkan wilayah berdasarkan tingkat kemiskinan. Klusterisasi ini dapat membantu pemerintah mengidentifikasi prioritas intervensi. Direkomendasikan agar pendekatan ini digunakan bersama metode lain untuk memperoleh hasil yang lebih komprehensif.

Ucapan Terima Kasih

Saya terima kasih dalam pembuatan jurnal” Penerapan K-Means Clustering untuk Pengelompokan Wilayah Berdasarkan Tingkat Kemiskinan di Indonesia” ini tidak lepas dari bimbingan dan bantuan dari berbagai pihak. Saya ucapkan terima kasih kepada dosen yang terkait. Yang telah membimbing dalam penelitian ini.

DAFTAR PUSTAKA

- [1] Arkham, D., & Swanjaya, D. (2020). K-Means Method For Clustering Public Service Assessment of Government Organization In Kediri City. *Prosiding SEMNAS INOTEK*, 155–160. <https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/79>
- [2] Bahauddin, A., Fatmawati, A., & Permata Sari, F. (2021). Analisis Clustering Provinsi Di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means. *Jurnal Manajemen Informatika Dan Sistem Informasi*, 4(1), 1–8. <https://doi.org/10.36595/misi.v4i1.216>
- [3] Dkk, A. K. (2015). *Indikator Kemiskinan dan Misklasifikasi Orang Miskin* (Ali Khomsa). november 2015.
- [4] Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (Jima-Ilkom)*, 3(1), 46–56. <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- [5] Mayasari, S. N., & Nugraha, J. (2023). Implementasi K-Means Cluster Analysis untuk Mengelompokkan Kabupaten/Kota Berdasarkan Data Kemiskinan di Provinsi Jawa Tengah Tahun 2022. *KONSTELASI: Konvergensi Teknologi Dan Sistem Informasi*, 3(2), 317–329. <https://doi.org/10.24002/konstelasi.v3i2.7200>
- [6] Nabila, H., Retno, D., & Saputro, S. (2022). Clustering Data Campuran Numerik dan Kategorik Menggunakan Algoritme Ensemble Quick RObust Clustering using linKs (QROCK). *Prisma, Prosiding Seminar Nasional Matematika*, 5(1), 716–720. <https://journal.unnes.ac.id/sju/index.php/prisma/article/view/54590>
- [7] prof. Dr. Ir. Keppi Sukei, M. (ed). (2015). *GENDER & KEMISKINAN DI INDONESIA* (Keppi Sukei). juli 2015.
- [8] Sepriyanti, N., Sani Nahampun, R., Zikri, M. H., Ambarani, I., & Rahmadayan, A. (2022). SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat Implementation of K-Means Clustering to Group Poverty Levels in Riau Province Penerapan K-Means Clustering

- Untuk Mengelompokkan Tingkat Kemiskinan di Provinsi Riau. *Seminar Nasional Penelitian Dan Pengabdian Masyarakat*, 59–65. <https://journal.irpi.or.id/index.php/sentimas>
- [9] Uddin, M. B., & Fatah, Z. (2024). *Gudang Jurnal Multidisiplin Ilmu Penerapan Data Mining Clustering K-Means Dalam Mengelompokkan Data Penduduk Penyandang Disabilitas*. 2(November), 86–94.