

Penerapan Data Mining Untuk Pengelompokan Kepadatan Penduduk Menurut Provinsi 2021 Menggunakan Algoritma *K-means* Dengan Rapid Miner

Nur Rizatul Mufidah^{1*}, Zaehol Fatah²

¹ Sistem Informasi, Universitas Ibrahimy Sukorejo Situbondo

² Sistem Informasi, Universitas Ibrahimy Sukorejo Situbondo

^{1*}nurrizatulmufidah06@email.com, ²zaeholfatah@email.com

Abstrak

Data mining merupakan bagian dari data analytics dan disiplin ilmu Data Science yang memiliki manfaat luas dan tepat guna. Kepadatan penduduk diberbagai provinsi akan mengakibatkan banyak dampak yang akan dialami oleh masyarakat, karena ketidaktahuan masyarakat akan lokasi kepadatan penduduk atau informasi tentang pengelompokan kepadatan penduduk disetiap provinsi, sehingga masyarakat yang memilih lokasi bertempat tinggal semuanya cenderung berada dalam pusaran permasalahan akibat dampak kepadatan penduduk disetiap provinsi yang ada. Untuk menghindari atau meminimalkan dampak kepadatan penduduk, maka perlu dilakukan pengelompokan kepadatan penduduk kedalam 3 kelompok (cluster), yaitu sangat padat(cluster 2), padat(cluster 0), dan sedang(cluster 1). Algoritma yang digunakan dalam penelitian ini adalah algoritma K-Means dengan rapid miner. Dengan hasil penggunaan algoritma K-Means dengan rapid miner menghasilkan bahwa dari 34 provinsi yang ada, terdapat 1 provinsi Sangat Padat (cluster 2), 6 provinsi padat (cluster 0), dan 27 kecamatan sedang (cluster 1). Sehingga dengan adanya cluster kepadatan penduduk ini akan memberikan pengetahuan baru kepada masyarakat yang melakukan urbanisasi, transmigrasi, dan imigrasi karena faktor pekerjaan, faktor ingin menetap, maupun faktor lainnya, sehingga menghindari atau meminimalisi dampak permasalahan karena faktor kepadatan penduduk.

Kata Kunci: Kepadatan penduduk, pengelompokan, Algoritma K-means, Rapid miner

PENDAHULUAN

Jumlah penduduk suatu daerah diukur dari jumlah masyarakat yang berdomisili didaerah tersebut, penambahan jumlah penduduk disuatu kota atau provinsi jika tidak di selesaikan dengan baik dan sistematis akan mengakibatkan dampak-dampak buruk. Pertambahan jumlah penduduk juga akan mempengaruhi pertumbuhan ekonomi dan kesejahteraan rakyat.[1]

Dengan banyaknya pertambahan jumlah penduduk di beberapa provinsi, diperlukan adanya pengelompokan jumlah penduduk berdasarkan provinsi sebagai salah satu upaya untuk mencegah kepadatan penduduk dan dampak-dampak buruk lainnya. Oleh karena itu, dibuatlah pengolahan data yang dapat menyelesaikan permasalahan tersebut yaitu menerapkan teknik data mining menggunakan algoritma *K-Means* terhadap data jumlah penduduk menurut provinsi. *K-Means* merupakan algoritma data clustering non hirarki yang berusaha mempartisi data yang ada dalam bentuk satu atau lebih cluster. Algoritma ini mempartisi data dalam cluster atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan dalam kelompok yang lain.[2]

Kepadatan penduduk akan menimbulkan banyak permasalahan, karena semakin besar tingkat penduduk, maka akan semakin besar memicu permasalahan, seperti kepadatan dan jumlah penduduk sangat signifikan mempengaruhi kapasitas ruas jalan dan volume lalu lintas. Peningkatan jumlah penduduk secara langsung dapat memicu meningkatnya kebutuhan akan penggunaan alat-alat transportasi seperti mobil dan motor secara tidak langsung dapat menyebabkan kemacetan, terlebih apabila kapasitas jalan yang ada tidak mampu mengakomodasi peningkatan jumlah kendaraan.[3]

Pengaruh kepadatan penduduk yang berperan kuat mempengaruhi sering terjadi tindak kriminal adalah aspek kepadatan penduduk karena banyaknya tingkat pengangguran maupun pekerjaan yang tidak menetap.[4] Pada penelitian yang dilakukan oleh Sonang Sonang dan tim menggunakan algoritma *K-Means* untuk mengelompokkan jumlah penduduk berdasarkan kategori usia. Hasil pengujian yang didapat yaitu kelompok jumlah penduduk tinggi berdasarkan usia ada 3 kecamatan sedangkan kelompok jumlah penduduk rendah berdasarkan usia ada 5 kecamatan.[5]

Pada penelitian yang dilakukan oleh Preddy Marpaung dan tim menggunakan algoritma K-means untuk mengelompokkan jumlah penduduk kabupaten Deli Serdang, hasil pengujian yang didapat bahwa dari 22 kecamatan yang ada, terdapat 3 kecamatan Sangat Padat (cluster1), 4 kecamatan padat (cluster2), dan 15 kecamatan sedang (cluster3). Juga pada penelitian yang dilakukan oleh Ririn Restu Aria dan tim menggunakan metode *K-Means* untuk

menentukan cluster penerima program bantuan, perhitungan menghasilkan 14 provinsi cluster rendah dan 20 provinsi cluster tinggi.[6]

Berdasarkan permasalahan dan dampak yang akan dialami masyarakat karena faktor kepadatan penduduk maka dalam penelitian ini perlu dilakukan pengelompokan kedalam beberapa kelompok (cluster) setiap provinsi yang ada, yaitu penduduk Sedang(C1), Padat(C0) dan Sangat Padat(C2). Dalam pengelompokan kepadatan penduduk menerapkan data mining menggunakan algoritma *K-means*. Algoritma *K-means* merupakan salah satu algoritma pengelompokan yang paling populer dan sederhana, dimana pertama kali diterbitkan pada tahun 1955 dan sampai sekarang *K-means* masih banyak digunakan.[7]

Dengan dilakukannya penelitian ini dapat memberikan pengetahuan baru kepada masyarakat, baik secara individu maupun kelompok yang melakukan urbanisasi, transmigrasi, dan imigrasi karena faktor pekerjaan, faktor ingin menetap, maupun faktor lainnya. Hasil dari penelitian ini sangat bermanfaat bagi masyarakat luas yaitu dengan adanya pengetahuan baru bagi masyarakat, bisa mendukung dalam pengambil keputusan lokasi bertempat tinggal, sehingga menghindari atau meminimalisir dampak permasalahan karena faktor kepadatan penduduk. Berdasarkan itu, dibuatlah judul penelitian ini “Penerapan Data Mining Untuk Pengelompokan Kepadatan Penduduk Menurut Provinsi 2021 Menggunakan Algoritma *K-means* Dengan Rapid Miner”.

METODE

Tahapan Penelitian

Penelitian dimulai dengan melakukan literatur terhadap data kependudukan setiap provinsi melalui situs resmi badan pusat statistik, kemudian dilanjutkan literatur dari jurnal terkait masalah utama dampak kepadatan penduduk yang akan dialami masyarakat dan juga algoritma yang sesuai untuk digunakan dalam penyelesaian masalah. Algoritma *K-Means* digunakan untuk mengelompokkan kepadatan penduduk kedalam kelompok (cluster) dengan kategori sangat padat, padat dan sedang. Tahapan yang akan dilakukan dalam penelitian ini secara keseluruhan dapat dilihat sebagai berikut:

Gambar 1. Tahapan Penelitian

Data Mining

Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar[8]. Data Mining juga dilakukan untuk menemukan pola dari data yang besar, data tersebut dapat disimpan untuk penyimpanan informasi, serta data warehouse[9]. Data mining atau tambang data merupakan metode, teknik, artificial intelligent dan mesin pembelajaran yang diekstraksi sehingga menghasilkan suatu pengetahuan dan informasi yang berguna yang tersimpan dalam suatu database besar[10].

Clustering

Clustering adalah sekumpulan objek data yang memiliki kemiripan satu dengan yang lain didalam cluster yang sama dan tidak mirip dengan objek didalam cluster yang lain[11], clustering merupakan alur proses tempat satu set subjek informasi ke dalam himpunan bagian yang disebut dengan klaster. klastering banyak dipakai dalam bermacam aplikasi ataupun web semacam misalnya pada business intelligence, identifikasi pola pandangan, mesin pencarian, aspek ilmu kesehatan, serta buat keamanan (security). Di dalam business intelligence, klaster dapat menata banyak informasi atau data ke dalam jumlah golongan[12].

K-Means

Algoritma *K-Means Clustering* adalah suatu metode penganalisaan data atau metode data mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan menggunakan metode dengan mengelompokkan data dengan cara

partisi. Metode K-Means mempunyai cara kerja yaitu data dibuat kedalam beberapa kelompok, data yang sama dibuat kedalam kelompok yang sama data yang berbeda kedalam satu kelompok yang berbeda pula[13].

Rapid Miner

Rapid Miner adalah sebuah analisis teks yang fokus didalam pekerjaan yang dilakukan oleh text mining, yang melibatkan penggalian pola dari kumpulan data besar dan menggabungkannya dengan Teknik statistik, kecerdasan buatan, dan basis data. Tujuan dari analisis teks ini adalah untuk menghasilkan informasi yang akurat dan bermakna dari teks yang sangat besar atau kompleks[14]. RapidMiner memakai bermacam metode deskriptif serta perkiraan dalam membagikan pengetahuan pada konsumennya alhasil bisa membuat ketetapan yang sangat pas serta cermat[12].

HASIL DAN PEMBAHASAN

Data Selection

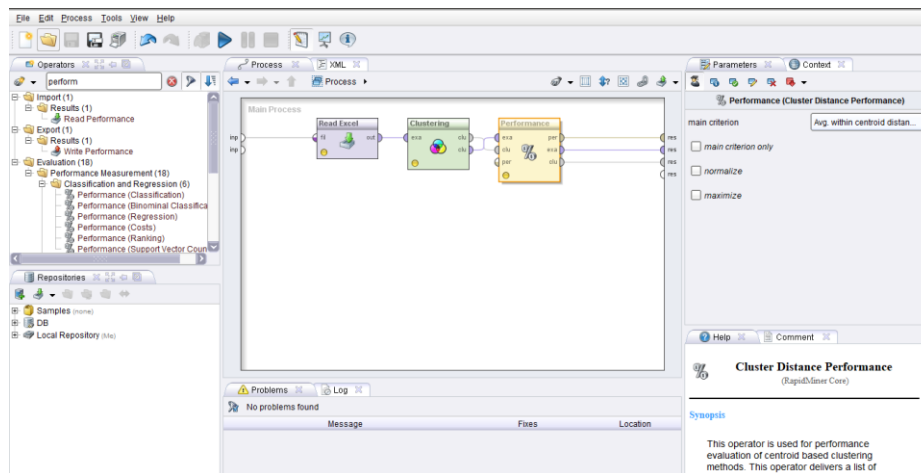
Adapun dataset yang digunakan dalam penelitian ini adalah jumlah penduduk pada tahun 2021 yang diambil dari 34 provinsi di Indonesia[15], sebagai berikut:

Tabel 1. Data Kepadatan Penduduk menurut Provinsi (jiwa/km2)2021

No	Provinsi	Kepadatan Penduduk (jiwa/km2)2021
1	Aceh	92
2	Sumatera Utara	205
3	Sumatera Barat	133
4	Riau	75
5	Jambi	72
6	Sumatera Selatan	93
7	Bengkulu	102
8	Lampung	262
9	KEP. Bangka Belitung	90
10	KEP. Riau	258
11	DKI Jakarta	15978
12	Jawa Barat	1379
13	Jawa Tengah	1120
14	DI Yogyakarta	1185
15	Jawa Timur	855
16	Banten	1248
17	Bali	755
18	Nusa Tenggara Barat	290
19	Nusa Tenggara Timur	111
20	Kalimantan Barat	37
21	Kalimantan Tengah	18
22	Kalimantan Selatan	106
23	Kalimantan Timur	30

24	Kalimantan Utara	9
25	Sulawesi Utara	190
26	Sulawesi Tengah	49
27	Sulawesi Selatan	196
28	Sulawesi Tenggara	70
29	Gorontalo	105
30	Sulawesi Barat	86
31	Maluku	40
32	Maluku Utara	41
33	Papua Barat	11
34	Papua	14

Data yang telah tersusun dimasukkan kedalam aplikasi *Rapid Miner* untuk melakukan pengujian pada gambar 1.

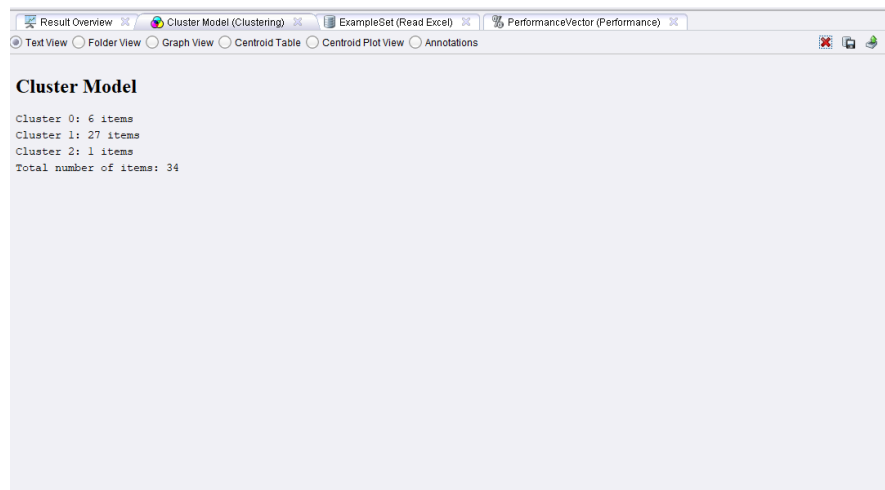


Gambar 1. Clustering K-Means dengan tools RapidMiner

Row No.	Provinsi	id	cluster	Kepadatan
1	ACEH	1	cluster_1	92
2	SUMATERA	2	cluster_1	205
3	SUMATERA	3	cluster_1	133
4	RIAU	4	cluster_1	75
5	JAMBI	5	cluster_1	72
6	SUMATERA	6	cluster_1	93
7	BENGKULU	7	cluster_1	102
8	LAMPUNG	8	cluster_1	262
9	KEP. BANGK.	9	cluster_1	90
10	KEP. RIAU	10	cluster_1	258
11	DKI JAKARTA	11	cluster_2	15978
12	JAWA BARAT	12	cluster_0	1379
13	JAWA TENG.	13	cluster_0	1120
14	DI YOGYAKARTA	14	cluster_0	1185
15	JAWA TIMUR	15	cluster_0	855
16	BANTEN	16	cluster_0	1248
17	BALI	17	cluster_0	755
18	NUSA TENG.	18	cluster_1	290
19	NUSA TENG.	19	cluster_1	111
20	KALIMANTAN	20	cluster_1	37
21	KALIMANTAN	21	cluster_1	18
22	KALIMANTAN	22	cluster_1	106
23	KALIMANTAN	23	cluster_1	30
24	KALIMANTAN	24	cluster_1	9

Gambar 2. ExampleSet Result

Pada gambar 2. menampilkan hasil dari *clustering* data, label cluster terbagi menjadi 3 kelompok yaitu cluster_0, cluster_1, dan cluster_2. Pembagian ini berdasarkan hasil dari centroid tiap masing-masing data dengan jarak terdekat.



Gambar 3. Cluster Model

Pada gambar 3. Hasil pembagian data terdapat tiap *cluster*, cluster 0 terdapat 6 provinsi, cluster 1 27 provinsi, cluster 2 terdapat 1 provinsi, total dari dataset yang diuji adalah 34 provinsi.

Evaluasi

Berdasarkan hasil implementasi algoritma *K-Means* maka didapatkan hasil performance yang peneliti lakukan dengan *centroid distance* -13758.899, *centroid distance_cluster_0* -6466.090, *centroid distance_cluster_1* -0.000, *centroid distance_cluster_2* -47643.222 dan *davies bouldin* dengan hasil -0.175 seperti pada gambar 2.



Gambar 4. Data PerformanceVector

Berdasarkan hasil yang telah diuji pada proses *clustering K-Means*, cluster_0 terdapat 6 provinsi yaitu cluster padat, cluster_1 terdapat 27 provinsi yaitu cluster sedang, cluster_2 terdapat 1 provinsi yaitu cluster sangat padat. Berikut hasil pengujian *Clustering K-Means*

Tabel 2. Hasil Pengujian *Clustering*

Id	Provinsi	Cluster	Kepadatan Penduduk (jiwa/km ²)2021
1	Aceh	cluster_1	92
2	Sumater Utara	cluster_1	205
3	Sumatera Barat	cluster_1	133
4	Riau	cluster_1	75
5	Jambi	cluster_1	72
6	Sumatera Selatan	cluster_1	93
7	Bengkulu	cluster_1	102

8	Lampung	cluster_1	262
9	KEP.Bangka Belitung	cluster_1	90
10	KEP.Riau	cluster_1	258
11	DKI Jakarta	cluster_2	15978
12	Jawa Barat	cluster_0	1379
13	Jawa Tengah	cluster_0	1120
14	DI Yogyakarta	cluster_0	1185
15	Jawa Timur	cluster_0	855
16	Banten	cluster_0	1248
17	Bali	cluster_0	755
18	Nusa Tenggara Barat	cluster_1	290
19	Nusa Tenggara Timur	cluster_1	111
20	Kalimantan Barat	cluster_1	37
21	Kalimantan Tengah	cluster_1	18
22	Kalimantan Selatan	cluster_1	106
23	Kalimantan Timur	cluster_1	30
24	Kalimantan Utara	cluster_1	9
25	Sulawesi Utara	cluster_1	190
26	Sulawesi Tengah	cluster_1	49
27	Sulawesi Selatan	cluster_1	196
28	Sulawesi Tenggara	cluster_1	70
29	Gorontalo	cluster_1	105
30	Sulawesi Barat	cluster_1	86
31	Maluku	cluster_1	40
32	Maluku Utara	cluster_1	41
33	Papua Barat	cluster_1	11
34	Papua	cluster_1	14

KESIMPULAN

Berdasarkan penelitian yang sudah dilakukan dapat menyimpulkan dengan menerapkan algoritma K-Means, dapat mengelompokkan kepadatan penduduk kedalam tiga kelompok (cluster), yaitu cluster_0 daerah penduduk padat terdapat 6 provinsi, cluster_1 daerah penduduk sedang terdapat 27 provinsi dan cluster_2 daerah penduduk sangat padat terdapat 1 provinsi. Dengan adanya pengetahuan pengelompokan kepadatan penduduk ini, akan meminimalkan dampak akibat kepadatan penduduk bagi setiap orang yang mau bertempat tinggal di beberapa provinsi di Indonesia, untuk kedepannya perlu dilakukan penelitian pengelompokan kepadatan penduduk berdasarkan kota, kabupaten maupun kecamatan. Berikut hasil dari pengelompokan kepadatan penduduk yang sudah dilakukan:

1. Cluster_0 terdapat 6 Provinsi yaitu: Bali, Jawa Timur, Jawa Tengah, DI Yogyakarta, Banten, Jawa Barat
2. Cluster_1 terdapat 27 Provinsi yaitu: Kalimantan Utara, Papua Barat, Papua, Kalimantan Tengah, Kalimantan Timur, Kalimantan Barat, Maluku, Maluku Utara, Sulawesi Tengah, Sulawesi Tenggara, Jambi, Riau, Sulawesi Barat, KEP.Bangka Belitung, Aceh, Sumatera Selatan, Bengkulu, Gorontalo, Kalimantan Selatan, Nusa Tenggara

Timur, Sumatera Barat, Sulawesi Utara, Sulawesi Selatan, Sumatera Utara, KEP.Riau, Lampung, Nusa Tenggara Barat.

3. Cluster_2 terdapat 1 Provinsi yaitu: DKI Jakarta.

UCAPAN TERIMA KASIH

Peneliti mengucapkan banyak terimakasih kepada pihak pihak yang membantu menyelesaikan penelitian ini dari tahap awal sampai tahap akhir atau publish.

DAFTAR PUSTAKA

- [1] ... 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.
- [2] Y. F. S. Y. Damanik, S. Sumarno, I. Gunawan, D. Hartama, and I. O. Kirana, "Penerapan Data Mining Untuk Pengelompokan Penyebaran Covid-19 Di Sumatera Utara Menggunakan Algoritma K-Means," *J. Ilmu Komput. dan Inform.*, vol. 1, no. 2, pp. 109–132, 2021, doi: 10.54082/jiki.13.
- [3] M. I. Ali and M. R. Abidin, "Pengaruh Kepadatan Penduduk Terhadap Intensitas Kemacetan Lalu Lintas Di Kecamatan Rappocini Makassar," *Pros. Semin. Nas. Lemb. Penelit. Univ. Negeri Makassar*, pp. 68–73, 2019.
- [4] R. M. Sabiq and N. Nurwati, "Pengaruh Kepadatan Penduduk Terhadap Tindakan Kriminal," *J. Kolaborasi Resolusi Konflik*, vol. 3, no. 2, p. 161, 2021, doi: 10.24198/jkrk.v3i2.35149.
- [5] S. Sonang, A. T. Purba, and F. O. I. Pardede, "Pengelompokan Jumlah Penduduk Berdasarkan Kategori Usia Dengan Metode K-Means," *J. Tek. Inf. dan Komput.*, vol. 2, no. 2, p. 166, 2019, doi: 10.37600/tekinkom.v2i2.115.
- [6] R. R. A. Aria, S. Susilowati, and I. R. Rahadjeng, "Data Mining Menentukan Cluster Penerima Program Bantuan dengan Metode K-Means," *Remik*, vol. 7, no. 1, pp. 291–300, 2023, doi: 10.33395/remik.v7i1.12030.
- [7] P. Marpaung and N. Tarigan, "Data Mining for Determining Book Loan Patterns in-Library Using Apriori Algorithm," *Infokum*, vol. 8, no. 1, Desember, pp. 11–15, 2019, [Online]. Available: <http://infor.seaninstitute.org/index.php/infokum/index>
- [8] E. taufiq luthfi Kusri, *Algoritma Data Mining*. 2009. [Online]. Available: https://www.google.co.id/books/edition/Algoritma_Data_Mining/-Ojclag73O8C?hl=id&gbpv=1&dq=k-means+clustering+bahasa+indonesia&printsec=frontcover
- [9] A. Yogiarto, A. Homaidi, and Z. Fatah, "Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1720–1728, 2024, doi: 10.33379/gtech.v8i3.4495.
- [10] Yulia and M. Silalahi, "Penerapan Data Mining Clustering Dalam Mengelompokan Buku Dengan Metode K-Means," *Indones. J. Comput. Sci.*, vol. 10, no. 1, 2021, doi: 10.33022/ijcs.v10i1.3008.
- [11] M. K. dkk Rani Rotul Muhima, *Kupas Tuntas Algoritma Clustering konsep, penghitungan manual dan program*. 2021. [Online]. Available: https://www.google.co.id/books/edition/KUPAS_TUNTAS_ALGORITMA_CLUSTERING_KONSEP/H55rEAAAQBAJ?hl=id&gbpv=1&dq=clustering+bahasa+indonesia&pg=PA2&printsec=frontcover
- [12] B. Provinsi, D. I. Indonesia, Y. Anggiat, J. R. Hermawan, T. C. Widiyanto, and S. Anggraeni, "DATA MINING CLUSTERING PENYEBARAN VIRUS COVID-19," vol. 6, no. 2, pp. 427–435, 2022.
- [13] Nurhayati, *Pemodelan K-Means Algoritma dan Big Data Analysis (Pemetaan Data Musytahiq)*. 2022. [Online]. Available: https://www.google.co.id/books/edition/PEMODELAN_K_MEANS_ALGORITMA_DAN_BIG_DATA/_bJmEAAAQBAJ?hl=id&gbpv=1&dq=k-means+clustering+bahasa+indonesia&pg=PA129&printsec=frontcover
- [14] M. Jundanuddin *et al.*, "Implementasi Data Mining Pada Penilaian Kinerja Guru Menggunakan Metode K-Nearest Neighbor (K-NN) Studi Kasus Pada Mi Almagiri Silo-Jember Implementation Of Data Mining On Teacher Performance Appraisal Using K-Nearest Neighbor (K-NN) Method Case Study At Mi Almagiri Silo-Jember," vol. 13, no. 105, pp. 869–880, 2024.
- [15] "Kepadatan Penduduk menurut Provinsi (jiwa/km2), 2021," Badan Pusat Statistik. Accessed: Mar. 31, 2023. [Online]. Available: <https://www.bps.go.id/id/statistics-table/2/MTQxIzI=/kepadatan-penduduk-menurut-provinsi--jiwa-km2-.html>