

Pengelompokkan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means

Sena Wijayanto^{*1}, M Yoka Fathoni^{*2}

^{*1,2}Program Studi S1 Sistem Informasi, Institut Teknologi Telkom Purwokerto
Jl. DI Panjaitan No.128, Kec. Purwokerto Selatan, Kab. Banyumas, Jawa Tengah, 53147
e-mail: *sena@ittelkom-pwt.ac.id, myokafathoni@ittelkom-pwt.ac.id

Abstrak

Padi merupakan salah satu sumber makanan pokok yang dikonsumsi masyarakat Indonesia. Namun kebutuhan beras masyarakat yang semakin meningkat tidak diimbangi dengan pertumbuhan area pertanian. Luas panen kota/kabupaten di provinsi Jawa Tengah di tahun 2020 mengalami penurunan dibandingkan tahun sebelumnya. Selain itu produksi padi juga mengalami penurunan. Oleh karena itu provinsi Jawa Tengah perlu mengelompokkan kota/kabupaten penghasil padi untuk mengetahui daerah yang memiliki potensi dalam memproduksi padi. Metode yang dapat digunakan dalam mengelompokkan kota/kabupaten penghasil padi adalah clustering. Clustering dapat mengelompokkan data tersebut berdasarkan kesamaan jenis data ke dalam cluster. Pada penelitian ini dilakukan pengelompokkan kota/kabupaten penghasil padi di Jawa Tengah dengan metode clustering. Clustering yang digunakan adalah dengan menggunakan algoritma K-Means. Hasil yang didapatkan dari metode ini yaitu 3 cluster antara lain C0 menggambarkan daerah yang hasil produktivitas padinya sedang terdiri dari 12 daerah. C1 menggambarkan daerah yang hasil produktivitas padinya rendah terdiri dari 18 daerah. Serta C2 menggambarkan daerah yang hasil produktivitas padinya tinggi terdiri dari 12 daerah.

Kata kunci—K-Means, Clustering, Data Mining, Padi

Abstract

Rice is one of the staple food sources consumed by the people of Indonesia. But the growing need for rice in the Community is not accompanied by a growth in the agricultural area. The harvested area of cities/districts in Central Java Province in 2020 decreased compared to the previous year. In addition, rice production has also decreased. Thus, the central province of Java needs to classify the rice-producing towns/regencies to discover the areas that have the potential to produce rice. The method that can be used for classifying rice-producing cities/regencies is clustering. Clustering can group data based on the similarity of data types in clusters. In this study, rice-producing cities/regencies in Central Java were grouped using the clustering method. The clustering method is the K-Means algorithm. The results obtained from this method are 3 clusters. C0 describing an area with moderate rice productivity consisting of 12 regions. C1 describing an area with low rice productivity composed of 18 regions. And C2 is describing a high paddy productivity area made up of 12 regions.

Keywords— K-Means, Clustering, Data Mining, Paddy

1. PENDAHULUAN

Salah satu sumber makanan pokok yang sebagian besar dikonsumsi oleh masyarakat Indonesia adalah beras [1]. Tanaman padi (*Oryza sativa L.*) merupakan tanaman penghasil beras. Tanaman padi tumbuh pada daerah yang memiliki iklim tropis atau sub

tropis [2]. Kondisi geografis Indonesia yang mendukung serta memiliki iklim tropis sangat cocok untuk membudidayakan padi. Sehingga pada tahun 2018, Indonesia berhasil menjadi negara penghasil beras terbanyak di dunia [3].

Untuk memenuhi kebutuhan pangan negara Indonesia, produksi padi perlu dioptimalkan. Hal ini bertujuan agar pemerintah dapat berhasil memenuhi kebutuhan beras di masyarakat sehingga pemerintah tidak perlu lagi mengimpor beras dari luar negeri. Namun pada praktiknya, pemerintah masih mengimpor beras dari luar negeri untuk pemenuhan kebutuhan beras. Hal ini dikarenakan kebutuhan beras pada masyarakat terus mengalami peningkatan tiap tahunnya yang disebabkan semakin meningkatnya laju pertumbuhan penduduk. Peningkatan pertumbuhan penduduk tersebut tidak diimbangi oleh pertumbuhan area pertanian. Area pertanian saat ini banyak alih fungsi menjadi lahan bangunan [1].

Provinsi Jawa Tengah terdiri dari 35 kota/kabupaten. Pada tahun 2018-2020, luas panen dan produksi padi semakin mengalami penurunan. Pada tahun 2018 provinsi Jawa Tengah memiliki luas panen sebesar 1.821.983,17 hektar. Pada tahun 2020 mengalami penurunan menjadi 1.684.746,24 hektar. Untuk produksi padi, pada tahun 2018 provinsi Jawa Tengah menghasilkan 10.499.588,23 ton. Namun pada tahun 2020 terjadi penurunan menjadi 9.586.910,98 ton [4]. Untuk mengoptimalkan produksi beras, provinsi Jawa Tengah dapat mengelompokkan kota atau kabupaten penghasil padi. Hal ini bertujuan untuk melihat dan mengetahui kota atau kabupaten yang memiliki potensi dalam memproduksi padi sekaligus mengetahui daerah yang memiliki produksi padi kurang maksimal.

Metode dalam data mining dapat menemukan pola-pola yang tidak terlihat dan menarik dari kumpulan data [5]. Salah satu metode data mining yang dapat digunakan untuk memetakan atau mengelompokkan data yang mirip adalah *clustering*. *Clustering* memiliki kelebihan dibandingkan metode data mining lain yaitu dapat mengklasifikasikan data meskipun tidak mempunyai pengetahuan sebelumnya [6]. *Clustering* membagi data serta mengelompokkan data tersebut berdasarkan kesamaan jenis data ke dalam *cluster* [7]. Terdapat beberapa algoritma yang termasuk dalam metode *clustering*, salah satu yang populer adalah K-Means.

Algoritma K-Means telah diterapkan dalam penyelesaian masalah di berbagai bidang. Di bidang pendidikan, algoritma K-Means digunakan untuk mengelompokkan data penerimaan mahasiswa baru [8]. Di bidang kesehatan algoritma K-Means digunakan untuk mengelompokkan data rekam medis pasien berdasarkan wilayah, jenis penyakit dan umur [5]. Pada bisnis, penentuan strategi marketing dapat ditemukan menggunakan algoritma K-Means [9]. Oleh karena itu penulis akan menggunakan algoritma K-Means untuk mengelompokkan daerah penghasil padi di Jawa Tengah sehingga diharapkan ditemukan pola-pola atau informasi tersembunyi dari data hasil produksi padi di Jawa Tengah.

2. METODE PENELITIAN

Pada penelitian ini terdapat beberapa tahapan dalam melaksanakan metode penelitian. Detail mengenai tahapan-tahapan akan dijelaskan di subbab selanjutnya.

2.1 Menganalisis Masalah

Pada tahap ini menentukan rumusan masalah yang akan dijadikan objek penelitian. Rumusan masalah ini nantinya digunakan untuk mencari alternatif pemecahan masalah dari masalah tersebut. Selain itu penentuan perumusan masalah dilakukan agar memahami ruang lingkup masalah dan langkah pemecahannya serta agar penelitian lebih terarah.

2.2 Studi Literatur

Pada tahap ini dilakukan pengumpulan berbagai macam literatur yang berkaitan dengan algoritma K-Means dan penggunaan software RapidMiner. Selain itu juga mengumpulkan dan mempelajari jurnal-jurnal pendukung mengenai padi dan algoritma K-Means. Literatur-literatur

tersebut digunakan untuk memperkuat metode yang akan digunakan dalam hal ini adalah algoritma K-Means.

2.3 Mengolah Data dengan Algoritma K-Means

Pada tahap ini, data yang dikumpulkan yaitu data produktivitas tanaman padi diolah menggunakan algoritma K-Means. Data produktivitas tanaman padi tersebut diseleksi terlebih dahulu lalu dimasukkan ke dalam rumus perhitungan. Dari hasil pengolahan data tersebut akan ditemukan hasil berupa pembagian *cluster*.

2.4 Mengolah Data dengan Aplikasi RapidMiner

Tahap ini melakukan pengolahan data menggunakan aplikasi RapidMiner. Data produktivitas tanaman padi pada tahap sebelumnya dianalisa dengan aplikasi RapidMiner menggunakan metode K-Means.

2.5 Menguji Hasil

Tahap ini dilakukan pengujian terhadap hasil pengolahan data dan perhitungan dengan algoritma K-Means yang dilakukan secara manual. Hasil tersebut dibandingkan dengan hasil yang didapatkan menggunakan RapidMiner sehingga dapat terlihat kesesuaian dan perbandingan antara hasil keputusan akhir dari perhitungan manual dengan menggunakan RapidMiner.

2.6 Menarik Kesimpulan

Pada tahap ini dilakukan penarikan kesimpulan dari proses yang telah dilakukan. Selain itu juga diberikan saran-saran yang bisa dijadikan untuk acuan di penelitian selanjutnya.

3. HASIL DAN PEMBAHASAN

Berikut adalah hasil dan pembahasan pada penelitian pengelompokan produktivitas tanaman padi di Jawa Tengah menggunakan metode *K-Means*.

3.1 Sumber Data

Data dalam penelitian ini diambil dari Badan Pusat Statistik (BPS) dimana data tersebut merupakan data luas panen, produksi, dan produktivitas padi menurut kabupaten/kota di provinsi Jawa Tengah (kuintal/hektar) tahun 2018-2020. Sebelum melakukan *clustering* data tersebut diseleksi terlebih dahulu yaitu hanya diambil data tahun 2020 saja seperti yang ditunjukkan pada tabel 1.

Tabel 1 Hasil Seleksi Data

Kabupaten / Kota	Luas Panen (ha)	Produksi (ton)	Produktivitas (ku/ha)
Kabupaten Cilacap	117296.51	761121.16	64.89
Kabupaten Banyumas	52929.85	292979.84	55.35
Kabupaten Purbalingga	29070.69	170241.62	58.56
Kabupaten Banjarnegara	19204.23	105737.42	55.06
Kabupaten Kebumen	73675.97	348910.62	47.36
Kabupaten Purworejo	50120.74	267240.47	53.32
Kabupaten Wonosobo	14694.42	78846.94	53.66
Kabupaten Magelang	36124.02	177662.43	49.18
Kabupaten Boyolali	49191.36	236823.5	48.14
Kabupaten Klaten	66729.3	370057.03	55.46

Kabupaten / Kota	Luas Panen (ha)	Produksi (ton)	Produktivitas (ku/ha)
Kabupaten Sukoharjo	46023.36	314992.36	68.44
Kabupaten Wonogiri	63141.93	353816.11	56.04
Kabupaten Karanganyar	43452.07	262399.7	60.39
Kabupaten Sragen	108953.93	723671.68	66.42
Kabupaten Grobogan	131929.86	805889.27	61.08
Kabupaten Blora	93986.48	477549.29	50.81
Kabupaten Rembang	31576.49	127772.83	40.46
Kabupaten Pati	102085.94	602806.71	59.05
Kabupaten Kudus	32051.02	180880.93	56.44
Kabupaten Jepara	40848.5	208619.49	51.07
Kabupaten Demak	106770.42	659420.83	61.76
Kabupaten Semarang	29466.41	167988.62	57.01
Kabupaten Temanggung	12827.19	79660.44	62.1
Kabupaten Kendal	33926.58	186593.02	55
Kabupaten Batang	30498.66	156003.61	51.15
Kabupaten Pekalongan	41176.44	193723.75	47.05
Kabupaten Pemalang	74155.86	408247.28	55.05
Kabupaten Tegal	65292.88	349850.75	53.58
Kabupaten Brebes	80256.77	477819.87	59.54
Kota Magelang	174.31	1057.13	60.65
Kota Surakarta	39.44	243.74	61.8
Kota Salatiga	672.69	4118.16	61.22
Kota Semarang	4165.43	23275.02	55.88
Kota Pekalongan	1628.49	8167.51	50.15
Kota Tegal	608	2721.85	44.77

3.2 Centroid Awal

Pada penerapan algoritma *K-Means*, ditentukan nilai *centroid* awal dari sumber data. Nilai *centroid* awal tersebut diambil secara acak [10]. Hasil *centroid* awal ditampilkan pada tabel 2.

Tabel 2 Centroid Awal

Kabupaten / Kota	Luas Panen (ha)	Produksi (ton)	Produktivitas (ku/ha)
Kabupaten Boyolali	49191.36	236823.5	48.14
Kabupaten Batang	30498.66	156003.61	51.15
Kabupaten Cilacap	117296.51	761121.16	64.89

3.3 Menentukan Jarak Euclidean

Langkah selanjutnya adalah menghitung jarak antara *centroid* dengan setiap data. Untuk menghitung jarak tersebut menggunakan rumus [11]:

$$D_{(x,y)} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (1)$$

Dimana x adalah data produktivitas tanaman padi yaitu luas panen, produksi dan produktivitas. Sedangkan y adalah *centroid* awal yang telah kita peroleh sebelumnya. Contoh perhitungan jarak antara data Kabupaten Banyumas dengan *centroid* pertama seperti di bawah ini:

$$D_{(x,y)} = \sqrt{(52929.85 - 49191.36)^2 + (292979.84 - 236823.5)^2 + (55.35 - 48.14)^2}$$

$$= 56280.64$$

Setiap data dihitung jarak dengan *centroid* sehingga mendapatkan hasil jarak euclidean seperti yang ditunjukkan pada tabel 3.

Tabel 3 Nilai Jarak Euclidean

C0	C1	C2
528702.5	611311	0
56280.64	138800.7	472545.6
69555.65	14309.44	597429.9
134472.3	51519.45	662683.9
114730.2	197680	414512.1
30431.17	112954.3	498428.2
161699.2	78758.65	689945.9
60587.02	22377.43	589078.1
0	82953.43	528702.5
134382.9	217098	394319.9
78233.03	159744.9	451786.2
117821.4	200487.8	410889.4
26212.24	107181.7	504158.8
490502.5	573063.9	38367.46
575049.1	657753.5	47099.03
244858.1	327753.5	284528.3
110464.2	28251.35	639122.9
369785.8	452501.7	159043.5
58509.51	24925.71	586468.7
29412.06	53624.16	557765.5
426501.9	509162.3	102243.6
71605.27	12029.38	599600.1
161315.2	78361.73	689421.9
52498.71	30780.88	580545.5
82953.43	0	611311
43838.65	39202.35	572480.6
173232	255993.8	355501.2
114168.4	196945	414545.2
242990.3	325640.3	285712.4
240807.9	157886	769035.1
241631.7	158710.1	769859.5
237709.6	154786.2	765933.8
218243.6	135315.6	746468.7
233550.4	150628.7	761786.2
239089.8	156169	767323.7

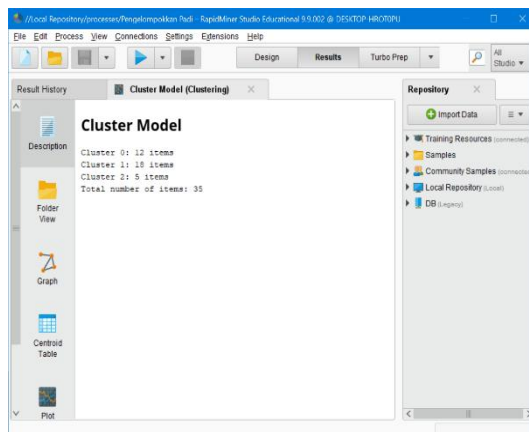
3.4 Clustering Data

Langkah berikutnya adalah menentukan letak *cluster* dari setiap data. Penentuan letak *cluster* ditentukan dengan cara memilih nilai yang paling kecil diantara C0, C1 dan C2. Pencarian *centroid* baru akan dilakukan kembali ketika semua cluster tidak berubah lagi [12]. Berdasarkan perhitungan menggunakan algoritma k-means menghasilkan 5 iterasi dengan hasil sebagai berikut:

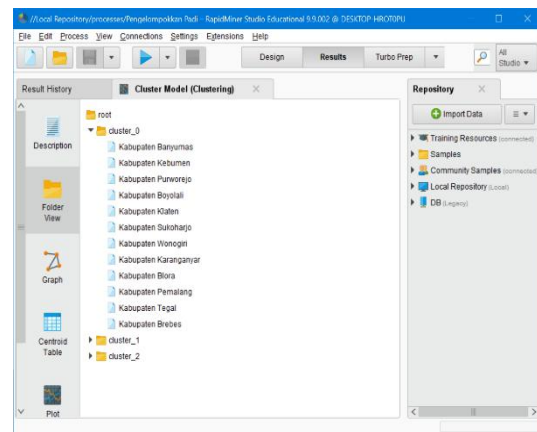
1. Cluster 0 (C0) : 12 anggota
2. Cluster 1 (C1) : 18 anggota
3. Cluster 2 (C2) : 5 anggota

3.5 Hasil Perhitungan RapidMiner

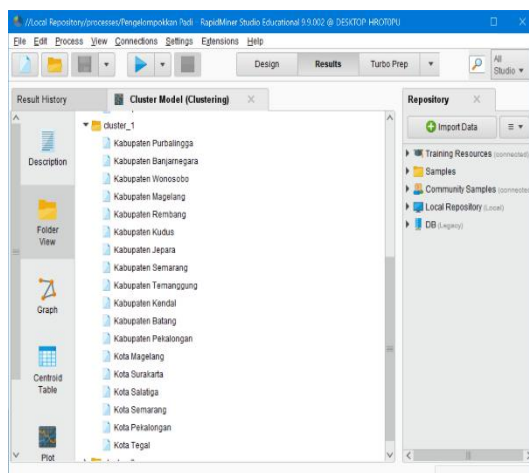
Proses selanjutnya adalah melakukan proses perhitungan dengan clustering K-Means pada data produktivitas tanaman padi sebanyak 37 data menggunakan aplikasi RapidMiner Studio 9.9. Hasil yang diperoleh dapat dilihat pada gambar 1 yaitu menampilkan keseluruhan cluster yang didapatkan. Gambar 2 menampilkan kota/kabupaten yang termasuk dalam cluster 0. Gambar 3 menampilkan kota/kabupaten yang termasuk dalam cluster 1. Gambar 4 menampilkan kota/kabupaten yang termasuk dalam cluster 2. Gambar 5 menampilkan visualisasi data cluster yang ditampilkan dalam bentuk diagram scater untuk melihat persebaran datanya.



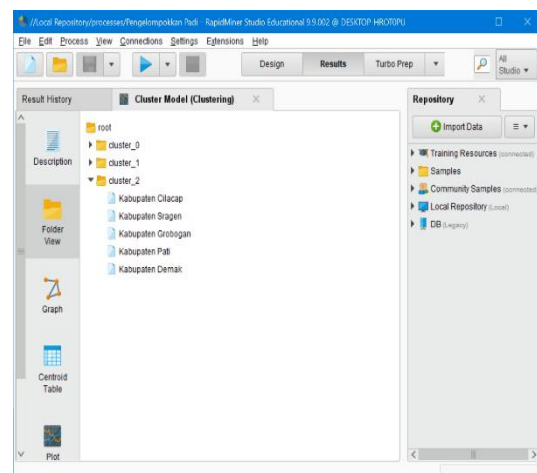
Gambar 1. Hasil cluster menggunakan RapidMiner



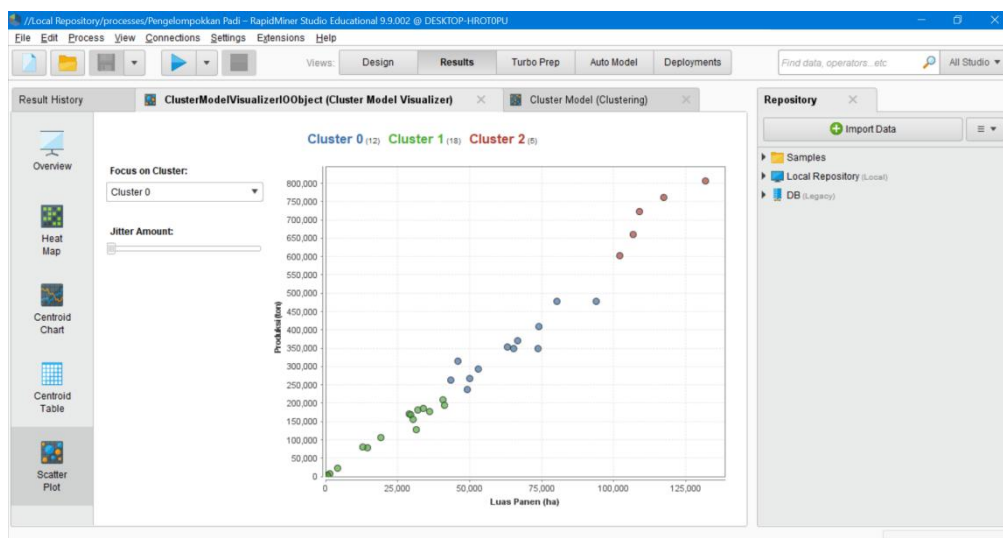
Gambar 2. Kabupaten/Kota yang masuk pada cluster 0



Gambar 3. Kabupaten/Kota yang masuk pada cluster 1



Gambar 4. Kabupaten/Kota yang masuk pada cluster 2



Gambar 5. Visualisasi data cluster

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan proses perhitungan menggunakan algoritma K-Means berhenti pada iterasi ke 5. Hasil cluster yang termasuk dalam C0 menggambarkan kota / kabupaten yang memiliki produktivitas padi sedang, C1 menggambarkan kota / kabupaten yang memiliki produktivitas padi rendah, dan C2 menggambarkan kota / kabupaten yang memiliki produktivitas padi tinggi. Kota / kabupaten yang termasuk dalam C0 berjumlah 12 daerah, yang termasuk dalam C1 berjumlah 18 daerah dan yang termasuk dalam C2 berjumlah 5 daerah.

Setelah dilakukan pengujian menggunakan aplikasi RapidMiner dengan data yang sama didapatkan hasil yang sama yaitu C0 menggambarkan daerah yang hasil produktivitas padinya sedang terdiri dari 12 daerah, C1 menggambarkan daerah yang hasil produktivitas padinya rendah terdiri dari 18 daerah, dan C2 menggambarkan daerah yang hasil produktivitas padinya tinggi terdiri dari 12 daerah.

5. SARAN

Terdapat beberapa saran yang dapat dijadikan sebagai acuan dalam penelitian berikutnya antara lain menggunakan metode *clustering* yang lain seperti Hierarchical Clustering, DBSCAN dll, kemudian membandingkannya dengan hasil yang didapatkan dengan K-Means. Selain itu, hasil proses data mining ini dapat digunakan sebagai bahan pertimbangan pengambilan keputusan lebih lanjut tentang kebijakan yang berkaitan dengan produksi padi di Jawa Tengah.

UCAPAN TERIMA KASIH

Pada kesempatan ini kami ucapkan banyak terima kasih kepada semua pihak yang telah memberikan kontribusinya di dalam penelitian ini. Terima kasih juga kepada JUPITER yang telah bersedia untuk mempublish jurnal tersebut.

DAFTAR PUSTAKA

- [1] M. J. Arifin, A. Basuki, and B. S. B. Dewantara, *Sistem Monitoring Cerdas untuk Mendeteksi Tingkat Pertumbuhan Tanaman Padi Menggunakan Drone*. Penerbit Qiara Media, 2021.
- [2] S. H. Y. Saragih, *BOTANI TANAMAN: Kajian Karakter Vegetatif dan Generatif Padi*. CV Literasi Nusantara Abadi, 2021.
- [3] F. Marisa et al., “Digitasi Produktivitas Panen Padi Berbasis K-Means Clustering,” *SMARTICS J.*, vol. 7, no. 1, pp. 21–26, 2021, doi: 10.21067/smartics.v7i1.5270.
- [4] BPS, “Luas Panen, Produksi, dan Produktivitas Padi Menurut Kabupaten/Kota di Provinsi Jawa Tengah (Kuintal/Hektar), 2018-2020,” 2020. <https://jateng.bps.go.id/indicator/53/463/1/luas-panen-produksi-dan-produktivitas-padi-menurut-kabupaten-kota-di-provinsi-jawa-tengah.html> (accessed Aug. 24, 2021).
- [5] R. Ordila, R. Wahyuni, Y. Irawan, and M. Yulia Sari, “PENERAPAN DATA MINING UNTUK PENGELOMPOKAN DATA REKAM MEDIS PASIEN BERDASARKAN JENIS PENYAKIT DENGAN ALGORITMA CLUSTERING (Studi Kasus : Poli Klinik PT.Inecda),” *J. Ilmu Komput.*, vol. 9, no. 2, pp. 148–153, 2020, doi: 10.33060/jik/2020/vol9.iss2.181.
- [6] C. Yuan and H. Yang, “Research on K-Value Selection Method of K-Means Clustering Algorithm,” *J 2019, Vol. 2, Pages 226-235*, vol. 2, no. 2, pp. 226–235, Jun. 2019, doi: 10.3390/J2020016.
- [7] M. Nishom and M. Y. Fathoni, “Implementasi Pendekatan Rule-Of-Thumb untuk Optimasi Algoritma K-Means Clustering,” *J. Inform. J. Pengemb. IT*, vol. 3, no. 2, pp. 237–241, 2018, doi: 10.30591/jpit.v3i2.909.
- [8] F. Yunita, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru,” *Sistemasi*, vol. 7, no. 3, p. 238, 2018, doi: 10.32520/stmsi.v7i3.388.
- [9] D. (Deni) Triyansyah and D. (Devi) Fitrihanah, “Analisis Data Mining Menggunakan Algoritma K-Means Clustering untuk Menentukan Strategi Marketing,” *InComTech J. Telekomun. dan Komput.*, vol. 8, no. 3, pp. 163–182, Oct. 2018, doi: 10.22441/INCOMTECH.V8I3.4174.
- [10] K. Fatmawati and A. P. Windarto, “Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (DBD) Berdasarkan Provinsi,” *Comput. Eng. Sci. Syst. J.*, vol. 3, no. 2, p. 173, 2018, doi: 10.24114/cess.v3i2.9661.
- [11] D. Darmansah, “Analisa Penyebab Kerusakan Tanaman Cabai Menggunakan Metode K-Means,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 2, pp. 126–134, 2020, doi: 10.35957/jatisi.v7i2.309.
- [12] Y. D. Darmi and A. Setiawan, “Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk,” *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2017, doi: 10.37676/jmi.v12i2.418.