

Penerapan Metode K-Means Clustering pada Hasil Produksi Beras di Wilayah Sumatera Utara

* Rafiqi Aidil Fitra

Universitas Negeri Medan

Email: rfqaidil.4211250001@mhs.unimed.ac.id

Rizki Risdah Sitorus

Universitas Negeri Medan

Email: rizkirisdahs@mhs.unimed.ac.id

Alamat : Jl. William Iskandar Ps. V, Kenangan Baru, Kec. Percut Sei Tuan, Kabupaten Deli Serdang, Sumatera Utara, Indonesia

*Korespondensi penulis: rfqaidil.4211250001@mhs.unimed.ac.id

Abstract. *This research aims to apply the method K-Means Clustering on rice production results in the region North Sumatra. K-Means Clustering is a data analysis technique that is useful for grouping data into several groups based on similar characteristics. This research uses rice production data from several districts/cities in North Sumatra as samples. The K Means method is used to group these regions into several clusters based on rice production they. The research results show that the K Means Clustering method can be used for identification rice production patterns in the North Sumatra region. The hope is, research results can be a valuable contribution, provide recommendations, and support initiatives for the North Sumatra provincial government in improving rice production throughout the region, with the aim of ensuring a more stable food supply for the community*

Keywords: *K-Means Clustering, Rice Production, North Sumatra, Data Analysis, Group*

Abstrak. Penelitian ini bertujuan untuk menerapkan metode K-Means Clustering pada hasil produksi beras di wilayah Sumatera Utara. K-Means Clustering adalah teknik analisis data yang berguna untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan kesamaan karakteristik. Penelitian ini menggunakan data produksi beras dari beberapa kabupaten/kota di Sumatera Utara sebagai sampel. Metode K-Means digunakan untuk mengelompokkan wilayah-wilayah tersebut ke dalam beberapa cluster berdasarkan produksi beras mereka. Hasil penelitian menunjukkan bahwa metode K-Means Clustering dapat digunakan untuk mengidentifikasi pola produksi beras di wilayah Sumatera Utara. Harapannya, hasil penelitian dapat menjadi kontribusi yang berharga, memberikan rekomendasi, serta mendukung inisiatif bagi pemerintah provinsi Sumatera Utara dalam meningkatkan produksi beras di semua wilayah, dengan tujuan memastikan pasokan makanan yang lebih stabil bagi masyarakat

Kata kunci: K-Means Clustering, Produksi Beras, Sumatera Utara, Analisis Data, Kelompok

LATAR BELAKANG

Penyempitan luas lahan pertanian menjadi lebih jelas karena laju pertumbuhan yang semakin tinggi, terutama di negara-negara berkembang seperti Indonesia, dan hal ini mengakibatkan penurunan produksi bahan pangan. Beras adalah makanan utama bagi mayoritas penduduk Indonesia, dan setiap tahunnya, konsumsi beras di negara ini terus bertambah sejalan dengan pertumbuhan populasi Indonesia. [1] Wilayah Sumatera Utara merupakan salah satu produsen beras terbesar di Indonesia. Meskipun Sumatera Utara memiliki potensi yang besar dalam produksi beras, masih terdapat tantangan yang perlu diatasi untuk meningkatkan hasil produksi dan efisiensi dalam sektor tersebut di wilayah ini.

Received Oktober 13, 2023; Revised November 02, 2023; Accepted Desember 07, 2023

* Rizki Risdah Sitorus, rizkirisdahs@mhs.unimed.ac.id

Penerapan teknik clustering, seperti K-Means Clustering, telah terbukti sukses di berbagai sektor. K-Means merupakan metode pengelompokan data non-hirarki yang berupaya membagi data ke dalam satu atau beberapa cluster. Fokusnya adalah menyusun data yang memiliki atribut serupa ke dalam kelompok yang sama, sementara data dengan karakteristik yang berbeda ditempatkan dalam kelompok yang berbeda. [2]

Pentingnya beras sebagai komoditas utama dalam konteks Indonesia tidak bisa dilewatkan, karena beras adalah sumber utama karbohidrat bagi masyarakat. Kestabilan produksi beras merupakan isu utama yang perlu diperhatikan oleh pemerintah, karena selain berperan sebagai bahan pokok, beras juga memiliki dampak signifikan pada sektor ekonomi negara dan memberikan lapangan kerja yang besar, terutama bagi jutaan rumah tangga petani di wilayah pedesaan. [3]

Peningkatan jumlah beras yang diimpor ke Indonesia mencerminkan adanya penurunan dalam aspek ekonomi negara ini sehubungan dengan ketersediaan pangan. Dari situ, dapat disimpulkan bahwa ketahanan pangan di Indonesia mengalami penurunan yang signifikan karena terbatasnya lahan pertanian dan produksi pangan yang tidak mampu mengakomodasi pertumbuhan penduduk yang semakin tinggi. [4] Dalam laporan dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara, terlihat bahwa produksi beras di wilayah ini terus mengalami penurunan setiap tahunnya, yang berkorelasi dengan lokasi produksi. Permintaan akan beras melebihi jumlah produksi, yang mengakibatkan sebagian masyarakat mengimpor beras dari provinsi lain, seperti Jawa Tengah.

Berdasarkan riset sebelumnya yang dilakukan oleh Agung, temuan menunjukkan bahwa pemanfaatan algoritma K-Means Clustering dapat menjadi rekomendasi yang bermanfaat dalam menganalisis isu-isu penjualan. Hal ini berpotensi menghasilkan data performa yang lebih akurat. Sebuah penelitian sebelumnya yang memfokuskan pada penentuan judul skripsi juga mencoba menggunakan algoritma K-Means Clustering sebagai solusi bagi permasalahan tersebut. Temuan dari penelitian tersebut menyiratkan bahwa mahasiswa dapat dengan lebih mudah dan cepat menemukan ide judul skripsi yang sesuai dengan kemampuan analisis mereka.

LANDASAN TEORI

Data Mining

Data mining, sebagai salah satu disiplin ilmu komputer yang luas, telah menjadi solusi umum untuk berbagai penelitian. Beragam algoritma, seperti K-means, K-Medoids, C4.5, dan Naïve Bayes Linear, digunakan dalam praktik data mining. Dalam konteks data mining, istilah seperti pattern recognition mengacu pada kemampuan data mining untuk mengenali pola yang tersembunyi dalam dataset. Selain itu, konsep Knowledge Discovery mencerminkan kemampuan data mining dalam menemukan pengetahuan yang tersembunyi di dalam kumpulan data, yang pada akhirnya dapat memberikan wawasan mendalam. [5] Secara umum, pengertian data mining mencakup proses identifikasi dan pengenalan data yang memiliki potensi tinggi, serta penggunaannya yang efektif dalam pengelolaan dataset besar. [6]

K-Means

K-Means, sebagai algoritma populer dalam pengelompokan data berdasarkan karakteristik serupa, menciptakan kelompok data yang dikenal sebagai Cluster untuk mengelompokkan objek berdasarkan ciri-cirinya. Algoritma ini terbatas pada atribut numerik karena memanfaatkan basis jarak untuk membagi data ke dalam sejumlah cluster tertentu. [7]

Clustering

Clustering adalah proses pengelompokan data ke dalam kelompok-kelompok (cluster) sedemikian rupa sehingga objek-objek yang serupa atau mirip ditempatkan dalam satu kelompok, sementara objek-objek yang berbeda atau tidak mirip ditempatkan dalam kelompok yang berbeda. Dalam setiap cluster, data-data diharapkan memiliki kesamaan atau kemiripan yang tinggi, dan hal ini sering diukur dengan menggunakan konsep jarak. Dalam konteks clustering, tujuan utamanya adalah untuk membuat jarak antar objek dalam satu cluster seminim mungkin, sehingga objek-objek dalam satu cluster memiliki kesamaan yang tinggi. Sebaliknya, jarak antar cluster diusahakan sejauh mungkin, yang berarti bahwa cluster-cluster yang berbeda harus memiliki perbedaan yang signifikan. Penjelasan ini juga menggarisbawahi bahwa dalam clustering terdapat parameter-parameter penting yang digunakan untuk mengukur dan merepresentasikan kesamaan ataupun perbedaan antar cluster, dan ini merupakan elemen penting dalam proses pengorganisasian data. [8]

METODE PENELITIAN

Tahapan Penelitian

Proses penelitian ini mengikuti beberapa langkah, dimulai dengan tahap identifikasi data. Pada tahap ini, peneliti menyelidiki literatur terkait untuk memahami dampak produksi beras terhadap ketahanan pangan dan penerapan algoritma K-Means Clustering dalam konteks penelitian sebelumnya, yang tercatat dalam jurnal, buku, dan karya ilmiah lainnya. Langkah ini membentuk landasan pengetahuan sebelum pengumpulan dan analisis data dimulai.

Setelah tahap identifikasi data, peneliti melakukan pengumpulan data secara daring melalui akses ke situs web BPS Sumatera Utara. Data yang diperoleh dari sumber ini menjadi bahan utama untuk analisis dalam penelitian. Selanjutnya, data yang telah terkumpul dianalisis sesuai dengan kebutuhan dan referensi yang telah ditetapkan, termasuk penerapan algoritma K-Means Clustering untuk mengelompokkan data dalam pola yang relevan dengan tujuan penelitian.

Data yang telah dikumpulkan kemudian diolah menggunakan algoritma K-Means Clustering, dan proses ini dilakukan melalui platform Google Colab sesuai dengan alur kerjanya. Algoritma ini membantu mengidentifikasi kelompok data yang memperlihatkan karakteristik serupa, khususnya dalam konteks produksi beras.

Langkah terakhir melibatkan evaluasi hasil pengujian data. Pada tahap ini, peneliti menganalisis hasil pengolahan data dan menyimpulkan temuan dari penelitian. Ini mencakup penarikan kesimpulan mengenai dampak produksi beras terhadap ketahanan pangan serta signifikansi penggunaan algoritma K-Means Clustering dalam konteks wilayah Sumatera Utara.

Data

Sumber data yang digunakan pada penelitian ini diambil melalui website resmi Badan Pusat Statistik (BPS) Sumatera Utara dengan data olah sebanyak 34 data. Website resmi BPS adalah situs web yang dikelola oleh Badan Pusat Statistik Indonesia yang bertanggung jawab untuk mengumpulkan, mengolah, menganalisis, dan menyediakan data statistik resmi tentang berbagai aspek kehidupan di Indonesia. Website resmi BPS merupakan sumber data resmi yang dapat diakses oleh publik dan digunakan oleh pemerintah, peneliti, akademisi, dan masyarakat umum. Pada penelitian ini data yang dipakai adalah jumlah hasil produksi beras di wilayah Sumatera Utara rentang tahun 2020-2022.

HASIL DAN PEMBAHASAN

Analisa Data

Pada tahap ini akan mengolah data yang didapat dari BPS SUMATERA UTARA dari tahun 2020 sampai 2022. Data akan diproses menggunakan penelitian yang ditunjukkan dalam Tabel 1.

Tabel 1. Daftar Produksi Beras di Wilayah Sumatera Utara

<i>Kabupaten/Kota Regency/Municipality</i>	Produksi Padi - Produksi Beras (ton)		
	2020	2021	2022
Nias	20 194,50	20 971,35	18 999,80x
Mandailing Natal	42 437,74	41 485,91	50 646,41x
Tapanuli Selatan	51 848,99	50 454,43	59 269,85x
Tapanuli Tengah	23 928,40	30 380,05	27 274,28x
Tapanuli Utara	62 913,44	70 299,18	79 057,19x
Toba Samosir	60 586,14	61 514,08	61 456,84x
Labuhan Batu	33 208,82	34 156,79	47 977,79x
Asahan	35 010,19	31 927,93	36 015,45x
Simalungun	99 754,03	93 161,88	85 202,71x
Dairi	20 150,90	17 425,35	22 207,17x
Karo	33 007,88	40 054,88	39 613,06x
Deli Serdang	179 847,64	185 339,80	188 636,45x
Langkat	79 795,29	71 761,12	63 337,17x
Nias Selatan	26 365,98	32 978,64	26 949,97x
Humbang Hasundutan	32 179,44	28 401,62	43 286,30x
Pakpak Bharat	2 125,35	2 567,96	3 250,48x
Samosir	21 173,46	24 314,91	23 700,96x
Serdang Bedagai	169 684,38	155 031,76	166 313,18x
Batu Bara	42 194,07	39 683,46	40 755,76x
Padang Lawas Utara	19 683,16	16 786,03	16 172,30x
Padang Lawas	16 672,67	15 102,61	15 748,55x
Labuhan Batu Selatan	356,32	284,64	235,67x
Labuhan Batu Utara	32 028,68	22 008,10	46 006,17x
Nias Utara	18 105,25	22 964,63	23 191,26x
Nias Barat	6 020,31	6 939,58	6 016,16x
Kota Sibolga	–	–	–
Kota Tanjung Balai	243,61	226,03	218,72x
Kota Pematang Siantar	7 221,72	6 921,10	6 694,00x
Kota Tebing Tinggi	1 526,94	1 368,40	1 283,66x
Kota Medan	2 868,31	3 051,06	3 070,46x
Kota Binjai	4 490,89	4 242,43	3 594,47x
Kota Padangsidempuan	11 190,64	10 112,27	10 260,02x
Kota Gunungsitoli	7 619,72	7 690,84	6 319,79x
Sumatera Utara	1 164 434,86	1 149 608,82	1 222 762,05x

Data yang digunakan untuk penelitian ini adalah jumlah produksi beras dimana ada 34 kabupaten/kota di wilayah Sumatera Utara dari tahun 2020 sampai 2022.

Menentukan Centroid

Sebelum memproses kerja algoritma K-Means Clustering, langkah pertama adalah menentukan centroid seperti Gambar 1.

```
#inisialisasi cluster awal
c1_initial = dataframe[dataframe['produksi_beras'] == dataframe['produksi_beras'].max()][['produksi_beras']].mean()
c2_initial = dataframe[dataframe['produksi_beras'] == dataframe['produksi_beras'].min()][['produksi_beras']].mean()

[ ] print('nilai untuk centroid c1 = ', c1_initial)

nilai untuk centroid c1 = 188636.45

[ ] print('nilai untuk centroid c2 = ', c2_initial)

nilai untuk centroid c2 = 218.72
```

Gambar 1. Menentukan centroid

Dari gambar di atas kita dapat menentukan centroid rata – rata tertinggi yang disimbolkan (C1) dan centroid rata - rata terendah yang disimbolkan (C2). Selanjutnya dapat ditentukan kabupaten/kota mana saja yang mempunyai rata – rata centroid tertinggi dan terendah pada Gambar 2 untuk centroid rata – rata tertinggi dan Gambar 3 untuk centroid rata – rata terendah.

	Kabupaten_kota	produksi_beras	tahun	Cluster
0	Nias	20194.50	2020.0	1
1	Nias	20971.35	2021.0	1
2	Nias	18999.80	2022.0	1
3	Mandailing Natal	42437.74	2020.0	1
4	Mandailing Natal	41485.91	2021.0	1
..
94	Padang Sidempuan	10112.27	2021.0	1
95	Padang Sidempuan	10260.02	2022.0	1
96	Padang Sidempuan	11452.45	2020.0	1
97	Padang Sidempuan	11190.64	2021.0	1
98	Padang Sidempuan	10229.92	2022.0	1

[99 rows x 4 columns]

Gambar 2. Centroid rata-rata tertinggi

DataFrame for Cluster 0:

	Kabupaten_kota	produksi_beras	tahun	Cluster
24	Simalungun	99754.03	2020.0	0
33	Deli Serdang	179847.64	2020.0	0
34	Deli Serdang	185339.80	2021.0	0
35	Deli Serdang	188636.45	2022.0	0
51	Serdang Berdagai	169684.38	2020.0	0
52	Serdang Berdagai	155031.76	2021.0	0
53	Serdang Berdagai	166313.18	2022.0	0

Gambar 3. Centroid rata-rata terendah

KESIMPULAN

Kesimpulan dari penelitian ini adalah rentang tahun 2020 sampai 2022 kabupaten/kota Nias, Mandailing Natal, dan Padang Sidempuan memiliki centroid rata-rata tinggi. Sedangkan

pada kabupaten/kota Simalungun, Deli Serdang, dan Serdang Bedagai memiliki centroid rata-rata rendah. Hasil penelitian menunjukkan bahwa metode K-Means Clustering dapat digunakan untuk mengidentifikasi pola produksi beras di wilayah Sumatera Utara. Harapannya, hasil penelitian dapat menjadi kontribusi yang berharga, memberikan rekomendasi, serta mendukung inisiatif bagi pemerintah provinsi Sumatera Utara dalam meningkatkan produksi beras di semua wilayah, dengan tujuan memastikan pasokan makanan yang lebih stabil bagi masyarakat.

DAFTAR REFERENSI

- [1] N. Suryani, R. Abdurrachim and N. Alindah, "Analisis Kandungan Karbohidrat, Serat Dan Indeks Glikemik Pada Hasil Olahan Beras Siam Unus Sebagai Alternatif Makanan Selingan Penderita Diabetes Mellitus," *Jurkessia*, vol. VII, no. 1, pp. 1-9, 2016.
- [2] C. J. Silalahi, A. Situmorang and J. F. Naibaho, "Implementasi Metode K-Means Clustering Untuk Memetakan Daerah Potensial Penghasil Padi di Provinsi Sumatera Utara," *Jurnal Ilmiah Teknik Informatika*, vol. II, no. 2, pp. 49-57, 2022.
- [3] Mardiyah, S., Fajar, M. Y., & Badruzzaman, F. H. (2022). Penggunaan Forecasting dan Goal Programming dalam Optimasi Perencanaan Produksi Beras. *Bandung Conference Series: Mathematics*, 2(1), 83–93.
- [4] Zaeroni, R., & Rustariyuni, S. (2016). Pengaruh Produksi Beras, Konsumsi Beras Dan Cadangan Devisa Terhadap Impor Beras Di Indonesia. *E-Jurnal Ekonomi Pembangunan Universitas Udayana*, 5(9), 993–1010.
- [5] N. A. R. I. & Z. A. Zulfa, "Jurnal Teknologi Dan Sistem Informasi(JTSl)," Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. , vol. 2, pp. 100 - 108, 2021.
- [6] J. *. S. S. Miharja, "Proceeding SENDIU," Penerapan Data Mining Penerimaan Karyawan Menggunakan Metode Naive bayes Classifier, pp. 579-583, 2021.
- [7] K. S. H. K. A. R. P. A. N. O. F. A. & A. Atros, *MAJALAH ILMIAH METHODODA*, 2022.
- [8] T. Tendean and W. Purba, "Analisis Cluster Provinsi Indonesia Berdasarkan Produksi Bahan Pangan Menggunakan Algoritma KMeans," *Jurnal Sains dan Teknologi*, vol. I, no. 2, pp. 5-11, 2020