https://doi.org/10.37817/ikraith-informatika.v11i1
P-ISSN : 2580-4316
E-ISSN : 2654-8054

PENERAPAN ALGORITMA FUZZY C-MEANS UNTUK PENGELOMPOKAN HASIL PERTANIAN DI KECAMATAN SIPAHUTAR

Yusuf R.S. Silalahi[1]
Swingly Purba[2]
Piala Mutiara[3]
Program Studi Teknik Informatika, Fakultas Teknologi Industri, Institut Sains dan Teknologi TD Pardede, Medan, Indonesia
Email:

yusufsilalahi678@gmail.com

swingly.purba61@gmail.com

pialamutiara78@gmail.com

ABSTRAK

Sektor pertanian merupakan salah satu sektor penting di Kecamatan Sipahutar, Kabupaten Tapanuli Utara, tetapi karakteristik produksi antar-komoditas menunjukkan perbedaan yang besar. Perbedaan luas lahan, hasil panen, dan produktivitas menyebabkan proses penentuan prioritas pengembangan tidak cukup dilakukan hanya berdasarkan pengamatan subjektif. Penelitian ini bertujuan menerapkan algoritma Fuzzy C-Means (FCM) untuk mengelompokkan hasil pertanian berdasarkan kemiripan karakteristik dan menyediakan informasi pendukung bagi perencanaan pembinaan pertanian. Data penelitian bersumber dari data hasil pertanian Kecamatan Sipahutar periode 2020–2025, terdiri atas 108 observasi dari 18 komoditas dengan tiga atribut numerik, yaitu luas lahan, hasil panen, dan produktivitas. Tahapan penelitian meliputi pembersihan data, standardisasi Z-Score, pengujian beberapa nilai jumlah cluster, proses iteratif FCM, evaluasi validitas internal, dan interpretasi hasil. Pengujian pada rentang dua sampai tujuh cluster menghasilkan konfigurasi terpilih sebanyak tiga cluster. Pada model ini diperoleh Fuzzy Partition Coefficient sebesar 0,8429, Silhouette Score sebesar 0,5919, Davies-Bouldin Index sebesar 0,4176, Xie-Beni Index sebesar 0,1051, Partition Entropy sebesar 0,2673, dan WCSS sebesar 7,876. Proses komputasi mencapai konvergensi pada sekitar iterasi ke-29 berdasarkan log implementasi utama. Pada tingkat observasi tahunan, komposisi cluster terdiri atas 6 observasi tinggi, 37 sedang, dan 65 rendah. Pada tingkat komoditas dominan, Nenas berada pada cluster tinggi; Jahe, Kunyit, Serai, Laos/Lengkuas, Kubis, Kangkung, dan Labu Siam menjadi komoditas yang perlu dipantau atau berpotensi ditingkatkan; sementara komoditas hortikultura lainnya dominan berada pada cluster rendah. Hasil tersebut menunjukkan bahwa FCM dapat digunakan untuk memetakan heterogenitas hasil pertanian dan menerjemahkannya menjadi rekomendasi pengembangan berbasis data.

Kata kunci: Fuzzy C-Means; clustering; data mining; hasil pertanian; Sipahutar.

ABSTRACT

The agricultural sector is important in Sipahutar District, North Tapanuli Regency, yet agricultural commodities exhibit substantial differences in land area, yield, and productivity. These differences make objective, data-based grouping necessary for agricultural development planning. This study applies the Fuzzy C-Means (FCM) algorithm to cluster agricultural products according to their characteristics and to produce supporting information for agricultural development decisions. The dataset consists of 108 annual observations from 18 commodities recorded during 2020–2025, using land area, harvested production, and productivity as numerical attributes. The workflow includes data cleaning, Z-Score standardization, testing candidate cluster numbers, iterative FCM computation, internal cluster validation, and interpretation. Testing two to seven clusters selected three clusters. The selected model achieved an FPC of 0.8429, a Silhouette Score of 0.5919, a Davies-Bouldin Index of 0.4176, an Xie-Beni Index of 0.1051, a Partition Entropy of 0.2673, and WCSS of 7.876. The main implementation log reached convergence at approximately iteration 29. At the annual-observation level, six observations belonged to the high cluster, 37 to the medium cluster, and 65 to the low cluster. At the dominant-commodity level, Pineapple was

P-ISSN : 2580-4316
E-ISSN : 2654-8054
https://doi.org/10.37817/ikraith-informatika.v11i1

classified as high; Ginger, Turmeric, Lemongrass, Galangal, Cabbage, Water Spinach, and Chayote were identified as commodities requiring monitoring or improvement; and most other horticultural commodities were dominated by the low cluster. These findings demonstrate that FCM can map agricultural heterogeneity and translate clustering results into data-based development recommendations.

Keywords: Fuzzy C-Means; clustering; data mining; agricultural products; Sipahutar District.

1. PENDAHULUAN

Pertanian mempunyai peran strategis dalam mendukung ketahanan pangan, pendapatan masyarakat, dan pembangunan wilayah. Pada tingkat lokal, pengelolaan pertanian membutuhkan informasi yang mampu menjelaskan perbedaan kinerja antar-komoditas secara terukur. Kecamatan Sipahutar di Kabupaten Tapanuli Utara memiliki beragam komoditas pertanian dengan skala lahan, volume panen, dan produktivitas yang tidak seragam. Variasi tersebut dapat dipengaruhi oleh kondisi lahan, pola budidaya, input produksi, cuaca, teknologi, serta kemampuan pengelolaan petani.

Permasalahan yang dihadapi adalah belum tersedianya pengelompokan komoditas yang menggabungkan beberapa indikator pertanian secara objektif. Pengambilan keputusan yang hanya mengandalkan rekomendasi atau pengamatan umum berisiko menghasilkan prioritas bantuan yang kurang tepat. Komoditas yang mempunyai potensi tinggi dapat terlewat, sedangkan komoditas dengan produktivitas rendah belum tentu memperoleh intervensi yang sesuai. Oleh sebab itu, diperlukan pendekatan data mining yang dapat menemukan pola alami dalam dataset hasil pertanian.

Clustering merupakan teknik unsupervised learning yang mengelompokkan objek berdasarkan kemiripan karakteristik. Penelitian ini menggunakan Fuzzy C-Means karena setiap objek tidak selalu mempunyai batas keanggotaan yang tegas. Dalam konteks pertanian, suatu komoditas dapat memiliki karakteristik yang berada di antara dua tingkat produktivitas. FCM merepresentasikan kondisi tersebut melalui derajat keanggotaan, sehingga informasi transisi antarkelompok tetap dapat dianalisis [1].

Penelitian terdahulu telah menerapkan FCM untuk pengelompokan produksi pangan dan hasil panen di berbagai wilayah [2] [3]. Namun, penelitian yang secara khusus memetakan 18 komoditas hasil pertanian Kecamatan Sipahutar menggunakan kombinasi luas lahan, hasil panen, dan produktivitas, kemudian menerjemahkan hasilnya menjadi kategori rekomendasi komoditas, masih terbatas. Kontribusi penelitian ini adalah membangun pemetaan cluster pertanian lokal, mengevaluasi model dengan beberapa indeks validitas, serta menyusun interpretasi prioritas pengembangan yang dapat menjadi masukan bagi pemangku kepentingan.

2. METODOLOGI PENELITIAN

2.1 Data dan Unit Analisis

Data yang dianalisis berasal dari rekapitulasi hasil pertanian Kecamatan Sipahutar periode 2020–2025 sebagaimana digunakan dalam skripsi. Dataset memuat 18 jenis tanaman pada enam tahun pengamatan sehingga menghasilkan 108 baris observasi. Atribut Jenis Tanaman dan Tahun digunakan sebagai identitas, sedangkan atribut yang masuk ke dalam perhitungan clustering adalah Luas Lahan (ha), Hasil Panen (ton), dan Produktivitas (ton/ha). Produktivitas dihitung sebagai perbandingan hasil panen terhadap luas lahan sesuai data penelitian.

Komponen	Keterangan
Unit analisis	Observasi komoditas per tahun
Periode	2020–2025
Jumlah komoditas	18 jenis tanaman
Jumlah observasi	108 baris
Fitur clustering	Luas lahan, hasil panen, produktivitas
Sumber data	BPS Kabupaten Tapanuli Utara dan rekapitulasi penelitian

2.2 Preprocessing dan Standardisasi

Preprocessing dilakukan melalui fungsi load_and_clean(). Tahap ini mencakup konversi format angka Indonesia menjadi bilangan numerik, penghapusan spasi berlebih pada label komoditas, pemeriksaan baris kosong atau nilai null, dan verifikasi jumlah observasi. Setelah data bersih, fitur numerik distandardisasi dengan Z-Score agar perbedaan skala tidak menyebabkan hasil panen yang bernilai besar mendominasi jarak Euclidean. Standardisasi dilakukan dengan persamaan $z=(x-\mu)/\sigma$, dengan x sebagai nilai

https://doi.org/10.37817/ikraith-informatika.v11i1
P-ISSN : 2580-4316
E-ISSN : 2654-8054

asli, μ sebagai rata-rata fitur, dan σ sebagai simpangan baku.

2.3 Algoritma Fuzzy C-Means

FCM meminimalkan fungsi objektif berbobot melalui pembaruan pusat cluster dan matriks keanggotaan secara bergantian [1]. Parameter utama yang digunakan adalah jumlah cluster c=3, fuzzifier m=2, toleransi error 0,005, maksimum iterasi 1.000, dan random seed 42. Matriks keanggotaan awal dibangkitkan secara acak tetapi dapat direproduksi karena seed ditetapkan.

Pusat cluster dihitung dengan persamaan v_j = Σ_i(u_ij^m x_i) / Σ_i(u_ij^m). Selanjutnya, derajat keanggotaan diperbarui berdasarkan jarak Euclidean data terhadap pusat cluster. Iterasi dihentikan apabila perubahan fungsi objektif atau keanggotaan berada di bawah toleransi, atau apabila jumlah iterasi mencapai batas maksimum. Label Rendah, Sedang, dan Tinggi diberikan setelah cluster diurutkan berdasarkan rata-rata produktivitasnya; dengan demikian, nomor cluster algoritmik tidak diperlakukan sebagai urutan tingkat kinerja.

2.4 Penentuan Jumlah Cluster dan Validasi

Jumlah cluster diuji pada c=2 sampai c=7. Model dievaluasi menggunakan FPC, Silhouette Score, Davies-Bouldin Index (DBI), Xie-Beni Index (XB), Partition Entropy (PE), dan WCSS/Elbow. FPC dan Silhouette yang lebih tinggi menunjukkan partisi yang lebih tegas dan terpisah, sedangkan DBI, XB, PE, dan WCSS yang lebih rendah menunjukkan struktur yang lebih kompak dan tidak terlalu tumpang tindih. Pemilihan akhir mempertimbangkan gabungan hasil indeks dan kebutuhan interpretasi tiga tingkat potensi pertanian.

3. HASIL DAN PEMBAHASAN

3.1 Statistik Deskriptif Dataset

Dataset memiliki variasi yang sangat besar. Rata-rata luas lahan adalah 108,29 ha, sedangkan median 10,02 ha dan maksimum 1.772,61 ha. Rata-rata hasil panen mencapai 8.864,93 ton, median 151,34 ton, dan maksimum 170.171,24 ton. Produktivitas mempunyai rata-rata 18,54 ton/ha, median 13,37 ton/ha, serta rentang 0–96 ton/ha. Selisih mean dan median pada luas lahan dan hasil panen menunjukkan kemencengan ke kanan yang kuat, terutama karena karakteristik Nenas yang menjadi outlier dominan dalam skala lahan dan produksi.

Variabel	Mean	Median	Minimum	Maksimum	Simpangan baku
Luas lahan (ha)	108,29	10,02	0,00	1.772,61	372,02
Hasil panen (ton)	8.864,93	151,34	0,00	170.171,24	35.995,26
Produktivitas (ton/ha)	18,54	13,37	0,00	96,00	20,90

Kondisi tersebut memperkuat alasan penggunaan standardisasi. Tanpa standardisasi, fitur hasil panen akan memiliki kontribusi jarak yang jauh lebih besar dibandingkan produktivitas. Nilai nol pada beberapa observasi juga dipertahankan sebagai informasi kondisi aktual, misalnya komoditas yang tidak ditanam atau tidak menghasilkan panen pada tahun tertentu.

3.2 Evaluasi Jumlah Cluster

	FPC	Silhouette	DBI	XB	PE	WCSS
	0,9926	0,9312	0,0834	0,0036	0,0215	21,0023
terpilih	0,8429	0,5919	0,4176	0,1051	0,2673	78,8760
	0,7641	0,5227	0,4727	0,1404	0,4281	48,8354
	0,7400	0,5376	0,5314	0,1711	0,4977	33,3906
	0,7370	0,5338	0,4565	0,1335	0,5160	25,5303
	0,7283	0,5113	0,5443	0,1057	0,5313	21,1553

Secara matematis, c=2 menghasilkan nilai indeks internal paling baik pada beberapa metrik. Namun, dua cluster hanya membedakan kelompok rendah dan tinggi sehingga tidak merepresentasikan kebutuhan praktis untuk membedakan komoditas tingkat sedang. Pada c=3, seluruh metrik masih menunjukkan struktur yang layak, titik siku WCSS berada pada c=3, dan hasilnya sesuai dengan kerangka kebijakan yang membedakan kategori rendah, sedang, dan tinggi. Oleh karena itu, c=3 dipilih sebagai konfigurasi yang paling informatif secara substantif, bukan semata-mata berdasarkan satu indeks.

3.3 Hasil Pengelompokan

P-ISSN : 2580-4316
E-ISSN : 2654-8054
https://doi.org/10.37817/ikraith-informatika.v11i1

Penerapan FCM pada 108 observasi menghasilkan tiga kelompok. Berdasarkan pengurutan rata-rata produktivitas, cluster tinggi ditandai oleh produktivitas sangat tinggi dan skala lahan besar; cluster sedang ditandai produktivitas menengah dengan skala lahan relatif lebih kecil; sedangkan cluster rendah ditandai produktivitas yang lebih rendah dan mencakup sebagian besar komoditas hortikultura.

Kategori	Jumlah observasi tahunan	Proporsi	Interpretasi umum
Tinggi	6	5,56%	Didominasi Nenas; produktivitas dan skala produksi sangat tinggi
Sedang	37	34,26 %	Komoditas dengan produktivitas menengah dan peluang peningkatan
Rendah	65	60,19 %	Komoditas yang memerlukan pembinaan dan intensifikasi

Nenas konsisten berada pada cluster tinggi dari 2020 sampai 2025, dengan produktivitas 96 ton/ha dan luas lahan tahunan sekitar 1.529,48–1.772,61 ha. Pada cluster sedang, Jahe, Kunyit, Serai, dan Laos/Lengkuas menunjukkan produktivitas yang relatif lebih baik dibandingkan sebagian besar sayuran. Kubis, Kangkung, dan Labu Siam pada beberapa observasi berada pada kelompok sedang atau menunjukkan perbaikan, sehingga lebih tepat diposisikan sebagai komoditas yang perlu dipantau daripada langsung dikategorikan unggulan.

Sebagian besar observasi cluster rendah berasal dari Cabai Keriting, Cabai Rawit, Bayam, Bawang Daun, Bawang Merah, Buncis, Kacang Panjang, Ketimun, Tomat, dan Kentang. Hasil ini tidak berarti komoditas tersebut tidak bernilai secara ekonomi, melainkan menunjukkan bahwa kombinasi indikator luas lahan, hasil panen, dan produktivitasnya berada di bawah pola cluster tinggi dalam dataset penelitian.

3.4 Analisis Derajat Keanggotaan dan Rekomendasi

Keunggulan FCM terlihat pada objek yang berada di dekat batas cluster. Buncis, Bawang Merah, Kacang Panjang, Ketimun, dan Tomat memiliki derajat keanggotaan yang menunjukkan posisi transisi pada sebagian tahun. Informasi ini penting karena komoditas tersebut tidak seharusnya diperlakukan sama dengan komoditas yang sangat tegas berada di cluster rendah. Derajat keanggotaan dapat digunakan sebagai dasar verifikasi lapangan dan penentuan prioritas intervensi.

Komoditas	Cluster dominan	Rata-rata produktivitas (ton/ha)	Keanggotaan rata-rata	Tren	Rekomendasi
Nenas	Tinggi	96,00	sangat tinggi	Stabil	Layak dikembangkan
Jahe	Sedang	21,05	0,9153	Meningkat	Perlu pemantauan
Kunyit	Sedang	24,73	0,9509	Stabil	Perlu pemantauan
Serai	Sedang	27,06	0,8833	Meningkat	Perlu pemantauan
Laos/Lengkuas	Sedang	18,93	0,8227	Stabil	Perlu pemantauan
Kubis	Rendah	4,76	0,9085	Meningkat	Perlu pemantauan
Cabai Keriting	Rendah	5,22	0,9225	Stabil	Perlu perhatian khusus
Cabai Rawit	Rendah	6,16	0,9628	Menurun	Perlu perhatian khusus
Tomat	Rendah	15,60	0,8334	Menurun	Perlu perhatian khusus
Kentang	Rendah	17,03	0,7902	Menurun	Perlu perhatian khusus

Kategori rekomendasi dalam tabel tersebut bersifat pendukung, bukan keputusan final. Untuk Dinas Pertanian, komoditas berkeanggotaan borderline dapat menjadi sasaran verifikasi awal karena relatif berpeluang merespons intervensi. Pola Nenas dapat dipelajari sebagai praktik usaha tani unggulan, tetapi tidak dapat langsung digeneralisasi kepada komoditas lain karena perbedaan skala usaha dan karakteristik agroklimat. Komoditas rendah dengan keanggotaan sangat tegas memerlukan intensifikasi, bibit unggul, perbaikan teknik budidaya, pengendalian hama dan penyakit, serta pendampingan penyuluh.

3.5 Pembahasan Kualitas dan Keterbatasan Model

Nilai Silhouette 0,5919 menunjukkan pemisahan cluster yang cukup baik, sedangkan FPC 0,8429 menunjukkan keanggotaan yang relatif tegas pada model terpilih. Nilai DBI dan XB yang lebih rendah dibanding konfigurasi cluster lebih banyak mendukung adanya kekompakan yang masih memadai. Walaupun demikian, hasil model sangat dipengaruhi oleh outlier Nenas. Karena itu, kesimpulan tentang cluster rendah dan sedang perlu dibaca sebagai pola relatif terhadap dataset yang digunakan, bukan sebagai ukuran absolut kesejahteraan atau nilai ekonomi komoditas.

Penelitian ini juga memiliki keterbatasan pada jumlah dan jenis fitur. Model belum memasukkan curah hujan, jenis tanah, irigasi, modal, harga, biaya produksi, dan faktor manajerial. Selain itu, analisis utama memperlakukan observasi tahunan sebagai titik data, sehingga dinamika deret waktu belum menjadi komponen eksplisit dalam pembentukan cluster. Validasi eksternal melalui pakar pertanian dan perbandingan dengan K-Means atau hierarchical clustering juga belum dilakukan.

4. KESIMPULAN

Algoritma Fuzzy C-Means dapat diterapkan untuk mengelompokkan hasil pertanian di Kecamatan Sipahutar menggunakan luas lahan, hasil panen, dan produktivitas setelah preprocessing serta standardisasi Z-Score. Pengujian jumlah cluster c=2 sampai c=7 memilih tiga cluster yang secara substantif dapat diberi label Tinggi, Sedang, dan Rendah. Model terpilih menghasilkan FPC 0,8429, Silhouette Score 0,5919, DBI 0,4176, XB 0,1051, PE 0,2673, dan WCSS 7,8760, dengan konvergensi pada sekitar iterasi ke-29 pada implementasi utama.

Pada tingkat observasi tahunan, cluster rendah berjumlah 65 observasi, cluster sedang 37 observasi, dan cluster tinggi 6 observasi. Nenas menjadi komoditas unggulan yang konsisten berada pada cluster tinggi. Jahe, Kunyit, Serai, Laos/Lengkuas, Kubis, Kangkung, dan Labu Siam merupakan komoditas yang perlu dipantau atau dikembangkan secara bertahap. Komoditas seperti Cabai, Bawang, Bayam, Buncis, Kacang Panjang, Ketimun, Tomat, dan Kentang memerlukan perhatian khusus karena dominan berada pada cluster rendah.

Derajat keanggotaan fuzzy memperkaya hasil dengan menunjukkan komoditas yang berada pada posisi transisi. Dengan demikian, hasil clustering dapat digunakan sebagai informasi pendukung penentuan prioritas bantuan, tetapi keputusan akhir tetap perlu mempertimbangkan verifikasi lapangan, kondisi sosial-ekonomi petani, dan faktor agroklimat.

5. SARAN

Dinas Pertanian Kabupaten Tapanuli Utara disarankan menggunakan hasil cluster sebagai masukan kuantitatif dalam menyusun prioritas bantuan dan menjadikan komoditas borderline sebagai sasaran verifikasi lapangan awal. Pemerintah daerah dapat mengkaji praktik usaha tani Nenas secara kualitatif sebelum menjadikannya rujukan bagi komoditas lain. Petani pada cluster rendah perlu memperoleh dukungan berupa bibit unggul, intensifikasi lahan, pengendalian organisme pengganggu tanaman, dan pendampingan teknis. Penelitian berikutnya perlu menambahkan variabel agroklimat dan manajerial, melakukan analisis sensitivitas dengan mengeksklusi outlier Nenas, memasukkan tren tahunan atau pendekatan time-series clustering, membandingkan FCM dengan algoritma lain, serta melibatkan validasi eksternal dari penyuluh pertanian.

6. UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Institut Sains dan Teknologi TD Pardede, dosen pembimbing, keluarga, serta pihak-pihak yang telah mendukung penyediaan data dan penyelesaian penelitian ini.

DAFTAR PUSTAKA

[1] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981.

[2] I. M. Nur, A. N. L. Syifa, M. Kharis, and S. H. Permatasari, “Implementasi metode Fuzzy C-Means dalam pengelompokkan hasil panen padi di Provinsi Bali,” *VARIANCE: Journal of Statistics and Its Applications*, vol. 5, no. 1, pp. 13–24, 2023. https://doi.org/10.30598/VARIANCEVOL5ISS1PAGE13-24

[3] A. Irianti and W. Firgiawan, “Pengelompokan kecamatan berdasarkan produktifitas hasil perkebunan di Kabupaten

Polewali Mandar menggunakan metode Fuzzy C-Means," *Journal of Computer and Information System (J-CIS)*, vol. 6, no. 2, pp. 40–50, 2023.

[4] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979. https://doi.org/10.1109/TPAMI.1979.4766909

[5] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973. https://doi.org/10.1080/01969727308546046

[6] J. Han, J. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2012.

[7] Badan Pusat Statistik Kabupaten Tapanuli Utara, *Kecamatan Sipahutar Dalam Angka 2025*. BPS Kabupaten Tapanuli Utara, 2025.