

CLUSTERING DATA JENIS WINE DENGAN METODE K-MEANS DAN C-MEANS

Rafael Arrelano

Program Studi Teknik Informatika, Teknik Informasi, Universitas Tarumanagara
Jl. Letjen S.Parman No.1, Jakarta Barat, DKI Jakarta, Indonesia 11410
e-mail: *rafael.arrelano@gmail.com*

ABSTRAK

Penelitian ini menggunakan teknik *clustering* untuk mengeksplorasi dan mengungkap pola-pola inheren dalam sebuah dataset yang berisi informasi tentang berbagai jenis *wine*. Dataset ini mencakup beragam atribut, termasuk properti kimia seperti '*Malic Acid*,' memberikan pandangan yang jelas tentang karakteristik *wine*. Dalam penelitian *clustering* ini, menggunakan 2 metode utama yaitu *K-Means* dan *C-Means* untuk menemukan *silhouette score* dari metode tersebut.

Kata kunci: *Clustering, Wine, Dataset, K-Means, C-Means.*

ABSTRACT

This research uses clustering techniques to explore and reveal inherent patterns in a dataset containing information about various types of wine. This dataset includes a wide range of attributes, including chemical properties such as 'Malic Acid,' providing a clear view of wine characteristics. In this clustering research, 2 main methods are used, namely K-Means and C-Means to find the silhouette score from this method.

Keywords: *Clustering, Wine, Dataset, K-Means, C-Means.*

1. PENDAHULUAN

Industri *wine*, sebagai bagian integral dari budaya dan industri minuman, terus berkembang pesat dengan munculnya berbagai jenis dan varietas *wine*. Keanekaragaman karakteristik yang terkandung dalam setiap jenis *wine* menciptakan tantangan unik dalam memahami dan mengklasifikasikan berbagai produk. Dalam konteks ini, *Clustering* data menjadi alat yang esensial untuk mengeksplorasi pola-pola dan hubungan yang mungkin ada di antara berbagai jenis *wine*.

Wine merupakan minuman beralkohol yang terbuat dari hasil fermentasi anaerob jus buah anggur tanpa kehadiran O₂[21]. Keseimbangan sifat alami yang terkandung pada buah anggur dapat menyebabkan buah tersebut difermentasi tanpa penambahan gula, asam, *enzyme*, maupun nutrisi lain. Pembuatan *wine* dengan cara fermentasi jus buah anggur ini menggunakan khamir tertentu yang kemudian kandungan gula yang ada pada buah anggur tersebut akan dikonsumsi oleh yeast (ragi) dan mengubahnya menjadi *alcohol*. Jenis anggur yang berbeda dan strain khamir yang digunakan, tergantung pada jenis *wine* yang akan di produksi[18]. Dalam memproduksi *wine*, komposisi yang digunakan harus mempunyai kandungan nutrisi tinggi, mempunyai keasaman yang tinggi sehingga dapat menghambat pertumbuhan mikroba yang tidak diinginkan, kandungan gula cukup tinggi dan aroma yang sedap, oleh itu kualitas *wine* harus diutamakan. Salah satu cara yang dapat digunakan untuk memprediksi kualitas dari sebuah *wine* dapat digunakan dengan cara klasifikasi data menggunakan data mining. Klasifikasi bertujuan untuk memprediksi kelas dari suatu objek yang belum diketahui sebelumnya dan diukur secara obyektif dan subyektif. Data mining adalah gabungan atau perpaduan model faktual dan *machine learning*

Penelitian ini berfokus pada penerapan teknik *clustering* untuk mengelompokkan *wine* berdasarkan karakteristik tertentu. Kumpulan data yang digunakan mencakup atribut kimia, dengan '*Malic Acid*' sebagai salah satu contoh[3], yang memungkinkan pembentukan kelompok yang lebih baik berdasarkan jenis bahan dasar pembuatan *wine*. Langkah pertama dalam penelitian ini adalah

memuat dan memvisualisasikan dataset *wine* untuk memahami distribusi dan pola yang mungkin terkandung dalam data. Kami kemudian menerapkan Metode *K-means* untuk memberikan pemahaman awal tentang kemungkinan kelompok atau klasifikasi yang dapat diidentifikasi[1]. Selanjutnya, untuk memperluas analisis, mengimplementasikan metode *C-means* untuk *clustering*. Metode ini memungkinkan penyesuaian lebih halus terhadap ketidakpastian inherent dalam data, mempertimbangkan kemungkinan bahwa *wine* tertentu dapat memiliki keanggotaan dalam lebih dari satu kelompok. Hasil dari analisis clustering ini kemudian dievaluasi menggunakan metrik evaluasi, termasuk *silhouette score*, untuk memberikan gambaran kuantitatif tentang efektivitas metode *clustering* yang digunakan[2].

K-Means dan *C-Means* jika dilihat dari beberapa riset sebelumnya yang dilakukan Mustakim dan Ekawati mampu memberikan hasil cluster terbaik. Kedua metode tersebut samasama memiliki hasil yang signifikan serta mempunyai beberapa perbedaan dalam hal pola cluster. Pengujian performansi algoritma adalah salah satu cara untuk menemukan metode mana yang lebih baik yang digunakan dalam *clustering*. Di dalam penelitian ini akan ditunjukkan perbandingan pengklasteran data *user knowledge modeling* menggunakan metode *K-Means* dan *Fuzzy C-Means*. Penelitian ini bertujuan untuk membandingkan dua algoritma *clustering* yaitu *K-Means* dan *Fuzzy C-Means* untuk melihat kemampuan dari setiap algoritma dalam *clustering* sehingga mendapatkan hasil algoritma terbaik. Manfaat dari penelitian ini adalah agar menjadi referensi bagi penelitian selanjutnya dalam pemelihan algoritma *clustering*[12].

2. TINJAUAN LITERATUR

Penelitian ini bertujuan untuk mendalami pemahaman tentang jenis-jenis *wine* melalui penerapan teknik *clustering* pada dataset yang mencakup berbagai atribut, khususnya fokus pada karakteristik kimia seperti '*Malic Acid*'. Pertama, penelitian ini bertujuan untuk mengidentifikasi pola dan hubungan yang mungkin ada di antara data, memungkinkan kita untuk merinci struktur inherent dalam dataset. Selanjutnya, penelitian ini akan mengeksplorasi potensi klasifikasi atau kelompok berdasarkan karakteristik kimia, menyediakan landasan bagi pemahaman lebih mendalam tentang variasi dalam jenis *wine*[4].

Dalam melakukan pra-pemrosesan data pada dataset jenis *wine*, beberapa langkah krusial telah diimplementasikan. Pertama, dilakukan pemilihan fitur, di mana fitur-fitur yang paling relevan untuk analisis clustering dipilih, dengan fokus utama pada atribut[17]. Langkah kedua melibatkan standarisasi data untuk mengatasi potensi masalah skala antar fitur. Oleh karena itu, dilakukan transformasi data menggunakan '*StandardScaler*' dari *scikit-learn*, memastikan bahwa distribusi fitur-fitur tersebut seragam. Setelah itu, dilakukan visualisasi data menggunakan *scatter plot* untuk memberikan gambaran jelas tentang distribusi sebelum proses *clustering*[2]. Langkah keempat melibatkan penanganan missing values dan pembersihan data, dengan mengidentifikasi dan menangani nilai-nilai yang hilang serta *outliers* yang mungkin mempengaruhi hasil *clustering*. Langkah terakhir, namun tidak kalah pentingnya, adalah ekstraksi fitur jika diperlukan untuk meminimalisir kesalahan data atau mengekstrak informasi yang lebih relevan[16]. Dalam konteks ini, tujuan lainnya adalah membandingkan efektivitas algoritma *clustering*, baik *K-means* maupun *C-means* untuk membentuk kelompok - kelompok yang bermakna. Evaluasi menggunakan *matrix* seperti *silhouette score* diharapkan memberikan dasar perhitungan skor yang terbaik.

3. METODE PENELITIAN

Dari hasil penelitian dan pembahasan yang telah dilakukan, diperlukan rancangan sistem beserta metode dan juga sumber data untuk membuat rancangan tersebut yang akan dijelaskan pada sub bab berikut:

3.1 Dataset

Dalam penelitian *clustering* yang dilakukan, diperlukan dataset yang diperoleh melalui *website* www.kaggle.com yang berupa dataset berisikan jenis - jenis kandungan *wine*.

	Alcohol	Malic_Acid	Ash	Ash_Alcanity	Magnesium	Total_Phenols	Flavanoids	Nonflavanoid_Phenols
0	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28
1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26
2	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30
3	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24
4	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39

Gambar 1. Atribut dari dataset *wine*.

Dapat dilihat pada Gambar 1 merupakan contoh dataset *wine* yang berisikan atribut mengenai jenis - jenis kandungan yang terdapat pada *wine* yang terdiri *Alcohol*, *Malic acid*, *Ash*, *Ash Alcanity*, *Magnesium*, *Total phenols*, *Flavanoids*, *Non Flavanoids*.

3.2 K-Means

K-Means Clustering adalah metode partisi yang bertujuan untuk mengelompokkan data ke dalam beberapa kluster. Pada langkah awal, ditentukan jumlah kluster (*k*) yang diinginkan. Algoritma kemudian secara iteratif menetapkan setiap data ke dalam kluster yang memiliki pusat terdekat, dan menghitung kembali pusat kluster tersebut. Proses ini berulang hingga tidak ada perubahan signifikan dalam penempatan data atau mencapai jumlah iterasi maksimum yang ditentukan. Dapat dilihat pada Gambar 2 merupakan kode atau fungsi untuk metode *K-Means* dan dapat dilihat pada Gambar 3 merupakan hasil grafik dari atribut "*Malic acid*" menggunakan metode *K-Means*[5].

Algoritma *K-Means* termasuk dalam kategori algoritma *partitional*, karena prinsip *K-Means* bergantung pada penentuan jumlah awal kelompok dengan menetapkan nilai *centroid* awalnya [14]. Proses *K-Means* dilakukan secara berulang-ulang untuk membentuk dasar data kluster. Input dari algoritma ini adalah jumlah *cluster* awal yang diinginkan, dan outputnya adalah titik *centroid* akhir. Dalam metode *K-Means*, pola *k* dipilih sebagai titik awal *centroid* secara acak. Jumlah iterasi yang diperlukan untuk mencapai *centroid* kluster akan dipengaruhi oleh penentuan awal *cluster centroid* secara acak. Oleh karena itu, pengembangan algoritma dapat mempertimbangkan cara menentukan *centroid* kluster dengan memperhatikan kepadatan data awal yang tinggi, sehingga dapat meningkatkan kinerja algoritma secara keseluruhan [10].

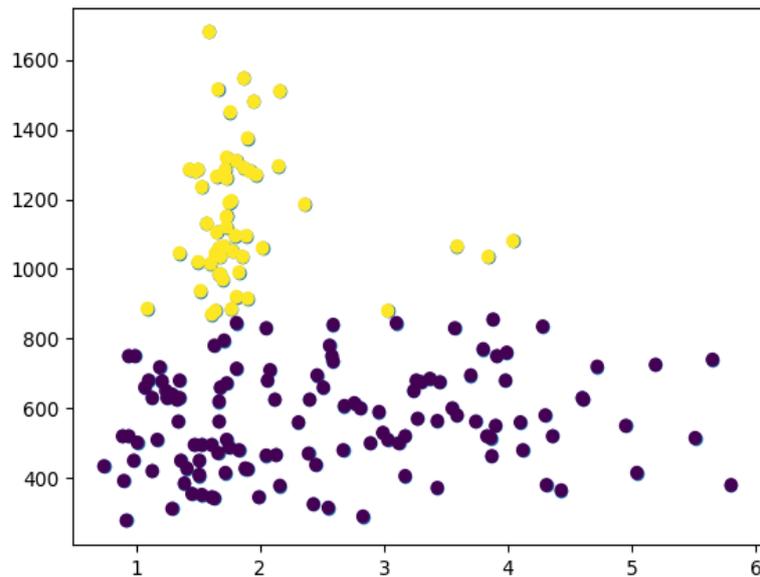
Dalam algoritma *K-Means* setiap data harus termasuk ke dalam *cluster* tertentu pada suatu tahapan proses. Pada awal algoritmanya, *K-Means* mengambil sebagian dari banyaknya komponen dari populasi untuk dijadikan pusat *cluster* awal. Pada tahap ini, pusat *cluster* dipilih secara acak dari sekumpulan populasi data. Berikutnya *K-Means* menguji masing-masing komponen di dalam populasi data dan menandai komponen tersebut ke salah satu pusat *cluster* yang telah didefinisikan tergantung dari jarak minimum antar komponen pada setiap pusat *cluster*. Posisi pusat *cluster* dihitung kembali hingga semua komponen data digolongkan dalam tiap-tiap *cluster* dan posisi pusat *cluster* akan dihitung kembali hingga seluruh komponen data digolongkan ke dalam tiap-tiap *cluster* dan terakhir akan membentuk posisi *cluster* baru.

```
plt.scatter(x_axis, y_axis)

kmeans = KMeans(n_clusters=2, init='random')
kmeans.fit(df)

pred = kmeans.predict(df)
pred
```

Gambar 2. *K- Means*



Gambar 3. Grafik Atribut *Malic Acid* dengan Metode *K-Means*

3.3 *C-Means*

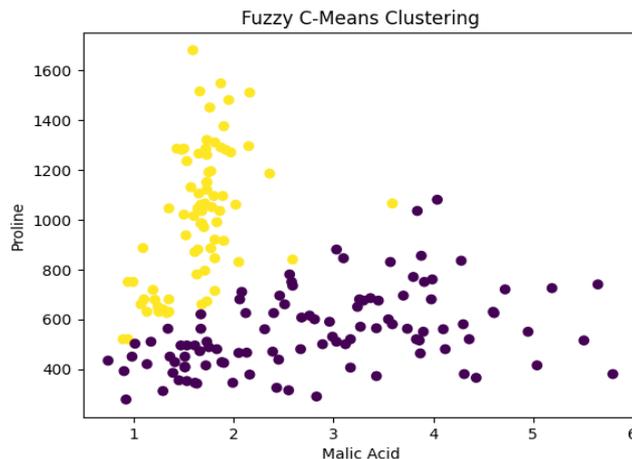
C-Means adalah varian dari *K-Means* yang memungkinkan setiap data memiliki derajat keanggotaan dalam semua kluster dengan nilai antara 0 dan 1. Algoritma *FCM* memperhitungkan ketidakpastian terhadap keanggotaan suatu data ke dalam kluster tertentu. Pada langkah pertama, derajat keanggotaan awal diinisialisasi secara acak[6]. Kemudian, pusat kluster dihitung dengan memperhitungkan derajat keanggotaan dan data ditempatkan dalam kluster yang memiliki pusat terdekat. Proses ini diulang hingga pusat kluster dan derajat keanggotaan. Dalam implementasi di atas, *C-Means clustering* menggunakan fungsi `fuzz.cluster.cmeans` dari library *scikit-fuzzy*. Data di-transformasi dan diproses melalui algoritma *C-Means*, dan hasilnya divisualisasikan dalam scatter plot bersamaan dengan data hasil *K-Means clustering*. Dapat dilihat pada Gambar 4 merupakan kode atau fungsi untuk menggunakan metode *C-Means* pada *clustering* dan dapat dilihat pada Gambar 5 merupakan hasil grafik atribut “*Malic acid*” dengan metode *C-Means*.

Logika *fuzzy* pertama kali diperkenalkan oleh Prof. Lotfi A. Zadeh pada tahun 1965. Dalam banyak hal, logika *fuzzy* digunakan sebagai suatu cara untuk memetakan permasalahan dari input menuju ke output yang diharapkan. Dalam logika *fuzzy* terdapat *fuzzy clustering* yang merupakan salah satu metode untuk menentukan kluster optimal dalam suatu ruang vektor yang didasarkan pada bentuk normal *Euclidian* untuk jarak antar vektor[15]. Dalam logika *fuzzy* terdapat metode yang sering digunakan untuk mengkluster data, yaitu metode *fuzzy c-means*. *Fuzzy c-means* adalah suatu metode pengklasteran data yang ditentukan oleh derajat keanggotaan.

```
scaler = StandardScaler()
df_scaled = scaler.fit_transform(df[['Malic_Acid', 'Proline']])

n_clusters = 2
cntr, u, u0, d, jm, p, fpc = fuzz.cluster.cmeans(
    df_scaled.T, n_clusters, 2, error=0.005, maxiter=1000, init=None
)
```

Gambar 4. *C-Means*



Gambar 5. Grafik Atribut *Malic Acid* dengan Metode *C-Means*

3.3 Skema Eksperimen

Skema eksperimen yang direncanakan untuk *clustering* pada dataset jenis *wine* melibatkan beberapa tahap utama. Pertama, dilakukan pra-pemrosesan data, di mana fitur-fitur yang paling relevan dipilih, dan data disesuaikan melalui standarisasi menggunakan `StandardScaler`. Visualisasi data awal dilakukan melalui scatter plot untuk memahami distribusi sebelum clustering. Langkah berikutnya adalah menerapkan dua metode clustering, yaitu *K-Means* dan *C-Means*, untuk membandingkan pendekatan eksklusif dan *fuzzy* dalam pembentukan kelompok. *K-Means* digunakan dengan jumlah kluster (*n-clusters*) sebanyak 2, sementara *C-Means* menggunakan fungsi `fuzz.cluster.cmeans` untuk memperoleh kelompok *fuzzy*.

Setelah kluster-kluster dibentuk, dilakukan evaluasi kualitas *clustering* menggunakan *silhouette score matrix* untuk kedua metode. *Silhouette score* memberikan gambaran tentang sejauh mana kelompok-kelompok tersebut terpisah dan konsisten. Evaluasi ini memberikan pemahaman lebih mendalam tentang performa dan karakteristik masing-masing metode *clustering* dalam konteks dataset *wine*. Penting untuk dicatat bahwa eksperimen ini tidak hanya bertujuan untuk mengelompokkan data, tetapi juga untuk memberikan wawasan baru terkait dengan jenis-jenis *wine* berdasarkan karakteristik tertentu. Hasil dari eksperimen ini diharapkan dapat memberikan kontribusi pada pemahaman industri *wine* serta memberikan dasar untuk analisis lanjutan terkait dengan preferensi konsumen dan pengembangan strategi bisnis.

3.4 Data Mining

Data mining menjadi semakin penting dalam era informasi saat ini, di mana jumlah data yang tersedia terus meningkat secara signifikan. Sumber data yang dapat dimanfaatkan untuk penambangan data mencakup beragam platform, seperti *database*, gudang data, web, repositori informasi lainnya, dan aliran data yang masuk ke dalam sistem dinamis [14]. Dalam konteks yang berkaitan dengan penambangan data, metode yang sering digunakan untuk mengelompokkan data adalah metode *K-Means*.

Metode *K-Means* merupakan sebuah algoritma yang memerlukan sejumlah parameter input, *k*, di mana algoritma ini bertujuan untuk membagi sekumpulan *n* objek ke dalam *k* *cluster*. Tujuan dari pembagian ini adalah untuk mencapai tingkat kemiripan yang tinggi antar anggota dalam satu *cluster*, dengan tingkat kemiripan yang sangat rendah terhadap anggota cluster lainnya. Proses pengukuran kemiripan anggota terhadap suatu cluster dilakukan dengan mengukur kedekatan objek terhadap nilai rata-rata pada *cluster*, yang dikenal sebagai *centroid cluster* [9]. Metode *K-Means* dikenal sebagai salah satu metode pengelompokan yang paling sederhana dan umum digunakan dalam dunia penambangan data. Keunggulan utama *K-Means* terletak pada kemampuannya untuk mengelompokkan data dalam jumlah besar dengan waktu komputasi yang relatif cepat dan efisien [10].

Meskipun demikian, seperti halnya dengan kebanyakan metode, *K-Means* memiliki kelemahan tertentu. Salah satu kelemahan utamanya adalah dalam menganalisis persebaran data, di mana metode ini cenderung kurang efektif dalam mengatasi distribusi data yang kompleks. Selain itu, *K-Means* sangat bergantung pada inisialisasi *centroid*, yang dapat mempengaruhi hasil akhir *clustering*. Metode ini hanya mempertimbangkan jarak data ke masing-masing *centroid* pada setiap *cluster*, sehingga keberhasilannya dapat dipengaruhi oleh pemilihan awal *centroid* yang optimal [11]. Meskipun demikian, kelebihan dan kelemahan tersebut menjadi poin penting yang perlu dipertimbangkan dalam penerapan *K-Means*, terutama dalam konteks penambahan data yang semakin berkembang.

4. HASIL DAN PEMBAHASAN

Hasil pembahasan dari eksperimen clustering pada dataset jenis *wine* memberikan wawasan yang sangat berharga terkait dengan struktur internal dan potensi klasifikasi berdasarkan karakteristik tertentu. Penggunaan algoritma *K-Means* dalam eksperimen ini menghasilkan pembagian data yang sangat terperinci ke dalam dua kluster eksklusif, memberikan kejelasan dalam identifikasi kelompok yang jelas dan terpisah. Sebaliknya, penerapan algoritma *C-Means* memberikan gambaran yang lebih kompleks dan relevan, dengan setiap data memiliki tingkat keanggotaan dalam setiap kluster [18]. Visualisasi hasil *clustering* secara grafis menunjukkan bahwa algoritma *K-Means* mampu menghasilkan pemisahan yang cukup jelas antara dua kelompok, sementara algoritma *C-Means* memberikan kelompok yang lebih tumpang tindih, mencerminkan tingkat ketidakpastian yang mungkin terkait dengan klasifikasi [12].

Evaluasi menggunakan *silhouette score* sebagai parameter penilaian menyokong temuan ini, di mana skor *siluet K-Means* cenderung lebih tinggi, menandakan tingkat pemisahan yang lebih baik antara kluster - kluster. Temuan ini menyiratkan bahwa, dalam konteks dataset *wine*, pendekatan eksklusif seperti *K-Means* mungkin lebih cocok untuk memahami pola-pola yang jelas dalam karakteristik *wine* [19]. Sementara itu, pendekatan inklusif seperti *C-Means* dapat memberikan wawasan tambahan ketika ketidakpastian atau tumpang tindih dalam klasifikasi diinginkan atau relevan. Meskipun demikian, penting untuk diakui bahwa kedua metode clustering memberikan pemahaman yang berharga terkait dengan diversitas dalam dataset *wine*. Pilihan antara *K-Means* dan *C-Means* dapat tergantung pada tujuan analisis yang diinginkan dan preferensi interpretatif peneliti.

Temuan ini, oleh karena itu, memberikan landasan yang kuat untuk analisis lebih lanjut terkait dengan preferensi konsumen terhadap jenis *wine* dan pengembangan strategi pemasaran yang lebih terarah dalam industri *wine*. Implikasi praktis dari hasil eksperimen ini dapat memberikan panduan berharga bagi pemangku kepentingan dalam industri *wine* untuk meningkatkan pemahaman tentang keberagaman karakteristik *wine* dan menyusun strategi yang lebih efektif dalam memenuhi ekspektasi konsumen.

5. KESIMPULAN

Dalam mengakhiri penelitian ini, dapat disimpulkan bahwa penerapan metode clustering pada dataset jenis *wine* dengan menggunakan algoritma *K-Means* dan *C-Means* memberikan wawasan yang berharga terkait dengan struktur internal data. *K-Means* menunjukkan kecenderungan untuk memberikan pembagian yang eksklusif dan jelas antara kelompok-kelompok, dengan membentuk kluster yang terpisah secara tegas. Sebaliknya, *C-Means* menghasilkan kelompok yang lebih fleksibel atau *fuzzy*, mencerminkan tingkat keanggotaan yang variatif dari setiap data dalam kelompoknya.

Pentingnya pemilihan metode clustering menjadi semakin nyata ketika hasil evaluasi menggunakan skor siluet diperhatikan. Algoritma *K-Means* mendemonstrasikan kinerja yang lebih unggul, dengan nilai *silhouette score* mencapai 0.660522 atau 66.05%. Skor yang mendekati 1 atau

100% menandakan tingkat pemisahan kluster yang sangat baik. Di sisi lain, metode *C-Means*, meskipun memberikan informasi yang berharga tentang tingkat keanggotaan dalam kelompok, menunjukkan *silhouette score* yang lebih rendah, yaitu sebesar 0.3786 atau 37.86%.

Analisis mendalam menegaskan bahwa *K-Means* memiliki keunggulan dalam menghasilkan pemisahan kluster yang lebih tajam dan jelas. Namun, penting untuk dicatat bahwa pilihan antara *K-Means* dan *C-Means* harus disesuaikan dengan tujuan analisis dan preferensi interpretatif peneliti. Jika kejelasan dalam pemisahan kluster menjadi prioritas, maka *K-Means* mungkin menjadi pilihan yang lebih tepat. Di sisi lain, jika keinginan untuk memahami tingkat keanggotaan yang lebih nuansa dalam kelompok adalah fokus utama, maka *C-Means* dapat memberikan perspektif yang lebih dinamis.

Selain itu, hasil penelitian ini membuka peluang untuk pengembangan lebih lanjut terkait dengan pemahaman karakteristik *wine* dan preferensi konsumen. Analisis ini dapat menjadi dasar untuk strategi pemasaran yang lebih terarah, dengan mempertimbangkan variasi dalam kelompok *wine* dan kecenderungan konsumen. Dengan demikian, kesimpulan ini memberikan landasan yang kokoh untuk penelitian lanjutan dan aplikasi praktis dalam industri *wine*.

UCAPAN TERIMAKASIH

Saya ingin mengucapkan terima kasih kepada Tuhan yang Maha Esa karena telah melancarkan penulisan penelitian ini dan kepada pembaca yang telah meluangkan waktu untuk membaca penelitian ini. Dan semoga kerja keras dapat menghasilkan dampak positif bagi dunia ilmiah dan masyarakat secara umum.

DAFTAR PUSAKA

- [1] Dhuhita, W. M. P. (2015). *Clustering Menggunakan Metode K-Means Untuk Menentukan Status Gizi Balita*.
- [2] Herlawati, H., & Handayanto, R. T. (2013). *Klasifikasi Dinamis dengan Modifikasi Algoritma Fuzzy C-means (Fcm)*.
- [3] Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). *Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan*.
- [4] Sulastrri H, Gufroni AI. Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia. *Jurnal Teknologi dan Sistem Informasi*. 2017
- [5] Priyatman, H., Sajid, F., & Haldivany, D. (2019). *Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa*.
- [6] Ningrat, D. R., Di Asih, I. M., & Wuryandari, T. (2016). *Analisis cluster dengan algoritma K-Means dan Fuzzy C-Means clustering untuk pengelompokan data obligasi korporasi*.
- [7] Agustina, N., & Prihandoko, P. (2018). *Perbandingan Algoritma K-Means dengan Fuzzy C-Means Untuk Clustering Tingkat Kedisiplinan Kinerja Karyawan*.
- [8] Agusta, Y. (2007). K-means–penerapan, permasalahan dan metode terkait. *Jurnal Sistem dan informatika*
- [9] Setiaji, G. G., Khoirudin, K., & Vydia, V. (2019). *Komparasi Metode Clustering K-Means dan Fuzzy C-Means Untuk Mempredesi Ketepatan Waktu Lulus*.
- [10] Setiaji, G. G., Khoirudin, K., & Vydia, V. (2019). *Komparasi Metode Clustering K-Means dan Fuzzy C-Means Untuk Mempredesi Ketepatan Waktu Lulus*.
- [11] Praja, A., Lubis, C., & Herwindiati, D. E. (2017). *Deteksi Penyakit Diabetes dengan Metode Fuzzy C-Means Clustering dan K-Means Clustering*.
- [12] Darmi, Y. D., & Setiawan, A. (2016). Penerapan metode
- [13] Syarif, R., Furqon, M. T., & Adinugroho, S. (2018). Perbandingan Algoritme K-Means Dengan Algoritme Fuzzy C Means (FCM) Dalam Clustering Moda Transportasi Berbasis GPS. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(10), 4107-4115.
- [14] Butarbutar, N., Windarto, A. P., Hartama, D., & Solikhun, S. (2017). Komparasi kinerja algoritma fuzzy c-means dan k-means dalam pengelompokan data siswa berdasarkan prestasi nilai akademik siswa. *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, 1(1), 46-55.

- [16] Andika, T. H., & Hafiz, A. (2018, November). Analisis perbandingan segmentasi citra menggunakan metode k-means dan fuzzy c-means. *In Prosiding Seminar Nasional Darmajaya* (Vol. 1, No. 1, pp. 237-246).
- [17] Firdaus, H. S., Nugraha, A. L., Sasmito, B., & Awaluddin, M. (2021). Perbandingan metode fuzzy c-means dan k-means untuk pemetaan daerah rawan kriminalitas di kota semarang. *Elipsoida: Jurnal Geodesi dan Geomatika*, 4(01), 58-64.
- [18] Rahmati, R. (2021). Analisis Cluster dengan Algoritma K-means, Fuzzy C-means dan Hierarchical Clustering (Studi Kasus: Indeks Pembangunan Manusia Tahun 2019). *JIKO (Jurnal Informatika dan Komputer)*, 5(2), 73-80.
- [19] Dwididanti, S., Anggoro, D. A., & Sutanto, M. H. (2022). Analisis Perbandingan Algoritma Bisecting K-Means dan Fuzzy C-Means pada Data Pengguna Kartu Kredit. *Emitor: Jurnal Teknik Elektro*, 22(2), 110-117.
- [20] Putri, A. L. R., & Dwidayati, N. (2021). Analisa perbandingan k-means dan fuzzy c-means dalam pengelompokan daerah penyebaran COVID-19 Indonesia. *Unnes Journal of Mathematics*, 50-55.
- [21] Praditya, N. W. P. (2023). Prediksi Kualitas Red Wine dan White Wine Menggunakan Data Mining. *Journal Software, Hardware and Information Technology*, 3(2), 25-33.