

Penggunaan Metode *Rank Order Centroid* dalam Penentuan Nilai *Centroid* (Studi Kasus : *Dataset Biji Gandum*)

Ricky Dwi Saputro ¹, Dwi Retnoningsih ², Hardika Khusnuliawati ³

^{1,2,3} Program Studi Informatika, Fakultas Sains Teknologi dan Kesehatan, Universitas Sahid Surakarta

ABSTRACT

Clustering is a data processing in data mining, namely a clustering process by separating a set of data into smaller groups or clusters based on similar characteristics. A clustering method in data mining is K-Means. The K-Means method processes data into the desired number of clusters and the data will be placed into clusters based on the proximity of the centroid or distance to each cluster. The conventional K-Means method has the disadvantage, namely the initial centroid selection is random so it can produce different results. Therefore, initial centroid selection can improve accuracy in the clustering process. Determining the initial centroid can be done in various ways such as Rank Order Centroid (ROC). ROC is a method that gives a score to each criterion according to its ranking appropriate to score based on its priority level. This research compares the evaluation results with the Davies Bouldin Index (DBI) method, which is a cluster validation method for grouping methods. The results of the clustering method using wheat grain data show that a DBI K-Means value using ROC of 0.33350 with 4 iterations in 2 clusters.

Keywords: Clustering, K-Means, centroid, ROC

ABSTRAK

Clustering merupakan salah satu contoh proses pengolahan data pada data mining, yaitu proses clustering dengan cara memisahkan sekumpulan data menjadi kelompok-kelompok atau cluster yang lebih kecil berdasarkan kesamaan karakteristik. Salah satu metode clustering dalam data mining adalah K-Means, metode K-Means mengolah data menjadi sejumlah cluster yang diinginkan dan data tersebut akan ditempatkan ke dalam cluster berdasarkan kedekatan centroid atau jarak ke masing-masing cluster. Metode K-Means konvensional mempunyai kekurangan yaitu pemilihan centroid awal bersifat acak sehingga dapat menghasilkan hasil yang berbeda-beda, sehingga pemilihan centroid awal dapat meningkatkan akurasi dalam proses clustering. Penentuan centroid awal dapat dilakukan dengan berbagai cara, salah satunya adalah Rank Order Centroid (ROC) yaitu suatu metode yang memberikan bobot pada setiap kriteria sesuai rankingnya yang akan diberikan nilai berdasarkan tingkat prioritasnya. Penelitian ini membandingkan hasil evaluasi dengan metode Davies Bouldin Index (DBI), yaitu metode validasi cluster untuk metode pengelompokan. Hasil dari metode clustering menggunakan data biji gandum diperoleh nilai DBI K-Means menggunakan ROC sebesar 0.33350 dengan 4 kali iterasi pada 2 cluster.

Kata Kunci: Clustering, K-Means, centroid, ROC

1. PENDAHULUAN

Data mining adalah proses mengekstraksi dan mengidentifikasi informasi bermanfaat dan pengetahuan dari berbagai basis data besar menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* (Pradana, 2019). Salah satu fungsi data mining yaitu *clustering*, *clustering* adalah proses dimana sekumpulan data atau objek akan dipisah kedalam kelompok atau *cluster* yang lebih kecil berdasarkan kesamaan ciri yang dimiliki (Putra, 2017). Salah satu jenis *clustering* yang dapat digunakan untuk mengelompokkan data ke dalam beberapa *cluster* berdasarkan tingkat kemiripan antar data yaitu metode K-Means. Metode K-Means merupakan salah satu metode yang populer karena mudah dalam implementasi ke berbagai jenis kasus dan atribut data, selain kemudahan implementasi metode K-Means juga masuk dalam 10 besar metode data mining yang paling digunakan (Nishom, 2018).

K-Means *clustering* adalah Metode pengelompokan yang membagi data ke dalam kelompok yang memiliki karakteristik sama akan dikelompokkan kedalam kelompok yang sama sehingga *dataset* tidak tumpang tindih Sirait (2017). Algoritma K-Means konvensional secara umum dalam penentuan jumlah *cluster* dilakukan oleh *user* dan dalam penentuan nilai *centroid* untuk tiap *cluster* dilakukan secara acak kemudian data akan ditempatkan ke dalam *cluster* berdasarkan kedekatan jarak yang ada (Putra, 2017). Menurut Salsabila (2019) pada metode *clustering* K-Means memiliki kekurangan dalam pemilihan *centroid* awal yang acak dapat membuat hasil yang berbeda-beda, maka disarankan menggunakan metode yang cocok dalam pemilihan *centroid* awal sehingga akan meningkatkan akurasi pada proses *clustering*. Sedangkan menurut Gustientiedina (2019) metode *clustering* K-Means perlu dilakukan penentuan nilai *centroid* terbaik agar proses *clustering* data dapat dilakukan lebih valid.

Rank Order Centroid (ROC) merupakan metode yang sederhana, dapat menghasilkan nilai bobot sesuai kriteria yang digunakan, kinerja tertinggi dalam hal identifikasi alternatif dan dapat mengatasi pembobotan atribut dalam pengambilan keputusan. Kelebihan ROC adalah dapat memperlihatkan kriteria yang lebih penting atau diprioritaskan sampai akhir kriteria dan dalam menentukan nilai bobot didapat dari urutan tingkat prioritas kriteria dimulai dari urutan pertama, kedua, ketiga dan seterusnya (Panjaitan, 2021).

Penelitian ini menggunakan *dataset seeds* atau biji gandum, karena penggunaan *dataset seeds* bertujuan mengembangkan pengelompokan biji gandum berdasarkan kelompok yang diinginkan menggunakan metode K-Means, sehingga individu atau petani atau perusahaan akan lebih mudah dalam mengelompokkan berbagai biji yang berbeda. Penelitian ini memprioritaskan biji yang memiliki kriteria area biji yang besar untuk prioritas utama dalam pemilihan biji.

Penelitian ini akan memperbaiki kelemahan algoritma K-Means yang penentuan nilai *centroid* awal secara acak, sehingga pada penelitian ini akan membahas penentuan nilai *centroid* awal dengan menggunakan metode *Rank Order Centroid* dan K-Means konvensional dengan studi kasus *seed dataset* atau data biji gandum.

2. METODE PENELITIAN

2.1. Clustering

Menurut Pradana (2019), *clustering* adalah proses mengelompokkan data data menjadi beberapa kelompok, sehingga objek yang memiliki karakteristik yang hampir sama atau sama akan di kelompokkan ke dalam satu kelompok dan objek yang tidak memiliki banyak persamaan atau berbeda akan di kelompokkan ke kelompok lain.

Menurut Olivia (2019), *clustering* dengan pendekatan partisi (*Partition-Based Clustering*) adalah proses pengelompokan data dengan menyaring data yang sudah dianalisis ke dalam *cluster* yang ada. *Clustering* dengan pendekatan hirarki (*Hierarchical Clustering*) adalah proses pengelompokan data yang mirip ditempatkan pada hirarki yang bedekatan sedangkan yang tidak diletakan pada hirarki yang berjauhan.

2.2. K-Means

Menurut Olivia (2019), Metode K-Means pertama kali diperkenalkan oleh MacQueen JB pada tahun 1976. K-Means merupakan suatu proses partisi data ke dalam bentuk satu atau lebih *cluster*, sehingga satu *cluster* memiliki data yang karakteristiknya sama dan data yang memiliki karakteristik yang berbeda akan di kelompokkan di *cluster* yang berbeda.

Langkah-langkah *clustering* dalam algoritma K-Means :

1. Tentukan jumlah *cluster* (k) pada *dataset*.
2. Tentukan nilai pusat (*centroid*).

Penentuan nilai *centroid* pada tahap awal dilakukan secara acak, sedangkan pada tahap terasi digunakan rumus rumus sebagai berikut :

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (1)$$

3. Pada masing-masing record, hitung jarak terdekat dengan *centroid* menggunakan rumus Euclidean distance .

$$De = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (2)$$

4. Kelompokkan objek berdasarkan jarak ke *centroid* terdekat.
5. Ulangi langkah ke-3 hingga langkah ke-4, lakukan iterasi hingga *centroid* bernilai optimal.

2.3. Rank Order Centroid (ROC)

Rank Order Centroid (ROC) dapat dihasilkan melalui tingkat kepentingan atau prioritas dari kriteria. Teknik ROC memberikan bobot pada setiap kriteria sesuai dengan ranking yang akan diberikan nilai berdasarkan tingkat prioritas (Ghazali, 2021) dan dirumuskan sebagai berikut :

$$W_k = \frac{1}{k} \sum_i^k = 1\left(\frac{1}{i}\right) \quad (3)$$

2.4. Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) adalah salah satu metode pengukuran validitas *cluster* pada suatu metode pengelompokan, kedekatan data antara titik pusat *cluster* dari *cluster* yang diikuti akan dijumlah sebagai kohesi. Nilai DBI yang lebih kecil atau rendah memiliki akurasi yang lebih akurat (Muhima, 2023). Tahapan perhitungan DBI sebagai berikut :

1. Sum of Square Within *Cluster* (SSW) metrik kohesi dalam sebuah *cluster* ke-*i*

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} \quad (4)$$

2. Sum of Square Between *Cluster* (SSB) metrik untuk separasi antara dua *cluster*

$$SSB_{i,j} = d(c_i, c_j) \quad (5)$$

3. Rasio seberapa baik nilai perbandingan antara *cluster* ke-*i* dan *cluster* ke-*j*.

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (6)$$

4. Kemudian nilai Davies Bouldin Index (DBI)

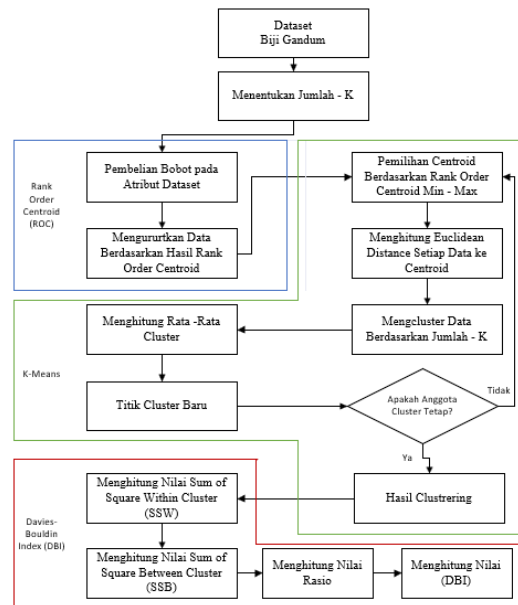
$$DBI = \frac{1}{k} \sum_{j=1}^k \max (R_{ij}) \quad (7)$$

2.5. Kerangka Penelitian

Untuk mendapatkan hasil pengelompokan yang optimal, dilakukan dengan tahapan seperti yang ditampilkan pada Gambar 1. Proses awal dengan memasukan *dataset* biji gandum, setelah *dataset* dimasukan proses selanjutnya menentukan jumlah *k* atau *cluster*. Kemudian masuk kedalam metode ROC dalam menentukan nilai *centroid* awal, proses ROC awalnya dengan memberi nilai bobot pada atribut *dataset* kemudian mengurutkan data berdasarkan hasil ROC. Setelah nilai ROC didapatkan proses kemudian memilih nilai *centroid* awal berdasarkan tertinggi - terendah ROC pada proses sebelumnya. Setelah nilai *centroid* awal didapatkan kemudian menghitung jarak antara objek dengan *centroid* menggunakan rumus euclidean distance (2).

Kemudian mengelompokkan data berdasarkan kedekatan data dengan *centroid* awal. Menghitung rata-rata dari data yang berada pada *centroid* yang sama untuk menentukan titik *centroid* baru. Jika titik *centroid* baru memiliki anggota *cluster* yang sama maka hasil

pengelompokan sudah ditemukan jika tidak maka akan mengulangi dari proses menentukan *centroid* baru dengan menggunakan persamaan (1). Setelah mendapatkan hasil pengelompokan proses selanjutnya adalah menghitung nilai akurasi pengelompokan menggunakan DBI, dengan proses pertama menghitung nilai SSW menggunakan formula (4). Proses kedua menghitung nilai SSB menggunakan formula (5) dan kemudian menghitung nilai Rasio menggunakan formula (6), setelah nilai SSW, SSB dan Rasio didapatkan kemudian menghitung nilai DBI menggunakan formula (7).



Gambar 1. Kerangka Penelitian

3. HASIL DAN PEMBAHASAN

3.1. Hasil Iterasi

Berdasarkan hasil proses *clustering* yang telah dilakukan diketahui bahwa pada proses *clustering 2 cluster* dalam K-Means dengan ROC membutuhkan 4 iterasi sedangkan K-Means konvensional membutuhkan 3 iterasi. Pada proses *clustering 3 cluster* dalam K-Means yang menggunakan ROC membutuhkan 4 iterasi sedangkan K-Means konvensional membutuhkan 7 iterasi. Pada proses *clustering 4 cluster* dalam K-Means yang menggunakan ROC membutuhkan 7 iterasi sedangkan K-Means konvensional membutuhkan 8 iterasi. Pada proses *clustering 5 cluster* dalam K-Means yang menggunakan ROC membutuhkan 5 iterasi sedangkan K-Means konvensional membutuhkan 5 iterasi.

Berdasarkan hasil perbandingan dari *clustering* data biji gandum menggunakan algoritma K-Means konvensional dan K-Means dengan ROC menghasilkan jumlah iterasi yang berbeda, dimana dari *cluster 3* dan *4* jumlah iterasi K-Means dengan ROC memiliki jumlah iterasi yang lebih sedikit dari K-Means konvensional. Sedangkan jumlah iterasi di *cluster 2* K-Means konvensional memiliki jumlah iterasi lebih sedikit dan di *cluster 5* keduanya memiliki jumlah iterasi yang sama atau dapat dilihat di tabel 1.

Tabel 1. Jumlah iterasi

Jumlah iterasi		
Cluster	K-Means & ROC	K-Means Konvensional
2	4	3
3	4	7
4	7	8

3.2. Hasil Nilai Akhir DBI

Berdasarkan hasil proses evaluasi diketahui bahwa hasil evaluasi DBI untuk proses *clustering 2 cluster* dalam K-Means yang menggunakan ROC memiliki nilai sebesar 0.33350 sedangkan K-Means konvensional memiliki nilai sebesar 0.34458 untuk proses *clustering 3 cluster* dalam K-Means yang menggunakan ROC memiliki nilai sebesar 0.75533 sedangkan K-Means konvensional memiliki nilai sebesar 0.75437 untuk proses *clustering 4 cluster* dalam K-Means yang menggunakan ROC memiliki nilai sebesar 0.72457 sedangkan K-Means konvensional memiliki nilai sebesar 0.89244 untuk proses *clustering 5 cluster* dalam K-Means yang menggunakan ROC memiliki nilai sebesar 0.94513 sedangkan K-Means konvensional memiliki nilai sebesar 0.85115. Hasil nilai akhir evaluasi menggunakan DBI dapat dilihat di tabel 2.

Tabel 2. Hasil nilai akhir DBI

Cluster	Nilai DBI	
	K-Means & ROC	K-Means Konvensional
2	0.33350	0.34458
3	0.75533	0.75437
4	0.72457	0.89244
5	0.94513	0.85115

Berdasarkan perbandingan hasil evaluasi *clustering* K-Means & ROC dan K-means konvensional menggunakan metode DBI yang dapat dilihat di Gambar 4.6 bahwa, hasil nilai akhir evaluasi *clustering 2 cluster* dari K-Means & ROC memiliki nilai lebih rendah, selisih 0,01108 dari hasil nilai akhir K-Means konvensional. Kemudian dalam hasil nilai akhir evaluasi *clustering 3 cluster* dari K-Means & ROC memiliki nilai lebih rendah, selisih 0,00096 dari hasil nilai akhir K-Means konvensional. Kemudian dalam hasil nilai akhir evaluasi *clustering 4 cluster* dari K-Means & ROC memiliki nilai lebih rendah, selisih 0,16787 dari hasil nilai akhir K-Means konvensional. Kemudian dalam hasil nilai akhir evaluasi *clustering 5 cluster* dari K-Means konvensional memiliki nilai lebih rendah, selisih 0,09398 dari hasil nilai akhir K-Means & ROC.

4. KESIMPULAN

Pengujian dari *clustering dataset* biji gandum menggunakan algoritma K-Means konvensional dan K-Means dengan ROC menghasilkan jumlah itersi lebih sedikit di K-Means dengan ROC, sedangkan K-Means konvensional menghasilkan lebih banyak iterasi. Hasil akhir DBI untuk K-Means dengan ROC menghasilkan nilai lebih sedikit dari K-Means konvensional, dengan menentukan *centroid* awal menggunakan ROC dapat mempengaruhi hasil iterasi dan hasil akhir DBI dan dengan penelitian menggunakan *dataset* biji gandum diharapkan individu atau petani maupun perusahaan dapat dijadikan sebagai model dalam *clustering* yang efektif dan akurat sesuai kriteria yang diinginkan. Berdasarkan hasil penelitian *dataset* biji gandum yang telah dilakukan menggunakan metode K-Means sesuai kriteria area biji gandum dapat disimpulkan bahwa K-Means dan ROC di 2 *cluster* adalah *cluster* terbaik karena hasil evaluasi DBI yang paling mendekati angka nol sebesar 0.33350 dan memiliki iterasi lebih rendah yaitu 4.

DAFTAR PUSTAKA

- Ghazali, A. (2021). Analisis Perbandingan Pembobotan Menggunakan *Fuzzy Logic* dan *Rank Order Centroid* (ROC) pada Metode *Simple Additive Weighting* (SAW). Medan: Teknik Informatika Universitas Sumatera Utara.
- Gustientiedina, Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means untuk *Clustering* Data Obat-Obat pada RSUD Pekanbaru. *Jurnal Nasional Teknologi dan Sistem Informasi* , 17-24.
- Muhima, R. R., Kurniawan, M., & Wardhana, S. R. (2023). *An Improved Clustering Based on K-Means for Hotspots Data*. *Indonesian Journal of Electrical Engineering and Computer Science*, 1109-1117.
- Nishom, M., & Wibowo, D. S. (2018). Implementasi Metode K-Means berbasis Chi-Square pada Sistem Pendukung Keputusan untuk Identifikasi Disparitas Kebutuhan Guru. *Jurnal Sistem Informasi Bisnis* , 187-194.
- Olivia, L. (2019). Pengelompokan Data Rekam Medis untuk Mengetahui Penyakit Endemi di suatu Daerah Menggunakan K-Means *Clustering*. Medan: Teknik Informatika Universitas Sumatra Utara .
- Panjaitan, E. A., & Desnelita, Y. (2021). Implementasi Metode *Rank Order Centroid* dan *Additive Ratio Assessment* dalam Penelitian Dosen. *Seminar Nasional Informatika*, 385-392.
- Pradana, C. C. (2019). Pengelompokan Data Evaluasi Pembelajaran Menggunakan Algoritma K-Means++ *Clustering*. Yogyakarta: Teknik Informatika Universitas Sanata Dharma.
- Putra, A. Z. (2017). Analisis Kinerja Metode Gabungan *Genetic Algorithm* dan K-Means *Clustering* dalam Penentuan Nilai *Centroid*. Medan: Teknik Informatika Universitas Sumatera Utara.
- Salsabila, N. (2019). Klasifikasi Barang Menggunakan Metode *Clustering* K-Means dalam Penentuan Prediksi Stok Barang : Studi Kasus UKM Mar'ah Jilbab Kediri. Malang: Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim.
- Sirait, K. (2017). Analisis Kinerja Algoritma K-Means Dengan Penentuan Nilai *Centroid* Awal Secara Random dan *KD Tree*. Medan: Teknik Informatika Universitas Sumatera Utara.