

# Perbandingan Metode Perhitungan Jarak pada Nilai Centroid dan Pengelompokan Data Menggunakan K-Means Clustering

Budi Hartono\*, Sri Eniyati, Kristophorus Hadiono

Fakultas Teknologi Informasi dan Industri, Program Studi Teknik Informatika, Universitas Stikubank Semarang, Indonesia

Email: <sup>1</sup>\*budihartono@edu.unisbank.ac.id, <sup>2</sup>eniyati@edu.unisbank.ac.id, <sup>3</sup>kristophorus.hadiono@edu.unisbank.ac.id

Email Penulis Korespondensi: budihartono@edu.unisbank.ac.id

Submitted: 23/03/2023; Accepted: 31/03/2023; Published: 31/03/2023

**Abstrak**—Penelitian ini akan mengamati proses pengelompokan data atau pembentukan kluster menggunakan kluster K-Means dengan tiga metode pengukuran jarak yaitu Euclidean distance, Manhattan distance, dan Minkowski distance. Pengamatan lebih difokuskan pada perubahan nilai centroid dan hasil pengelompokan data, serta jumlah iterasi yang diperlukan. Data percobaan berjumlah 20, 30, 40, dan 50 buah data yang dikelompokkan menjadi 2 kelompok. Penelitian ini juga merangkum penerapan kluster K-Means yang telah banyak digunakan diberbagai bidang, antara lain Kesehatan, Pendidikan, dan Kebencanaan. Hasil pengelompokan data dengan ketiga metode pengukuran jarak adalah tidak terlalu jauh berbeda, yaitu selisih tertinggi adalah 2 anggota data pada data uji 50. Iterasi terbanyak pada data uji 40 menggunakan Euclidean distance yaitu 7 iterasi, dan iterasi paling sedikit pada data uji 20 menggunakan Minkowski distance yaitu 3 iterasi. Pada data uji 50 dibutuhkan 4 iterasi. Jumlah data uji tidak berbanding lurus dengan jumlah iterasi yang dibutuhkan untuk mencapai kluster dalam keadaan stabil.

**Kata Kunci:** Pengelompokan Data; Euclidean Distance; Centroid; Kluster K-Means

**Abstract**—This study will observe the process of grouping data or forming clusters using K-Means clusters with three methods of measuring distances, namely Euclidean distance, Manhattan distance, and Minkowski distance. Observations are more focused on changing the centroid value and the results of grouping data, as well as the number of iterations required. Experimental data amounted to 20, 30, 40, and 50 pieces of data which were grouped into 2 groups. This research also summarizes the application of K-Means clusters which have been widely used in various fields, including Health, Education, and Disaster. The results of grouping data with the three distance measurement methods are not too much different, namely the highest difference is 2 members of the data on 50 test data. The most iterations on 40 test data use the Euclidean distance, namely 7 iterations, and the least iteration on 20 test data uses Minkowski distance i.e. 3 iterations. On the 50 test data it takes 4 iterations. The amount of test data is not directly proportional to the number of iterations needed to reach the cluster in a stable state.

**Keywords:** Data Cluster; Euclidean Distance; Centroid; K-Means Cluster

## 1. PENDAHULUAN

Mengolah data bertujuan untuk mengetahui pola sehingga memperoleh informasi yang tersembunyi dari data tersebut. Pengolahan data ini sering disebut sebagai data mining. Salah satu metode data mining adalah pengelompokan data atau kluster data. Metode kluster data yang banyak dipakai adalah kluster K-Means (k-means clustering). Metode ini membutuhkan masukan jumlah kluster (k) yang ingin dibentuk [4]. Kluster K-Means merupakan salah satu metode kluster data non-hirarki yang sederhana [1]. Jumlah kluster yang tepat tidak bisa diketahui sehingga nilai k ini dapat ditentukan dengan pertimbangan kebutuhan jumlah kelompok data.

Beberapa penelitian menerapkan metode kluster K-Means menggunakan perhitungan jarak seperti Euclidean distance, Manhattan distance dan Minkowski distance. Pada penelitian [6], perhitungan jarak menggunakan Euclidean, Manhattan, dan Cosine Similarity, dengan pengelompokan data menggunakan metode kluster K-Means. Metode Euclidean Distance dan Cosine Similarity dapat melakukannya dengan baik. Penelitian ini akan mengamati proses pembentukan kluster data menggunakan kluster K-Means dengan perhitungan jarak Euclidean, Manhattan, dan Minkowski distance.

Berikut ini beberapa penelitian mengenai penerapan metode kluster K-Means di beberapa bidang, yaitu bidang Kesehatan, Pendidikan, Kebencanaan, dan lain sebagainya. Pengelompokan tingkat penyebaran Covid-19 bertujuan untuk membuat strategi penanganan penyebaran Covid-19 diseluruh provinsi di Indonesia. Disini dibagi menjadi tiga buah kluster. Melalui informasi kluster ini dapat digunakan oleh pemerintah untuk mengambil keputusan strategis menekan tingkat penyebaran Covid-19 di Indonesia [9]. Metode kluster K-Means digunakan untuk mengelompokkan penyebab penyebaran penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Provinsi Riau, menjadi dua kluster [12]. Pada penelitian [18] bertujuan untuk mengetahui persentase usia penduduk di atas 15 tahun pada 34 provinsi di Indonesia yang merokok. Data dikelompokkan menjadi 2 kluster. Sumber data diperoleh dari situs Badan Pusat Statistik (BPS) dari tahun 2015 sampai dengan 2018. Informasi hasil kluster menjadi pertimbangan pemerintah untuk memberikan edukasi mengurangi persentase perokok di Indonesia.

Penelitian oleh Prianto dan kawan-kawan [10] tentang pengajaran bahasa Inggris pada jurusan S1 Teknik Informatika. Pengelompokan tingkat kemampuan mahasiswa menjadi tiga kategori. Fakultas dapat menentukan rata-rata tingkat kemampuan mahasiswa dan dosen dapat memberikan proses pengajaran yang lebih baik. Klusterisasi dilakukan untuk mengetahui peningkatan dan penurunan jumlah mahasiswa disetiap Program Studi [13]. Hasil penelitian ini menjadi pertimbangan dan evaluasi dalam menentukan strategi meningkatkan jumlah

mahasiswa baru disetiap program studi. Pada penelitian [23] metode kluster K-Means diterapkan untuk mengelompokkan data mahasiswa sebagai rekomendasi penerima bantuan Uang Kuliah Tunggal (UKT). Pendekatan kluster K-Means digunakan untuk memetakan kemampuan penggunaan teknologi informasi di Indonesia pada penduduk usia 15-59 tahun yang memiliki keterampilan di bidang Teknologi Informasi dan Komputer (TIK) [17].

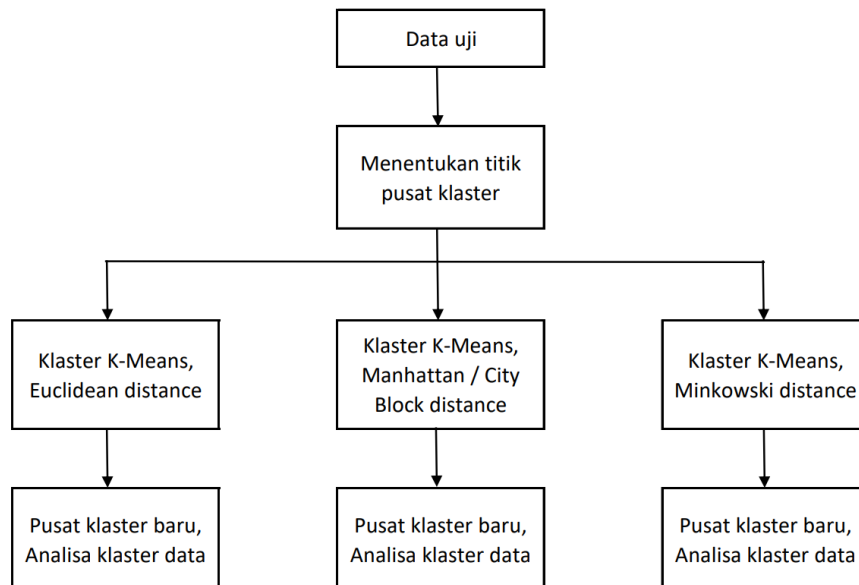
Metode kluster K-Means digunakan untuk pengelompokan episentrum gempa di Provinsi Bengkulu, dengan variabel lintang, bujur, dan magnitudo. Data gempa tektonik di Provinsi Bengkulu dan sekitarnya dari Januari 1970 hingga Desember 2015, diperoleh dari stasiun Badan Meteorologi Klimatologi dan Geofisika (BMKG) Kepahiang Bengkulu [11]. Pulau Jawa dengan populasi penduduk yang padat merupakan salah satu daerah rawan gempa bumi. Metode kluster K-Means digunakan untuk mengetahui dampak gempa bumi berdasarkan tingkat keparahannya. Dampak bencana dibagi menjadi 4 kluster. Data gempa bumi diperoleh dari Badan Nasional Penanggulangan Bencana (BNPB) untuk kurun waktu tahun 2012 sampai dengan 2021 [21].

Tujuan penelitian ini adalah mengamati perubahan nilai centroid dan hasil pengelompokan data, serta jumlah iterasi yang diperlukan untuk proses tersebut. Data percobaan berjumlah 50 buah dikelompokkan menggunakan kluster K-Means dengan beberapa macam metode perhitungan jarak yaitu Euclidean distance, City Block distance (atau Manhattan distance), dan Minkowski distance.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian diawali dengan menyiapkan data uji sebanyak 50 buah. Data uji ini akan dikelompokkan menjadi dua kelompok. Proses pengelompokan data menggunakan kluster K-Means menggunakan tiga metode perhitungan jarak yaitu Euclidean distance, Manhattan distance (City Block distance), dan Minkowski distance. Pada awal proses iterasi, centroid atau titik pusat kluster adalah sama untuk ketiga metode perhitungan jarak, yaitu nilai C1 (3, 18) dan C2 (23, 2). Nilai awal centroid ditentukan secara acak, untuk selanjutnya pada saat pembentukan kluster nilai ini akan berubah. Diagram alir tahapan penelitian dapat dilihat pada Gambar 1.



**Gambar 1.** Tahapan Penelitian

Proses pengelompokan data dikerjakan secara berulang (iterasi) sampai dengan diperoleh komposisi kelompok data yang sudah tidak berubah atau tetap. Diperoleh titik pusat kluster baru, serta menganalisa hasil kluster atau kelompok data.

### 2.2 Metode Pengukuran Jarak

Jarak Euclidean adalah metode pengukuran jarak (distance) data yang banyak digunakan. Berikut ini adalah beberapa penelitian mengenai penerapan beberapa metode perhitungan jarak. Metode Euclidean distance digunakan dalam sistem penghitung skor tembak secara otomatis, yaitu untuk mengukur jarak antara lubang tembak terhadap titik pusat sekaligus menghitung skor otomatis [2]. Mengevaluasi kinerja algoritma K-means pada dataset kanker serviks berdasarkan perhitungan jarak Euclidean dan Manhattan. Algoritma K-Means dengan metode perhitungan jarak Euclidean memiliki akurasi yang lebih baik dibandingkan dengan metode perhitungan jarak Manhattan [3]. Perbandingan pengukuran jarak (Euclidean, Manhattan, dan Minkowski) untuk pelabelan kluster. Pengukuran jarak Euclidean yang terbaik untuk diterapkan dalam kluster K-Means [5]. Pada penelitian

[7], membandingkan dua distance space yaitu antara Manhattan dan Euclidean. Euclidean distance adalah yang optimal digunakan dalam pengelompokan data.

Tiga buah persamaan berikut digunakan untuk menghitung jarak untuk data dengan dua buah fitur (p, q) [19]. Minkowski distance seperti pada Persamaan 1, Manhattan (City Block) distance dapat dilihat pada Persamaan 2, dan Euclidean distance dapat dilihat pada Persamaan 3.

$$d(p,q) = (\sum_{i=1}^n |p_i - q_i|^c)^{1/c} \tag{1}$$

$$d(p,q) = \sum_{i=1}^n |p_i - q_i| \tag{2}$$

$$d(p,q) = \sqrt{\sum_{k=1}^n (p_k - q_k)^2} \tag{3}$$

### 2.3 Kluster K-Means

Ada dua jenis kluster data yang sering digunakan dalam proses pengelompokan data, yaitu Hierarchical dan non-Hierarchical. K-means adalah metode kluster non-hierarki. Metode kluster K-Means berupaya untuk mengelompokkan data yang ada menjadi beberapa kelompok, dimana data dalam satu kelompok memiliki karakteristik yang sama dan memiliki karakteristik yang berbeda dengan data pada kelompok lainnya [15][20]. Metode ini banyak diterapkan diberbagai bidang kehidupan. Pada Tabel 1 mencatat beberapa studi kasus penelitian penerapan klusterisasi data.

**Tabel 1.** Ragam bidang penerapan klusterisasi

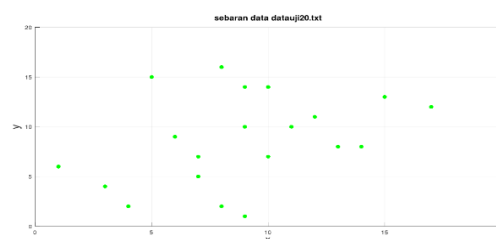
Bidang Penerapan	Studi kasus klusterisasi
Kesehatan	Penyebaran Covid-19 di Indonesia [9]
	Penyebaran penyakit ISPA di Provinsi Riau [12]
	Persentase perokok usia di atas 15 tahun [18]
Pendidikan	Kemampuan pelajar pada Bahasa Inggris [10]
	Mahasiswa baru memilih program studi [13]
	Penilaian Kedisiplinan Siswa [14]
Kebencanaan	Menentukan prioritas bantuan uang kuliah tunggal [23]
	Episentrum gempa bumi [11]
	Dampak gempa bumi di Pulau Jawa [21]
Teknologi - Perangkat Lunak	Pemetaan kemampuan penggunaan Teknologi Informasi [17]
	Kerusakan / cacat perangkat lunak [22]
Lain-lain	Pemetaan produksi tanaman Tomat di Indonesia [1]
	Minat nasabah asuransi [8]

Proses penyusunan data menjadi anggota kluster pada metode kluster K-Means dilakukan dengan cara meminimalkan variasi antara data dalam sebuah kluster dan memaksimalkan variasi antara data pada kluster yang berbeda [16]. Jumlah kluster yang paling tepat tidak bisa diketahui tetapi dapat ditentukan dengan mempertimbangkan jumlah kelompok data yang diinginkan.

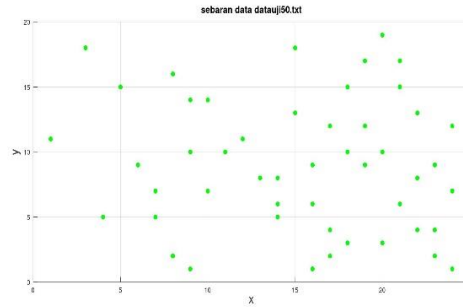
## 3. HASIL DAN PEMBAHASAN

### 3.1 Data Uji

Proses pengamatan dilakukan untuk jumlah data uji 20, 30, 40, dan 50 buah data. Data uji ini akan dibagi menjadi dua kluster yaitu C1 dan C2. Pada Gambar 2 dan Gambar 3 adalah sebaran data uji 20 dan data uji 50 buah data. Pengamatan proses pembentukan kluster data adalah pada area sumbu x (0 sampai dengan 25) dan sumbu y (0 sampai dengan 20), menyesuaikan dengan sebaran data uji. Pada proses iterasi pertama, titik pusat kluster (x, y) ditentukan secara subyektif yaitu C1 (3, 18) dan C2 (23, 2), dengan pertimbangan masih berada di daerah sebaran data uji. Dapat dilihat di Gambar 4 dan Gambar 5, pada gambar iterasi ke 1, posisi titik pusat C1 dan C2 berada di pinggir daerah sebaran data uji yaitu kiri atas dan kanan bawah. Hal ini agar dapat diamati berapa kali terjadi pergeseran atau perpindahan titik pusat kluster pada saat proses iterasi pembentukan kluster data.



**Gambar 2.** Data uji 20



**Gambar 3.** Data uji 50

**3.2 Nilai Centroid dan Iterasi**

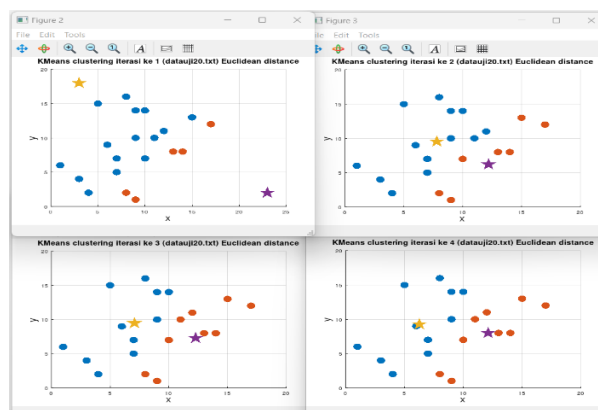
Hasil pengujian untuk mengetahui nilai centroid dan jumlah iterasi yang dibutuhkan untuk mencapai kluster dalam keadaan stabil dicatat pada Tabel 2. Proses iterasi terbanyak pada data uji 40 menggunakan Euclidean distance yaitu 7 kali iterasi. Jumlah iterasi paling sedikit pada data uji 20 menggunakan Minkowski distance yaitu 3 kali iterasi. Pada data uji 50, jumlah iterasi untuk mencapai kluster dalam keadaan stabil adalah sama yaitu 4 kali untuk ketiga metode pengukuran jarak.

Proses iterasi dan pergeseran centroid untuk data uji 20 menggunakan Euclidean distance dapat dilihat pada Gambar 4, dengan kluster stabil setelah empat kali iterasi. Pada Gambar 5 adalah proses iterasi dan pergeseran centroid untuk data uji 40 menggunakan Euclidean distance, dengan kluster stabil setelah iterasi ketujuh. Posisi centroid bergeser mendekati pusat data. Pada iterasi terakhir posisi centroid tidak berubah atau tetap karena sudah tidak terjadi lagi perubahan anggota pada setiap kluster. Perubahan data anggota sebuah kluster adalah data bertambah atau berkurang dalam sebuah kluster dikarenakan terjadi perpindahan data antar kluster.

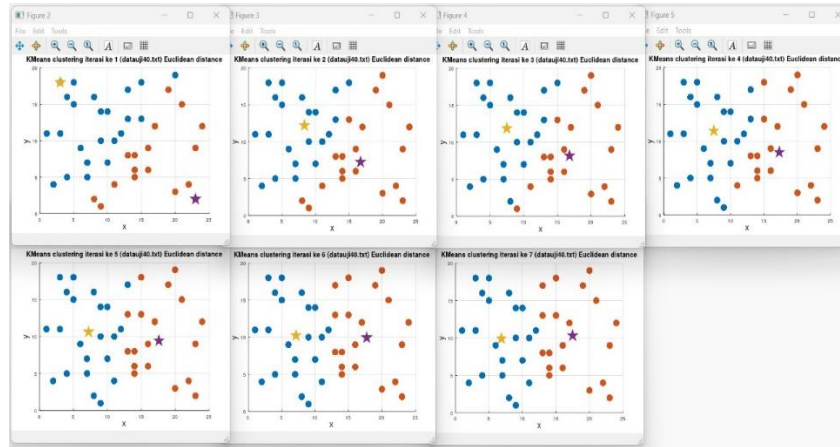
**Tabel 2.** Nilai centroid dan jumlah iterasi

Data Uji	Nilai Centroid	Euclidean Distance	Manhattan Distance	Minkowski Distance
Datauji20	C1	(6.27 , 9.27)	(6 , 5.8)	(6.27 , 9.27)
	C2	(12.11 , 8)	(11.8 , 11.6)	(12.11 , 8)
Datauji30	Iterasi	4	6	3
	C1	(7.76 , 9.58)	(8.11 , 9.5)	(7.76 , 9.58)
Datauji40	C2	(19.15 , 10.84)	(19.58 , 11.08)	(19.15 , 10.84)
	Iterasi	4	5	5
Datauji50	C1	(6.85 , 9.9)	(6.6 , 9.85)	(7.2 , 11.8)
	C2	(17.47 , 10.31)	(17.2 , 10.35)	(16.6 , 8.4)
Datauji50	Iterasi	7	6	4
	C1	(8.16 , 10.05)	(7.43 , 9.68)	(8.16 , 10.05)
Datauji50	C2	(19.21 , 8.28)	(18.91 , 8.55)	(19.21 , 8.28)
	Iterasi	4	4	4

Pada data uji 20, data uji 30, dan data uji 50 menggunakan Euclidean dan Minkowski distance, untuk iterasi terakhir nilai centroid (C1 dan C2) adalah sama. Jumlah data uji tidak berbanding lurus dengan jumlah iterasi yang dibutuhkan untuk mencapai kluster dalam keadaan stabil. Jumlah iterasi untuk mencapai kluster dalam keadaan stabil tidak secara langsung dipengaruhi oleh jumlah data, karena nilai awal centroid dan sebaran data juga turut mempengaruhi jumlah iterasi. Penelitian lebih lanjut masih diperlukan untuk dapat mengamati hal ini.



**Gambar 4.** Proses iterasi dan pergeseran centroid pada data uji 20



**Gambar 5.** Proses iterasi dan pergeseran centroid pada data uji 40

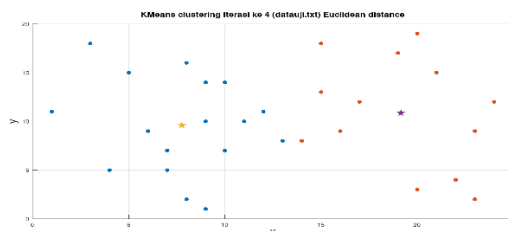
### 3.3 Hasil Pengelompokan Data

Kelompok data dari setiap data uji dicatat pada Tabel 3, yang dibagi menjadi dua kluster yaitu Kelas 1 dan Kelas 2. Pada data uji 20, data uji 30, dan data uji 50 untuk Euclidean dan Minkowski distance, jumlah anggota data pada kelas 1 dan kelas 2 adalah sama. Pada data uji 40 untuk Manhattan dan Minkowski distance, jumlah anggota data pada kelas 1 dan kelas 2 adalah sama.

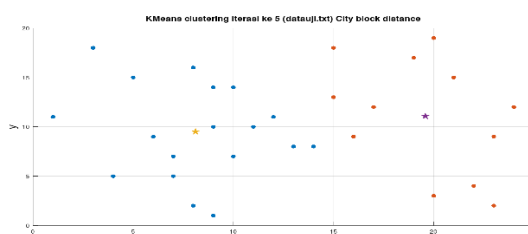
**Tabel 3.** Hasil kluster data

Data Uji	Kelompok Data	Euclidean Distance	Manhattan Distance	Minkowski Distance
Datauji20	Kelas 1	11	10	11
	Kelas 2	9	10	9
Datauji30	Kelas 1	17	18	17
	Kelas 2	13	12	13
Datauji40	Kelas 1	21	20	20
	Kelas 2	19	20	20
Datauji50	Kelas 1	18	16	18
	Kelas 2	32	34	32

Hasil pengelompokan data untuk ketiga metode pengukuran jarak adalah tidak terlalu jauh berbeda. Pada data uji 20, data uji 30, dan data uji 40 dilihat dari jumlah anggota data Kelas 1, selisih tertinggi adalah 1. Pada data uji 50 dilihat dari jumlah anggota data Kelas 1, selisih tertinggi adalah 2. Penelitian ini menggunakan dua kelompok data sehingga selisih tertinggi pada kelas yang lain (yaitu Kelas 2) adalah sama dengan Kelas 1. Penelitian lebih lanjut diperlukan untuk pengamatan pada lebih dari dua kelompok data. Hasil kluster K-Means menggunakan Euclidean distance, Manhattan distance (City Block), dan Minkowski distance, masing-masing dengan data uji 50 dapat dilihat pada Gambar 6, Gambar 7, dan Gambar 8.



**Gambar 6.** Hasil kluster K-Means menggunakan Euclidean distance dengan data uji 50



**Gambar 7.** Hasil kluster K-Means menggunakan Manhattan distance (City Block) dengan data uji 50



**Gambar 8.** Hasil kluster K-Means menggunakan Minkowski distance dengan data uji 50

## 4. KESIMPULAN

Hasil pengelompokan data menggunakan kluster K-Means dengan metode pengukuran jarak Euclidean distance, Manhattan distance, dan Minkowski distance adalah tidak terlalu jauh berbeda, yaitu selisih tertinggi adalah 2 anggota data pada data uji 50. Posisi centroid bergeser mendekati pusat data. Pada iterasi terakhir posisi centroid tidak berubah atau tetap karena sudah tidak terjadi lagi perubahan anggota pada setiap kluster. Proses iterasi terbanyak pada data uji 40 menggunakan Euclidean distance yaitu 7 iterasi. Jumlah iterasi paling sedikit pada data uji 20 menggunakan Minkowski distance yaitu 3 iterasi. Pada data uji 50 dibutuhkan 4 iterasi. Jumlah data uji tidak berbanding lurus dengan jumlah iterasi yang dibutuhkan untuk mencapai kluster dalam keadaan stabil. Jumlah iterasi untuk mencapai kluster dalam keadaan stabil tidak secara langsung dipengaruhi oleh jumlah data, karena nilai awal centroid dan sebaran data juga mempengaruhi jumlah iterasi.

## UCAPAN TERIMA KASIH

Terima kasih kepada Direktorat Penelitian Pengabdian Masyarakat dan Publikasi (DPPMP) Universitas Stikubank Semarang.

## REFERENCES

- [1] Syaifuddin, Ramlah, I. Hakim, Y. Berliana, and Nurhayati, "Pemetaan Produksi Tanaman Tomat di Indonesia Berdasarkan Provinsi Menggunakan Algoritma K-Means Clustering," *Journal of Computer System and Informatics (JoSYC)*, vol. 3, no. 4, pp. 222-228, 2022, doi: 10.47065/josyc.v3i4.2206.
- [2] M. Sarosa, M. Ridwan, I. Mahfudi, and M. B. Purwanto, "Penghitung Skor Tembak Otomatis menggunakan Metode Background Substraction dan Euclidean Distance," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 8, no. 1, pp. 140-146, 2022, doi: <http://dx.doi.org/10.26418/jp.v8i1.51265>.
- [3] S. Widodo, H. Brawijaya, and Samudi, "Clustering Kanker Serviks Berdasarkan Perbandingan Euclidean dan Manhattan Menggunakan Metode K-Means," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, pp. 687-694, 2021, doi: <http://dx.doi.org/10.30865/mib.v5i2.2947>.
- [4] R. G. Santosa, Y. Lukito, and A. R. Chrismanto, "Implementasi Metode TwoStep Clustering untuk Klasterisasi Karakteristik Akademik Mahasiswa," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 7, no. 2, pp. 121-132, 2021, doi: <http://dx.doi.org/10.26418/jp.v7i2.47735>.
- [5] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," *Jurnal Pengembangan IT (JPIT)*, vol. 4, no. 1, pp. 20-24, 2019, doi: <http://dx.doi.org/10.30591/jpit.v4i1.1253>.
- [6] M. S. Pangestu and M. A. Fitriani, "Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma K-Means," *Jurnal SAINTEKS*, vol. 19, no. 2, pp. 141-155, 2022, doi: <http://dx.doi.org/10.30595/sainteks.v19i2.14495>.
- [7] R. I. Fajriah, H. Sutisna, and B. K. Simpony, "Perbandingan Distance Space Manhattan Dengan Euclidean Pada K-Means Clustering Dalam Menentukan Promosi," *Indonesian Journal on Computer and Information Technology (IJCIT)*, vol. 4, no. 1, pp. 36-49, 2019, <https://ejournal.bsi.ac.id/ejurnal/index.php/ijcit/article/view/4630>.
- [8] J. Hutagalung and F. Sonata, "Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah Asuransi," *Jurnal Media Informatika Budidarma*, vol. 5, no. 3, pp. 1187-1194, 2021, doi: <http://dx.doi.org/10.30865/mib.v5i3.3113>.
- [9] E. Virantika, Kusnawi, and J. Ipmawati, "Evaluasi Hasil Pengujian Tingkat Clusterisasi Penerapan Metode K-Means Dalam Menentukan Tingkat Penyebaran Covid-19 di Indonesia," *Jurnal Media Informatika Budidarma*, vol. 6, no. 3, pp. 1657-1666, 2022, doi: <http://dx.doi.org/10.30865/mib.v6i3.4325>.
- [10] C. Prianto, Rd. Nuraini, and A. T. Wali, "Implementation of K-Means Methods In Clustering Students Ability Levels in English Language," *The IJICS (International Journal of Informatics and Computer Science)*, vol. 3, no. 2, pp. 49-58, 2019, doi: <http://dx.doi.org/10.30865/ijics.v3i2.1382>.
- [11] P. Novianti, D. Setyorini, and U. Rafflesia, "K-Means cluster analysis in earthquake epicenter clustering," *International Journal of Advances in Intelligent Informatics (IJAIN)*, vol. 3, no. 2, pp. 81-89, 2017, doi: <https://doi.org/10.26555/ijain.v3i2.100>.

- [12] N. Purba, Poningsih, and H. S. Tambunan, "Penerapan Algoritma K-Means Clustering Pada Penyebaran Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Provinsi Riau," *Journal of Information System Research (JOSH)*, vol. 2, no. 3, pp. 220-226, 2021, <https://ejournal.seminar-id.com/index.php/josh/article/view/736>.
- [13] H. Nopriandi and F. Haswan, "Analisis Klasterisasi Mahasiswa Baru dalam Memilih Program Studi dengan Menggunakan Algoritma K-Means," *Journal of Information System Research (JOSH)*, vol. 3, no. 4, pp. 666–671, 2022, doi: <https://doi.org/10.47065/josh.v3i4.1986>.
- [14] Q. I. Mawarni and E. S. Budi, "Implementasi Algoritma K-Means Clustering Dalam Penilaian Kedisiplinan Siswa," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 3, no. 4, pp. 522–528, 2022, doi: [10.30865/json.v3i4.4242](https://doi.org/10.30865/json.v3i4.4242).
- [15] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques 3rd Edition*, Morgan Kaufmann, 2011.
- [16] I.H. Witten, E. Frank, M.H. Hall, and C.J. Pal, *Data Mining Practical Machine Learning Tools and Techniques 4th Edition*, Morgan Kaufmann, Elsevier, 2016.
- [17] R. Watrianthos, R. Handayani, A. F. P. Akhir, Ambiyar, and U. Verawardina, "Penerapan Algoritma K-Means Pada Pemetaan Kemampuan Penggunaan Teknologi Informasi Remaja dan Dewasa di Indonesia," *Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 1, pp. 45–50, 2022, doi: [10.47065/josyc.v4i1.2264](https://doi.org/10.47065/josyc.v4i1.2264).
- [18] R. Kurniawan, Suhada, and R. Dewi, "Penerapan Algoritma K-Means Clustering Dalam Persentase Merokok Pada Penduduk Umur Di Atas 15 Tahun Menurut Provinsi," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 2, no. 2, pp. 178-186, 2021, doi: [10.30865/json.v2i2.2770](https://doi.org/10.30865/json.v2i2.2770).
- [19] Suyanto, E. Rachmawati, M. D. Sulistiyono, G. S. Wulandari, and M. Fachrie, *Explainable Artificial Intelligence Menggunakan Metode-Metode Berbasis Nearest Neighbors*, Penerbit Informatika Bandung, 2022.
- [20] Suyanto, *Data Mining untuk Klasifikasi dan Klasterisasi Data (edisi revisi)*, Penerbit Informatika Bandung, 2019.
- [21] A. Wahyu and Rushendra, "Klasterisasi Dampak Bencana Gempa Bumi Menggunakan Algoritma K-Means di Pulau Jawa," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 8, no. 1, pp. 175-179, 2022, doi: <http://dx.doi.org/10.26418/jp.v8i1.52260>.
- [22] R. Annisa, D. Rosiyadi, and D. Riana, "Improved point center algorithm for k-means clustering to increase software defect prediction," *International Journal of Advances in Intelligent Informatics (IJAIN)*, vol. 6, no. 3, pp. 328-339, 2020, doi: <https://doi.org/10.26555/ijain.v6i3.484>.
- [23] M. R. Kusnaldi, T. Gulo, and S. Aripin, "Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal," *Journal of Computer System and Informatics (JoSYC)*, vol. 3, no. 4, pp. 330–338, 2022, doi: [10.47065/josyc.v3i4.2112](https://doi.org/10.47065/josyc.v3i4.2112).