

Penggunaan Algoritma K-Means dalam Mengelompokkan Jumlah Usaha dan Hasil Pendapatan UMKM DKI Jakarta

The Use of the K-Means Algorithm in Grouping the Number of Businesses and Revenue Results of DKI Jakarta MSMEs

Sunarti*¹, Enok Tuti Alawiah², Omar Pahlevi³

¹²³Program Studi Sistem Informasi Kampus Kota Bogor, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

E-mail : sunarti.sni@bsi.ac.id*¹, enok.etw@bsi.ac.id², omar.opi@bsi.ac.id³

*Corresponding author

Received 14 October 2025; Revised 30 February 2025; Accepted 7 November 2025

Abstrak - UMKM Provinsi DKI Jakarta berkontribusi signifikan terhadap perekonomian daerah. Namun, ada permasalahannya pemetaan data lemah, segmentasi usaha kurang, belum adanya pemetaan yang jelas mengenai jumlah usaha dan pendapatan, keterbatasan akses data dan validitas informasi, serta belum adanya diferensiasi program pembangunan berdasarkan karakteristik usaha. Tujuan penelitian adalah menganalisis data jumlah usaha dan hasil pendapatan UMKM menggunakan metode K-means. Metode ini digunakan mengelompokkan jumlah usaha dan hasil pendapatan. Hasil klusterisasi menghasilkan tiga klaster wilayah: Klaster (1) yaitu kota administratif Jakarta Pusat, Jakarta Selatan, dan Jakarta Timur, memiliki jumlah usaha dan pendapatan tertinggi. Klaster(2) yaitu kota administratif Jakarta Barat dan Jakarta Utara, memiliki jumlah usaha dan pendapatan sedang. Klaster(3) yaitu kota administratif Kepulauan Seribu, memiliki jumlah usaha dan pendapatan yang rendah. Hasil evaluasi proses klusterisasi mempergunakan *Davies Bouldin Index (DBI)* bernilai -0.390. Hasil ini menunjukkan K-Means dapat memetakan jumlah usaha dan hasil pendapatan, sehingga memudahkan dalam menetapkan strategis kebijakan, mengembangkan usaha, dan menyusun strategi pemasaran berdasarkan karakteristik ekonomi setiap daerah.

Kata Kunci – UMKM, Provinsi DKI Jakarta, Klusterisasi, Metode K-Means

Abstract - MSMEs in DKI Jakarta Province contribute significantly to the regional economy. However, there are problems: weak data mapping, inadequate business segmentation, the absence of a clear mapping of the number of businesses and revenues, limited data access and information validity, and the absence of differentiation of development programs based on business characteristics. The purpose of this study is to analyze data on the number of businesses and revenues of MSMEs using the K-means method. This method is used to group businesses by number and revenue. The clustering results produce three regional clusters: Cluster(1), namely the administrative cities of Central Jakarta, South Jakarta, and East Jakarta, have the highest number of businesses and revenues. Cluster(2), namely the administrative cities of West Jakarta and North Jakarta, has a medium number of businesses and revenues. Cluster(3), namely the administrative city of the Seribu Islands, has a low number of businesses and revenues. The evaluation results of the clustering process using the *Davies Bouldin Index(DBI)* are -0.390. These results indicate that K-Means can effectively map the number of businesses and revenues, thereby facilitating the determination of strategic policies, business development, and marketing strategies tailored to the economic characteristics of each region.

Keywords – MSMEs, DKI Jakarta Province, Clustering, K-Means Method

1. PENDAHULUAN

Usaha Mikro, Kecil, dan Menengah (UMKM) salah satu sektor mempunyai peranan signifikan dalam perekonomian nasional maupun daerah. Menurut Undang-Undang No. 20 Tahun 2008, UMKM dikategorikan berdasarkan jumlah aset dan omzet tahunan. Di Provinsi DKI Jakarta, UMKM memiliki kontribusi besar dalam menciptakan lapangan kerja, meningkatkan pendapatan masyarakat, serta mendukung pertumbuhan ekonomi lokal. Berdasarkan data Badan Pusat Statistik (BPS), jumlah UMKM terus meningkat setiap tahun, sehingga memerlukan strategi pengelolaan dan pemberdayaan dengan tepat [1]. Akan tetapi pertumbuhan jumlah UMKM tersebut tidak diiringi dengan pemetaan data yang baik, sehingga seringkali pemerintah mengalami kesulitan dalam mengelompokkan skala usaha, hasil pendapatan maupun potensi pertumbuhannya. Kondisi ini menyebabkan berbagai program pendampingan, pelatihan, hingga bantuan modal tidak tepat sasaran pada kelompok usaha yang paling membutuhkan, sehingga efektivitas kebijakan pemberdayaan belum optimal [2].

Permasalahan penelitian dalam mengelompokkan jenis usaha dan hasil pendapatan UMKM DKI Jakarta ialah lemahnya pemetaan data dan kurangnya segmentasi usaha [3], [4], [5], belum adanya pemetaan jelas terkait jumlah usaha dan hasil pendapatan [6], adanya keterbatasan akses data dan validitas informasi [7], [8] dan kurangnya diferensiasi program pembinaan berdasarkan karakteristik usaha [9]. Dengan adanya kondisi yang ada maka menyebabkan program pemberdayaan dari pemerintah belum sepenuhnya efektif, karena tidak mampu mengidentifikasi kelompok UMKM yang benar-benar membutuhkan intervensi khusus sesuai kapasitas dan potensi pertumbuhannya. Dengan demikian, diperlukan pendekatan analisis data yang lebih sistematis untuk menghasilkan klustering yang objektif, sehingga kebijakan yang diambil dengan tepat sasaran dan berdaya guna bagi seluruh pelaku UMKM.

Tabel 1. Literatur-literatur Penelitian

<i>Research Problem (RP)</i>		<i>Literature Support</i>
RP1	Lemahnya pemetaan data dan kurangnya segmentasi usaha untuk para UMKM	Banyaknya program pembinaan dan pemberdayaan pemerintah belum tepat sasaran [3], [4], [5].
RP2	Belum adanya pemetaan jelas terkait jumlah usaha dan hasil pendapatan	Banyak pelaku UMKM belum terbiasa melaporkan perkembangan usaha dan pendapatannya [6].
RP3	Adanya keterbatasan akses data dan validitas informasi	Banyak pelaku UMKM belum memiliki kemampuan atau sarana digital yang memadai untuk melaporkan data usahanya [7], [8].
RP4	Kurangnya diferensiasi program pembinaan berdasarkan karakteristik usaha	Belum adanya pemetaan data yang detail, keterbatasan sumber daya, dan lemahnya kolaborasi lintas Lembaga usaha [9].

Untuk mengatasi permasalahan tersebut, peneliti menggunakan metode K-Means digunakan sebagai pendekatan analisis data. Dengan algoritma ini, UMKM dapat dikelompokkan berdasarkan jumlah usaha dan hasil pendapatan sehingga terbentuk pengelompokan dengan merepresentasikan kategori jumlah usaha dan hasil pendapatan [10], [11]. K-Means terbukti efektif dalam mengelompokkan data berdasarkan omset dan jumlah tenaga kerja, sehingga membantu pemerintah dalam menyusun strategi pembinaan yang lebih tepat sasaran, metode ini menunjukkan bahwa prinsip K-Means memberikan gambaran distribusi usaha berdasarkan pendapatan [12], [13]. Algoritma ini lebih efektif serta efisien saat proses pengelompokan data [14], [15]. Dengan pendekatan K-Means *Clustering*, data dapat dikelompokkan ke dalam beberapa kelompok data memiliki karakteristik serupa [16], [17], [18], [19], [20], [21]. *Clustering* salah satu teknik dalam data mining digunakan untuk mengelompokkan data ke dalam beberapa klaster, disetiap klaster berisi objek dengan karakteristik serupa dan berbeda dari klaster lainnya [22], [23].

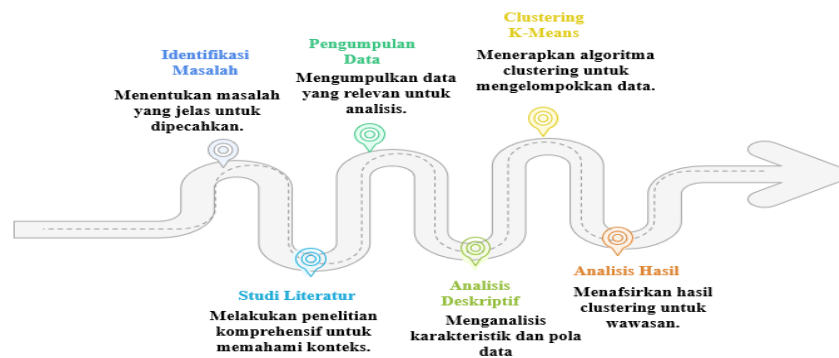
Tujuan daripada penelitian ialah menganalisis data jumlah usaha dan hasil pendapatan UMKM Provinsi DKI Jakarta untuk mengelompokkan dengan memanfaatkan metode K-means berdasarkan jumlah usaha dan hasil pendapatan, dimana hasil klasterisasi untuk memberikan

rekomendasi dasar pertimbangan dalam penyusunan kebijakan pemberdayaan UMKM yang lebih efektif.

Keterbaruan penelitian ini terletak pada implementasi algoritma *K-Means Clustering* untuk mengelompokkan UMKM Provinsi DKI Jakarta berdasarkan jumlah usaha dan hasil pendapatan. Pendekatan ini menghasilkan klaster yang lebih objektif sebagai acuan penyusunan strategi pembinaan dan kebijakan pemberdayaan. Berbeda dari penelitian terdahulu hanya berfokus pada omzet atau tenaga kerja, penelitian ini mengaitkan hasil klastering langsung dengan kebutuhan kebijakan pemerintah daerah, sehingga memiliki kontribusi akademis tetapi juga memberikan manfaat praktis bagi pemerintah daerah dalam meningkatkan efektivitas program pemberdayaan UMKM.

2. METODE PENELITIAN

Proses pengambilan data pada UMKM Provinsi DKI Jakarta yaitu berdasarkan enam wilayah Provinsi DKI Jakarta diantaranya Kabupaten Administrasi Kepulauan Seribu, Kota Administrasi Jakarta Barat, Kota Administrasi Jakarta Pusat, Kota Administrasi Jakarta Selatan, Kota Administrasi Jakarta Timur dan Kota Administrasi Jakarta Utara. Untuk data diperoleh dengan mengakses *website* Biro Pusat Statistik (BPS) <https://jakarta.bps.go.id/> mengenai UMKM Provinsi DKI Jakarta tahun 2024 berdasarkan jumlah usaha dan hasil pendapatan. Adapun langkah untuk melakukan penelitian, seperti gambar 1. sebagai berikut:



Gambar 1. Kerangka dari Penelitian.

2.1. Identifikasi Masalah

Permasalahan dalam pengelolaan UMKM Provinsi DKI Jakarta adalah lemahnya pemetaan data dan kurangnya segmentasi usaha, belum adanya pemetaan jelas terkait jumlah usaha dan tingkat pendapatan, adanya keterbatasan akses data dan validitas informasi serta kurangnya diferensiasi program pembinaan berdasarkan karakteristik usaha. Data UMKM sangat besar dan berragam mengakibatkan menyulitkan pemerintah dalam mengelompokkan skala usaha serta potensi pertumbuhannya. Kondisi ini menyebabkan berbagai program pembinaan, pelatihan, dan bantuan modal seringkali tidak tepat sasaran karena kurangnya segmentasi usaha berdasarkan karakteristik dan kapasitas masing-masing UMKM. Selain itu, keterbatasan akses data dan validitas informasi semakin menghambat pemerintah dalam menentukan prioritas intervensi. Di sisi lain, penelitian terdahulu menggunakan pendekatan *data mining* umumnya hanya berfokus pada klastering berdasarkan omzet atau tenaga kerja, namun belum banyak yang mengaitkannya secara langsung dengan strategi kebijakan pemberdayaan di tingkat daerah.

2.2. Studi Literatur

Tahapan ini melibatkan telaah terhadap *literatur*, artikel, dan hasil penelitian terkait yang menjadi dasar dalam memahami permasalahan penelitian.

2.3 Pengumpulan data

Data penelitian ini dikumpulkan dari UMKM Provinsi DKI Jakarta yang mencakup jumlah usaha dan hasil pendapatan. Data ini dipilih karena merepresentasikan kondisi UMKM, sehingga dapat digunakan sebagai dasar dalam proses pengelompokan menggunakan metode *clustering*.

2.4 Analisis Deskriptif

Tahapan analisis untuk memberikan gambaran awal terkait karakteristik data UMKM, baik dari segi jumlah usaha maupun hasil pendapatan. Pada tahap ini, data dikelompokkan menggunakan metode *K-Means Clustering* untuk menemukan pola distribusi serta membentuk klaster usaha yang lebih homogen sesuai dengan kategori tertentu.

2.5 Analisis Hasil

Pada analisa hasil berfokus pada evaluasi kualitas hasil pengelompokan. Untuk menilai validitas dan keakuratan klaster yang terbentuk digunakan metode *Davies Bouldin Index* (DBI). Nilai DBI yang rendah menunjukkan bahwa hasil pengelompokan lebih baik, karena klaster yang dihasilkan lebih homogen di dalam kelompok dan lebih berbeda antar kelompok. Pada tahap ini selanjutnya adalah tahap evaluasi yang bertujuan memastikan bahwa proses pengelompokan menggunakan algoritma K-Means berjalan secara akurat, relevan dengan data UMKM DKI Jakarta, serta mampu menghasilkan kelompok (cluster) yang bermakna untuk analisis kebijakan.

2.6 Data Mining

Data mining yakni suatu pendekatan analisis yang dipergunakan mengidentifikasi pola serta tren tersembunyi dari kumpulan data berukuran besar. Proses ini umumnya diawali dengan pemilihan data yang relevan, dilanjutkan dengan pencarian pola melalui penerapan berbagai teknik statistik dan komputasi, serta diakhiri dengan pemodelan data guna memahami fenomena yang sedang dikaji [24].

2.7 Clustering

Clustering adalah sebuah metode analisis data yang digunakan untuk mengelompokkan sekumpulan data berdasarkan kesamaan atribut tertentu, sehingga data dengan karakteristik serupa akan dikelompokkan dalam satu klaster, sementara data dengan perbedaan atribut ditempatkan pada klaster lain yang berbeda [25].

2.8 K-Means

Algoritma dasar yang umum digunakan dalam proses *clustering* meliputi beberapa tahapan, yaitu:

- Menentukan jumlah klaster dengan mengasumsikan setiap variabel sebagai satu kelompok, sehingga jumlah klaster (k) sama dengan jumlah variabel (n).
- Setelah jumlah klaster ditentukan, langkah selanjutnya adalah memilih *centroid* secara acak yang berfungsi sebagai titik pusat masing-masing klaster.
- Menghitung jarak setiap data terhadap pusat klaster menggunakan rumus jarak *Euclidean*.

Jarak *Euclidean* merupakan jarak lurus menghubungkan dua titik data yang dianalisis. Metode ini sering dipergunakan karena mampu menggambarkan jarak terpendek dalam ruang data antara dua titik yang dibandingkan. Tingkat kemiripan antar data dapat diukur melalui perhitungan jarak geometris bisa dilihat pada rumus berikut ini:

$$D_{xy} = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \dots + (X_n - Y_n)^2} \quad (1)$$

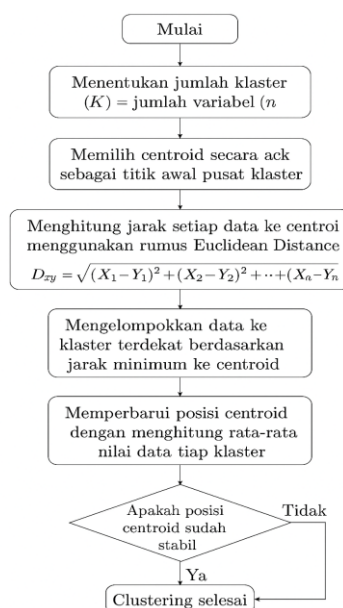
D_{xy} = Jarak data ke i ke pusat *cluster* i

X_i = Data ke i pada atribut data ke k

Y_i = Titik pusat ke y pada atribut ke k

- d. Setiap data dikelompokkan ke dalam kluster dengan jarak terdekat dari *centroid* yang ada. Proses ini dilakukan dengan membandingkan jarak data terhadap pusat setiap kluster.
- e. Posisi *centroid* kemudian diperbarui dengan menghitung nilai rerata dari seluruh data pada kluster tersebut.
- f. Jika posisi *centroid* sudah stabil atau tidak mengalami perubahan, maka proses *clustering* dianggap selesai. Namun, apabila masih terjadi pergeseran, maka iterasi dilanjutkan hingga tercapai kondisi konvergen.

Untuk lebih jelasnya tahapan Algoritma dalam proses *clustering* dapat dilihat pada gambar 2. dibawah ini:



Gambar 2. Tahapan Algoritma dalam Proses *Clustering*.

3. HASIL DAN PEMBAHASAN

Data UMKM dari Provinsi DKI Jakarta tahun 2024 dijadikan sumber utama dalam penelitian ini. Pengelompokan dilakukan menggunakan algoritma *K-Means*, diawali dengan tahap penetapan data sebagai objek yang akan dikelompokkan. Berikut adalah penetapan data sebagai objek yang akan dikelompokkan dijelaskan ditabel 2. berikut ini:

Tabel 2. Data Jumlah Usaha dan Hasil Pendapatan
Provinsi DKI Jakarta Tahun 2024

No	Wilayah	Jumlah Usaha	Hasil Pendapatan
1	Kabupaten Administrasi Kepulauan Seribu	32235	191888308196
2	Kota Administrasi Jakarta Barat	269602	942631160211
3	Kota Administrasi Jakarta Pusat	302058	1238854060881
4	Kota Administrasi Jakarta Selatan	511552	1599548491227
5	Kota Administrasi Jakarta Timur	406645	1350676087042
6	Kota Administrasi Jakarta Utara	239326	656426223662

Tabel 3. dibawah ini menjelaskan bagaimana melakukan perhitungan dengan algoritma *K-Means*. Tahapan berikut dilakukan adalah hasil dikumpulkan untuk jumlah *cluster*. Ada tiga *cluster* dari enam data. Selanjutnya, nilai *centroid* dipilih dengan *random*.

Tabel 3. *Centroid* Data Awal

Wilayah	Jumlah Usaha	Hasil Pendapatan
C1 (Kabupaten Adm. Kepulauan Seribu)	32235	191888308196
C2 (Kota Administrasi Jakarta Barat)	269602	942631160211
C3 (Kota Administrasi Jakarta Pusat)	302058	1238854060881

Penghitungan *centroid* paling dekat dengan penentuan pusat kelompok terdekat pada tiap data yang ada. Setelah itu, masing-masing data dialokasikan ke dalam kelompok dengan pusat terdekat. Tahap berikutnya adalah menghitung jarak antara setiap data dengan seluruh pusat kelompok menggunakan rumus berikut:

$$D_{xy} = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \dots (X_n - Y_n)^2} \quad (2)$$

Penghitungan *centroid* terdekat:

a. Kabupaten Administrasi Kepulauan Seribu

D_{11}

$$D_{11} = \sqrt{(32235 - 32235)^2 + (191888308196 - 191888308196)^2}$$

$D_{11}=0,00$

D_{12}

$$D_{12} = \sqrt{(32235 - 269602)^2 + (1918883081 - 96942631160211)^2}$$

$D_{12}= 750742852015.04$

D_{13}

$$D_{13} = \sqrt{(32235 - 302058)^2 + (191888308196 - 1238854060881)^2}$$

$D_{13}= 1046965752685.03$

b. Kota Administrasi Jakarta Barat

D_{21}

$$D_{21} = \sqrt{(269602 - 32235)^2 + (942631160211 - 1238854060881)^2}$$

$D_{21}= 750742852015.04$

D_{22}

$$D_{22} = \sqrt{(269602 - 269602 + (942631160211 - 942631160211)^2}$$

$D_{22}=0,00$

D_{33}

$$D_{33} = \sqrt{(269602 - 302058)^2 + (942631160211 - 1238854060881)^2}$$

$D_{33}= 296222900670.00$

Tabel 4 menunjukkan tahapan awal yang dipergunakan dalam proses pengelompokan sebagai dasar perhitungan untuk memperoleh hasil akhir dari keseluruhan analisis.

Tabel 4. Hasil Penghitungan Jarak Pusat *Cluster* Iterasi

No	Provinsi	Cluster			Jarak Terpendek
		C1	C2	C3	
1	Kabupaten Administrasi Kepulauan Seribu	0.00	750742852015.04	1046965752685.03	0.00
2	Kota Administrasi Jakarta Barat	750742852015.04	0.00	296222900670.00	0.00
3	Kota Administrasi Jakarta Pusat	1046965752685.03	296222900670.00	0.00	0.00
4	Kota Administrasi Jakarta Selatan	1407660183031.08	656917331016.05	360694430346.06	360694430346.06
5	Kota Administrasi Jakarta Timur	1158787778846.06	408044926831.02	111822026161.05	111822026161.05
6	Kota Administrasi Jakarta Utara	464537915466.05	286204936549.00	582427837219.00	286204936549.00

Penentuan *centroid* baru dilakukan dengan menghitung total nilai dari setiap klaster yang telah terbentuk, kemudian membaginya dengan jumlah anggota dalam klaster tersebut. Tabel 5. menunjukkan proses dimulainya iterasi kedua yang didasarkan pada hasil penghitungan jarak masing-masing objek pada iterasi pertama.

Tabel 5. *Centroid* Data Iterasi 2

Wilayah	Jumlah Usaha	Hasil Pendapatan
C1 (Kabupaten Administrasi Kepulauan Seribu)	32235	191888308196
C2 (Kota Administrasi Jakarta Pusat)	302058	1238854060881
C3 (Kota Administrasi Jakarta Selatan)	511552	1599548491227

Pada tabel 6. menunjukkan hasil dari pengelompokan daripada langkah 1. Tabel tersebut menunjukkan hasil perhitungan jarak terpendek setiap wilayah di Provinsi DKI Jakarta terhadap tiga *centroid* (C1, C2, dan C3) pada proses *clustering* menggunakan algoritma K-Means. Nilai jarak terpendek menentukan keanggotaan masing-masing wilayah dalam suatu klaster.

Tabel 6. Hasil Perhitungan Jarak Pusat *Cluster* Iterasi 2

No	Provinsi	Cluster			Jarak Terpendek
		C1	C2	C3	
1	Kabupaten Administrasi Kepulauan Seribu	0.00	1046965752685.03	1407660183031.08	0.00
2	Kota Administrasi Jakarta Barat	750742852015.04	296222900670.00	656917331016.05	296222900670.00
3	Kota Administrasi Jakarta Pusat	1046965752685.03	0.00	360694430346.06	0.00
4	Kota Administrasi Jakarta Selatan	1407660183031.08	360694430346.06	0.00	0.00
5	Kota Administrasi Jakarta Timur	1158787778846.06	111822026161.05	248872404185.02	111822026161.05
6	Kota Administrasi Jakarta Utara	464537915466.05	582427837219.00	943122267565.04	464537915466.05

Data UMKM Provinsi DKI Jakarta dikelompokkan berdasarkan jarak terdekat ke pusat *cluster*. Data yang paling dekat dengan *centroid* akan masuk ke dalam *cluster* tersebut. Posisi data pada iterasi kedua tercantum dalam tabel 7. dibawah ini:

Tabel 7. Posisi Data Dengan Tiap *Cluster* Pada Iterasi Ke-2

No	Provinsi	Cluster		
		C1	C2	C3
1	Kabupaten Administrasi Kepulauan Seribu	1		
2	Kota Administrasi Jakarta Barat		1	

3	Kota Administrasi Jakarta Pusat		1	
4	Kota Administrasi Jakarta Selatan			1
5	Kota Administrasi Jakarta Timur		1	
6	Kota Administrasi Jakarta Utara	1		

Penetapan *centroid* baru dilakukan dengan melakukan penjumlahan seluruh data per klaster yang telah ada, kemudian hasilnya dibagi dengan jumlah anggota dalam klaster tersebut. Setelah perhitungan jarak antar objek pada iterasi pertama selesai, proses dilanjutkan ke iterasi berikutnya berdasarkan hasil perhitungan yang ditampilkan pada Tabel 8.

Tabel 8. *Centroid* Data Iterasi 3

Wilayah	Jumlah Usaha	Hasil Pendapatan
C1 (Kabupaten Adm. Kepulauan Seribu)	32235	191888308196
C2 (Kota Administrasi Jakarta Timur)	406645	1350676087042
C3 (Kota Administrasi Jakarta Utara)	239326	656426223662

Hasil dari iterasi 3 di Tabel 9. bahwa proses pengelompokan mulai menunjukkan di mana pembagian wilayah berdasarkan kesamaan karakteristik jumlah usaha dan hasil pendapatan mulai terbentuk dengan jelas. Hal ini menandakan bahwa *centroid* baru telah menghasilkan pengelompokan yang lebih representatif terhadap data UMKM di setiap wilayah Propinsi DKI Jakarta dengan menunjukkan hasil dari pengelompokan, yang dapat dilihat sebagai berikut:

Tabel 9. Hasil Perhitungan Jarak Pusat *Cluster* Iterasi 3

No	Provinsi	Cluster			Jarak Terpendek
		C1	C2	C3	
1	Kabupaten Administrasi Kepulauan Seribu	0.00	1158787778846.06	464537915466.05	0.00
2	Kota Administrasi Jakarta Barat	750742852015.04	408044926831.02	286204936549.00	286204936549.00
3	Kota Administrasi Jakarta Pusat	1046965752685.03	111822026161.05	582427837219.00	111822026161.05
4	Kota Administrasi Jakarta Selatan	1407660183031.08	248872404185.02	943122267565.04	248872404185.02
5	Kota Administrasi Jakarta Timur	1158787778846.06	0.00	694249863380.02	0.00
6	Kota Administrasi Jakarta Utara	464537915466.05	694249863380.02	0.00	0.00

Melakukan penentuan posisi data UMKM Provinsi DKI Jakarta dengan jarak data yang paling sedikit daripada pusat *cluster*. Data kemudian tergabung dalam suatu kelompok jikalau jaraknya paling dekat dengan *centroid* tersebut. Tabel.10 memperlihatkan posisi data pada masing-masing kelompok dalam iterasi ketiga.

Tabel 10. Posisi Data Dengan Tiap *Cluster* Pada Iterasi Ke-3

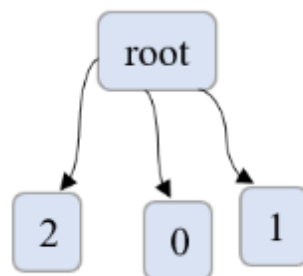
No	Provinsi	Cluster		
		C1	C2	C3
1	Kabupaten Administrasi Kepulauan Seribu	1		
2	Kota Administrasi Jakarta Barat			1

3	Kota Administrasi Jakarta Pusat		1	
4	Kota Administrasi Jakarta Selatan		1	
5	Kota Administrasi Jakarta Timur		1	
6	Kota Administrasi Jakarta Utara			1

Cluster pertama menggambarkan wilayah jumlah usaha dan hasil pendapatan dengan pertumbuhan ekonomi sedang yakni Kota Administrasi Jakarta Barat dan Kota Administrasi Jakarta Utara. *Cluster* kedua mempunyai jumlah usaha dan hasil pendapatan kategori tertinggi pertumbuhan ekonomi adalah Kota Administrasi Jakarta Pusat, Kota Administrasi Jakarta Selatan dan Kota Administrasi Jakarta Timur. *Cluster* 0 merepresentasikan wilayah dengan perekonomian rendah yakni Kabupaten Administrasi Kepulauan Seribu.

Dalam proses analisis data, peneliti menggunakan aplikasi RapidMiner sebagai alat bantu untuk melakukan pengelompokan data (*clustering*) menggunakan metode K-Means. Melalui tahapan analisis ini, diperoleh hasil berupa pembentukan tiga kluster utama, yaitu *Cluster* 0, *Cluster* 1, dan *Cluster* 2. Ketiga kluster tersebut mewakili kategori tingkat karakteristik yang berbeda, di mana *Cluster* 0 menunjukkan kelompok dengan nilai terendah, *Cluster* 1 merepresentasikan kelompok dengan nilai sedang, sedangkan *Cluster* 2 menggambarkan kelompok dengan nilai tertinggi.

Proses penentuan posisi data dalam setiap kluster dilakukan secara sistematis pada langkah ketiga analisis, yaitu saat RapidMiner menghitung jarak antar data terhadap pusat kluster (*centroid*). Berdasarkan hasil perhitungan jarak tersebut, setiap data secara otomatis dialokasikan ke dalam kluster yang memiliki kedekatan paling tinggi dengan pusatnya. Dengan demikian, hasil pengelompokan ini memberikan gambaran yang lebih jelas mengenai distribusi dan karakteristik data UMKM berdasarkan wilayah usaha dan hasil pendapatan di wilayah DKI Jakarta. Untuk selanjutnya pada Gambar 3. menunjukkan seberapa besar perbandingan antara *cluster* dengan semua data, yang dihasilkan dari pemecahan data induk atau *root set* yang menghasilkan turunan.



Gambar 3. *Root Set*.

Untuk menilai hasil klasterisasi serta mengevaluasi kualitas dari kluster yang terbentuk, digunakan *Davies Bouldin Index* (DBI). Indeks ini bertujuan untuk mengukur tingkat efektivitas metode evaluasi dalam menilai hasil proses klasterisasi pada data mining. Nilai yang diperoleh dari perhitungan *Davies Bouldin* menunjukkan seberapa baik kluster yang dihasilkan oleh metode tersebut. Hasil perhitungan menunjukkan bahwa evaluasi menggunakan *Davies Bouldin Index* (DBI) menghasilkan nilai sebesar -0.390. Semakin mendekati nol, nilai DBI menunjukkan kualitas kluster yang semakin baik. Dengan demikian, hasil tersebut mengindikasikan bahwa evaluasi kluster K-Means yang diperoleh tergolong baik. Untuk hasilnya pada gambar 4. dibawah ini:

```

PerformanceVector

PerformanceVector:
Avg. within centroid distance: -18189556096655096000000.000
Avg. within centroid distance_cluster_0: -0.000
Avg. within centroid distance_cluster_1: -20478316426254510000000.000
Avg. within centroid distance_cluster_2: -22726901242473850000000.000
Davies Bouldin: -0.390
    
```

Gambar 4. Hasil *Cluster Distance Performance* 3 Klaster.

4. KESIMPULAN

Metode K-Means dapat dipergunakan untuk mengelompokkan UMKM di DKI Jakarta berdasarkan jumlah usaha dan hasil pendapatan. Hasil klasterisasi menghasilkan tiga klaster wilayah yakni Klaster pertama Kota administratif Jakarta Pusat, Jakarta Selatan, dan Jakarta Timur, memiliki jumlah usaha dan pendapatan tertinggi. Klaster kedua Kota administratif Jakarta Barat dan Jakarta Utara, memiliki jumlah usaha dan pendapatan sedang. Klaster ketiga Kota administratif Kepulauan Seribu, memiliki jumlah usaha dan pendapatan terendah. Hasil evaluasi proses klasterisasi mempergunakan *Davies Bouldin Index (DBI)* bernilai -0.390. Hasil ini menunjukkan K-Means dapat memetakan jumlah usaha dan hasil pendapatan. Hasil Klastering bisa memberikan wawasan dalam memahami karakteristik dan kebutuhan UMKM, sehingga dapat membantu pemerintah dan pihak terkait dalam merumuskan kebijakan dan program dengan tepat sasaran untuk pengembangan UMKM. Penelitian selanjutnya dapat dilakukan dengan menggunakan data yang lebih lengkap dan fitur yang lebih kompleks, serta menggunakan metode *clustering* lainnya untuk membandingkan hasil klastering.

DAFTAR PUSTAKA

- [1] H. Kholisoh, B. P. Priyadi, and H. Purnaweni, "Implementasi Program Jakpreneur Dalam Pengembangan Usaha Mikro Kecil dan Menengah (UMKM) Di Kota Administrasi Jakarta Pusat," *J. Monas*, vol. 17, pp. 1–20, 2021, doi: <https://doi.org/10.14710/jppmr.v13i2.43688>.
- [2] E. F. Lamaile, "Pengaruh Pertumbuhan UMKM, Indeks Keterbukaan Perdagangan, Pertumbuhan Pendapatan Per Kapita, dan Indeks Pembangunan Manusia Terhadap Pengentasan Kemiskinan Di 5 Negara ASEAN," *J-REMA (Jurnak Ris. Ekon. Manaj. dan Akuntansi)*, vol. 1, pp. 167–186, 2022, doi: <https://doi.org/10.25170/jrema.v1i3.4593>.
- [3] D. A. Ariani and F. Nugroho, "mplementation of Integrated Entrepreneurship Development Policy within the Sustainable Livelihoods Framework for Vulnerable Groups in DKI Jakarta Province," 2025. doi: <https://doi.org/10.63922/ajmesc.v5i02.1357>.
- [4] S. B. M. Putri, S. Raudah, and A. Maryati, "Implementasi Program Pemberdayaan, Usaha Menengah, Usaha Kecil dan Usaha Mikro Kecil (UMKM) Pada Dinas Koperasi Usaha Kecil Menengah Perindustrian dan Perdagangan Kabupaten Hulu Sungai Utara," *J. MSDM (Manajemen Sumber Daya Manusia)*, vol. 2, pp. 114–127, 2025.
- [5] Sudrajat, T. Hidayat, F. Wulandari, R. Multinsyah, and A. A. Jufry, "Optimalisasi Digital Marketing untuk Pengembangan UMKM Citruk di Desa Ranca Iyuh, Kabupaten Tangerang," *Abdi Dharma*, vol. 5, no. 2, pp. 15–30, 2025, doi: [10.31253/ad.v5i1.3561](https://doi.org/10.31253/ad.v5i1.3561).
- [6] S. Sari, "Strategi Pengembangan Usaha Mikro Kecil dan Menengah Berbasis Collaborative Governance Desa Pananrang Kecamatan Mattiro Bulu," *J. Manaj. dan Perbank. Syariah*, vol. 6, pp. 16–29, 2024.
- [7] I. Iin, R. Fadila, A. Rizki Rinaldi, and F. Fathurrohman, "Penerapan Data Mining Dalam Mengelompokkan Jumlah Umkm Berdasarkan Kabupaten Kota Menggunakan K-Means

- Clustering,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1446–1450, 2024, doi: 10.36040/jati.v8i2.8427.
- [8] Triyanto, M. Nuryatno, Susi Dwi Mulyani, Erie Riza Nugraha, Falendra Jovanka, and Edeltrudis Ertin Resi, “Transformasi Digital – Strategi Umkm Untuk Berdaptasi Di Era Industri 4.0 Melalui Pelatihan Dan Pendampingan,” *J. Abdikaryasakti*, vol. 5, no. 1, pp. 25–42, 2025, doi: 10.25105/v5i1.22611.
 - [9] P. Puntoriza and C. Fibriani, “Analisis Persebaran UMKM Kota Malang Menggunakan Cluster K-means,” *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 86–94, 2020, doi: 10.33633/joins.v5i1.3469.
 - [10] U. Linarti, A. H. Soleliza Jones, L. Zahrotun, and A. Rahmawati, “Penerapan Metode K-Medoids Guna Pengelompokan Data Usaha Mikro, Kecil dan Menengah (UMKM) Bidang Kuliner Di Kota Yogyakarta,” 2024. doi: 10.55338/jikomsi.v7i1.2194.
 - [11] N. Hafizah, A. Lia Hananto, F. Nurapriani, and E. Novalia, “Segmentasi Nasabah Umkm Berdasarkan Kinerja Dan Keuntungan Menggunakan K-Means Clustering,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 5, pp. 8661–8665, 2025, doi: 10.36040/jati.v9i5.15056.
 - [12] Nursyifa and R. N. Fahmi, “Implementasi Metode K-Means Clustering dalam Analisis Persebaran UMKM Di Jawa Barat,” *J. Inf. Syst.*, vol. 6, pp. 211–220, 2021.
 - [13] A. Fauzan, N. Suarna, I. Ali, and H. Susana, “Penerapan Algoritma K-Means Clustering Untuk Meningkatkan Model Pengelompokan Dan Kinerja Jaringan Wi-Fi Secara Optimal,” *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 2, 2025, doi: 10.23960/jitet.v13i2.6272.
 - [14] F. Handayani, “Aplikasi Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokan Mahasiswa Berdasarkan Gaya Belajar,” *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 46–63, 2022, doi: 10.34010/jati.v12i1.6733.
 - [15] N. Bili, R. Thimotius Abineno, and A. A. Pekuwali, “Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Performa Siswa Pada Pembelajaran Bahasa Indonesia(Studi Kasus: SD INPRES WAINGAPU 3),” *SATI Sustain. Agric. Technol. Innov.*, pp. 523–537, 2024.
 - [16] E. A. Saputra and Y. Nataliani, “Analisis Pengelompokan Data Nilai Siswa Untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means,” *J. Inf. Syst. Informatics*, vol. 3, pp. 424–439, 2021.
 - [17] Y. Pratiwi and Mulyawan, “Implementasi Algoritma K-Means untuk Menentukan Angka Harapan Hidup Berdasarkan Tingkat Propinsi,” *J. Tek.*, vol. 1, pp. 1–11, 2023, doi: <https://doi.org/10.56211/blendsains.v1i4.233>.
 - [18] D. Marcelina, A. Kurnia, and Terttiaavini, “Analisis Klaster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering,” *MALCOM Indones. J. Mach. Learn. Comput. Sciene*, vol. 3, pp. 293–301, 2023, doi: <https://doi.org/10.57152/malcom.v3i2.952>.
 - [19] F. Khalish, N. M. Piranti, and O. Martadireja, “Implementasi Data Mining Menggunakan Teknik Clustering dengan Metode K-Means,” *JiIP (Jurnal Ilm. Ilmu Pendidikan)*, vol. 8, pp. 5392–5397, 2025, doi: <https://doi.org/10.54371/jiip.v8i5.7874>.
 - [20] Q. Adawiyah, S. Defit, and Sumijan, “Penerapan Algoritma K-Means Clustering untuk Mengelompokkan Rekomendasi Metode Kontrasepsi Berbasis Machine Learning,” *J. KomtekInfo*, vol. 11, no. 4, pp. 300–305, 2024, doi: 10.35134/komtekinfo.v11i4.563.
 - [21] F. Handayanna and Sunarti, “Penerapan K-Means untuk Mengklasifikasikan Tingkat Keterampilan Teknologi Informasi dan Komunikasi di Pulau Jawa,” *J. Inf. Syst.*, vol. 10, pp. 71–80, 2025, doi: <https://doi.org/10.33633/joins.v10i1.12917>.
 - [22] W. Sudrajat, I. Cholid, and J. Petrus, “Penerapan Algoritma K-Means Clustering untuk Pengelompokan UMKM Menggunakan Rapidminer,” *JUPITER*, vol. 14, 2022, doi: <https://doi.org/10.5281/4467/5.jupiter.2022.04>.
 - [23] M. Baha’uddin and Z. Fatah, “Penerapan Data Mining Clustering K-Means Dalam Mengelompokkan Data Penduduk Penyandang Disabilitas,” *Gudang J. Multidisiplin Ilmu*,

- vol. 2, no. 10, pp. 86–94, 2024, [Online]. Available: <https://gudangjurnal.com/index.php/gjmi>
- [24] N. Ameliana, N. Suarna, and W. Prihaartono, “Analisis Data Mining Pengelompokan UMKM Menggunakan Algoritma K-Means Clustering Di Provinsi Jawa Barat,” 2024.
- [25] M. Mustika, Y. Ardilla, A. Manuhutu, N. Ahmad, and Imanuddin, *Data Mining dan Aplikasinya*, Pertama. Bandung: Widia Bhakti Persada Bandung, 2021.