Jurnal Jaring SainTek
Vol.8, No.1, April 2026, pp. 37 - 44
e-ISSN : 2656-9485

Analisis *Clustering* Pelaku Usaha UMKM Kota Bekasi Menggunakan Algoritma K-Means

Michelle Ledisty Aisha[1], Rafika Sari[2,*], Ratna Salkiawati[3]
[1,2,3]Informatika, Ilmu Komputer, Universitas Bhayangkara Jakarta Raya, Indonesia
e-mail: [1]202110715254@mhs.ubharajaya.ac.id, [2,*]**rafika.sari@dns.ubharajaya.ac.id,**
[3]ratna_tind@dsn.ubharajaya.ac.id

Abstract
Micro, Small, and Medium Enterprises (MSMEs) are the backbone of Indonesia's economy, employing over 97% of the workforce. However, the absence of data-driven classification hampers the formulation of effective support policies. This study aims to cluster MSMEs in Bekasi City based on business capital, revenue, and subdistrict location using the K-Means Clustering algorithm. A total of 7,923 entries were analyzed after cleaning the initial 7,964 dataset. The optimal number of clusters was determined using the Davies-Bouldin Index (DBI), with the best score of 0.5940 achieved at K=3. The clustering was conducted in two stages: manual calculation on a sample of 138 data points and automated processing via Python for the full dataset. The manual process converged at the 3rd iteration, where each step involved calculating Euclidean distances to cluster centroids, followed by centroid updates based on the mean of cluster members. The results classified MSMEs into three categories: small-scale, medium-scale, and large-scale enterprises. Small-scale MSMEs, typically with capital and revenue below IDR 5 million, are concentrated in Mustika Jaya, Pondok Gede, and Bantar Gebang. In contrast, large-scale MSMEs with higher financial figures are mostly found in Bekasi Selatan, Medan Satria, and Rawalumbu. This clustering model offers a practical foundation for designing more targeted, spatially informed, and adaptive MSME development policies.

Keywords: *K-Means, clustering, MSMEs, Bekasi*

Abstrak
Usaha Mikro, Kecil, dan Menengah (UMKM) merupakan pilar ekonomi nasional yang menyerap lebih dari 97% tenaga kerja. Namun, belum adanya klasifikasi yang berbasis data terhadap pelaku UMKM menyulitkan penyusunan kebijakan pembinaan yang tepat. Penelitian ini bertujuan untuk mengelompokkan UMKM di Kota Bekasi berdasarkan modal, omzet, dan lokasi kecamatan menggunakan algoritma K-Means Clustering. Data sebanyak 7.923 baris digunakan setelah proses pembersihan dari 7.964 data awal. Penentuan jumlah cluster optimal menggunakan metode Davies-Bouldin Index (DBI) yang menghasilkan nilai terbaik pada K=3 (DBI = 0,5940). Proses clusterisasi dilakukan secara dua tahap, yaitu perhitungan manual terhadap 138 sampel data dan perhitungan otomatis menggunakan Python untuk seluruh data. Pada perhitungan manual, proses iterasi menunjukkan kondisi konvergen pada iterasi ke-3. Setiap iterasi melibatkan kalkulasi jarak Euclidean antara data dan centroid, diikuti pembaruan centroid berdasarkan rata-rata nilai anggota cluster. Hasil clusterisasi membagi UMKM menjadi tiga kategori: skala kecil, menengah, dan besar. UMKM skala kecil umumnya memiliki modal < Rp5 juta dan omzet < Rp5 juta, serta tersebar di Mustika Jaya, Pondok Gede, dan Bantar Gebang. Sementara itu, UMKM skala besar dominan di Bekasi Selatan, Medan Satria, dan Rawalumbu dengan modal dan omzet relatif tinggi. Model pengelompokan ini diharapkan dapat digunakan sebagai dasar dalam perumusan kebijakan pembinaan UMKM berbasis spasial yang lebih efisien, adaptif, dan tepat sasaran.

Kata Kunci: *K-Means, clustering, UMKM, Bekasi*

PENDAHULUAN

Usaha Mikro, Kecil, dan Menengah (UMKM) memiliki peranan strategis dalam perekonomian Indonesia, dengan kontribusi terhadap Produk Domestik Bruto (PDB) nasional mencapai lebih dari 60% dan menyerap lebih dari 97% tenaga kerja [1]. Kota Bekasi sebagai salah satu pusat pertumbuhan ekonomi di Jawa Barat menunjukkan potensi besar dalam pengembangan UMKM. Namun, perkembangan tersebut belum merata di seluruh kecamatan, yang ditandai dengan ketimpangan jumlah pelaku usaha, variasi besar modal, serta omzet yang tidak seragam.

Ketimpangan tersebut menjadi tantangan dalam penyusunan kebijakan pembinaan dan pengembangan UMKM yang tepat sasaran. Salah satu pendekatan yang dapat digunakan untuk mengatasi permasalahan ini adalah dengan melakukan analisis pengelompokan (clustering) berbasis data untuk memahami karakteristik pelaku usaha berdasarkan modal dan omzet. Pengelompokan ini memungkinkan pemangku kebijakan untuk memetakan segmen pelaku UMKM secara lebih objektif, sehingga strategi intervensi dapat

Submitted: **19/07/2025**; Revised: **19/08/2025**; Accepted: **19/09/2025**; Published: **30/04/2026**

disesuaikan dengan kebutuhan masing-masing kelompok [2].

Algoritma K-Means Clustering dipilih dalam penelitian ini karena kemampuannya dalam mengelompokkan data numerik secara efisien berdasarkan kesamaan karakteristik [3]. Selain itu, metode Davies-Bouldin Index (DBI) digunakan untuk mengevaluasi jumlah cluster optimal, guna menghasilkan segmentasi UMKM yang lebih representatif [4]. Penelitian ini menggunakan data sekunder dari Dinas UMKM Kota Bekasi tahun 2019–2024, dengan fokus pada tiga variabel utama: modal, omzet, dan kecamatan tempat usaha.Tujuan dari penelitian ini adalah untuk membangun model klasifikasi yang mampu mengelompokkan pelaku UMKM di Kota Bekasi ke dalam beberapa cluster berdasarkan karakteristik finansial dan geografis. Hasil pengelompokan ini diharapkan dapat memberikan kontribusi dalam proses pengambilan keputusan strategis oleh pemerintah daerah, khususnya dalam mendukung keberlangsungan dan pertumbuhan UMKM secara lebih merata dan berkelanjutan.

Clustering adalah metode dalam pembelajaran tidak terawasi (*unsupervised learning*) di mana data yang dianalisis belum memiliki label atau target kelas tertentu. Tujuan dari proses ini adalah untuk mengelompokkan objek-objek seperti data, sampel, atau observasi ke dalam cluster berdasarkan pola atau karakteristik yang serupa. Objek-objek tersebut dikelompokkan ke dalam kelas atau kelompok tertentu melalui proses pengelompokan, baik secara fisik maupun abstrak, berdasarkan tingkat kesamaan dan kedekatannya.

Algoritma K-Means merupakan salah satu metode clusterisasi tertua dan paling umum digunakan, khususnya dalam skala aplikasi kecil hingga menengah karena kesederhanaan implementasinya. Sejak tahun 1950-an, algoritma ini telah banyak diteliti oleh para ahli seperti Lloyd (1957, 1982), Forgey (1965), Friedman dan Rubin (1967), serta MacQueen (1967). Prinsip dasar dari algoritma ini adalah meminimalkan jumlah kuadrat kesalahan (Sum of Squared Errors/SSE) antara titik data dengan sejumlah pusat cluster (centroid). Prosesnya diawali dengan pemilihan sejumlah centroid secara acak sebanyak *k*, kemudian setiap data ditempatkan ke cluster terdekat berdasarkan jarak tertentu. Setelah itu, posisi centroid diperbarui berdasarkan rata-rata nilai data dalam cluster tersebut. Langkah ini diulang hingga posisi centroid tidak lagi berubah atau telah mencapai kondisi konvergen. [4]

METODE PENELITIAN

Penelitian ini menggunakan pendekatan data mining dengan algoritma K-Means Clustering untuk melakukan pengelompokan terhadap pelaku UMKM di Kota Bekasi berdasarkan tiga variabel utama, yaitu modal usaha, omzet, dan lokasi kecamatan. Data yang digunakan merupakan data sekunder yang diperoleh dari Dinas UMKM Kota Bekasi, mencakup periode tahun 2019–2024 dengan total sebanyak 7.964 baris data, yang kemudian dibersihkan menjadi 7.923 data valid setelah proses *pre-processing*.

Metodologi penelitian mengikuti tahapan *Knowledge Discovery in Database* (KDD) yang terdiri atas: (i) ***Data Collection***, Pengumpulan data UMKM dari instansi pemerintah daerah yang mencakup informasi modal, omzet, dan kecamatan; (ii) ***Data Preparation***, Proses ini meliputi pembersihan data dari nilai kosong (missing value), penghapusan data duplikat, dan normalisasi data numerik agar berada pada skala yang seragam; (iii) ***Data Mining***, proses penting dimana metode cerdas diterapkan untuk mengidentifikasi pola atau membentuk model; (iv) ***Evaluation Model***, untuk mengidentifikasi pola atau model yang benar-benar menarik yang merepresentasikan pengetahuan; dan (v) ***Knowledge Presentation***, proses visualisasi dan teknik representasi pengetahuan untuk menyampaikan hasil analisis data kepada pengguna. [3]

K-Means dimulai dengan menentukan jumlah pusat *cluster* (*centroid*), kemudian mengelompokkan setiap data ke *centroid* terdekat berdasarkan jarak. Setelah pengelompokan awal, algoritma akan menghitung ulang posisi *centroid* berdasarkan rata-rata data dari setiap *cluster*. Proses ini diulang hingga posisi centroid tidak berubah secara signifikan atau hingga kondisi konvergen tercapai [5]. Langkah-langkah cara kerja K-Means:

a. Tentukan jumlah cluster (K) yang diinginkan.
b. Pilih centroid awal secara acak ambil K titik data secara acak sebagai pusat cluster awal.

c. Kelompokkan data berdasarkan jarak terdekat ke masing-masing centroid menggunakan jarak Euclidean.

$$d(p,q)\sqrt{((p_1-q_1)^2+(p_1-q_2)^2+\cdots+(p_n-q_n)^2} \quad (1)$$

d. Hitung ulang centroid setiap cluster sebagai rata-rata dari seluruh data dalam cluster.

$$M_K=\left(\frac{1}{n_k}\right)\sum_{i=1}^{n_k} x_{ik} \quad (2)$$

e. Hitung kesalahan kuadrat dalam cluster (disebut within-cluster variation / WCV), yaitu total jarak kuadrat antara setiap data dengan pusat cluster-nya. Nilai ini menunjukkan seberapa rapat atau tersebar data dalam satu cluster

$$e_k^2=\sum_{i=1}^{n_k}(x_{ik}-M_k)^2 \quad (3)$$

f. Hitung total kesalahan dari seluruh cluster

$$e_k^2=\sum_{i=1}^{n_k}(x_{ik}-M_k)^2 \quad (4)$$

g. Ulangi proses pengelompokan dan perhitungan centroid secara berulang hingga posisi centroid stabil (konvergen) [6].

Selanjutnya untuk mengevaluasi model yang telah dibuat, akan dihitung metode Davies-Bouldin Index (DBI) yang diperkenalkan oleh David L.Davies dan Donald W. Bouldin pada tahun 1979. Metode yang digunakan untuk mengevaluasi kualitas cluster. DBI dihitung dengan membandingkan rata-rata jarak dalam cluster dengan jarak antar cluster terdekat, sehingga membantu mengukur seberapa baik data dikelompokkan [5]. Davies-Bouldin Index (DBI) merupakan salah satu metrik yang digunakan untuk mengevaluasi kualitas hasil *clustering* dengan melihat seberapa baik objek-objek dikelompokkan berdasarkan kesamaan karakteristik atau fitur dalam data.[7]

$$DBI=\frac{1}{N}\sum_{i=0}^{n}\max R_{ij}(5) \quad (5)$$

Nilai dalam persamaan (5) dihitung berdasarkan dua aspek utama, yakni tingkat kedekatan antar data dalam satu cluster (kohesi) dan jarak antar cluster yang berbeda (separasi). Pengelompokan yang optimal ditandai dengan kohesi yang rendah dan separasi yang tinggi.

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan algoritma K-Means Clustering dalam proses pemodelan untuk mengelompokkan data pelaku Usaha Mikro, Kecil, dan Menengah (UMKM) di Kota Bekasi berdasarkan tiga variabel utama, yaitu modal usaha, omzet, dan lokasi kecamatan. Proses pemodelan dilakukan dengan memanfaatkan platform Google Colab menggunakan bahasa pemrograman Python, serta mengikuti kerangka kerja Knowledge Discovery in Database (KDD) yang meliputi tahapan seleksi data, praproses data, transformasi, data mining, dan evaluasi.

Tahapan pemodelan dimulai dengan menentukan jumlah cluster (K) yang optimal menggunakan metode Elbow Method dan Davies-Bouldin Index (DBI) sebagai metrik evaluasi. Data UMKM yang digunakan berasal dari Dinas UMKM Kota Bekasi periode 2019–2024, dengan total sebanyak 7.923 entri, serta dilakukan juga perhitungan manual terhadap 138 sampel untuk memastikan konvergensi dan validasi hasil pemodelan.

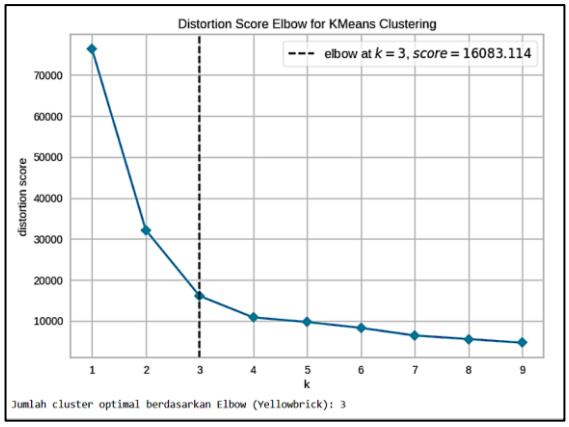
Setelah nilai K ditentukan, algoritma K-Means menginisialisasi centroid secara acak, kemudian menghitung jarak setiap data ke masing-masing centroid menggunakan metode Euclidean Distance, menentukan keanggotaan cluster berdasarkan jarak terdekat, dan memperbarui posisi centroid pada setiap iterasi hingga mencapai kondisi konvergen. Dalam proses manual, konvergensi tercapai pada iterasi ke-3, sedangkan pada proses komputasi dengan keseluruhan data, iterasi berlangsung hingga 10 kali.

Penerapan algoritma K-Means bertujuan untuk mengidentifikasi pola distribusi UMKM berdasarkan karakteristik finansial dan geografis, sehingga hasil pengelompokan dapat memberikan informasi strategis dalam pengambilan kebijakan pemberdayaan dan pengembangan UMKM di Kota Bekasi. Nilai Davies-Bouldin Index (DBI) sebesar 0,5940 menunjukkan bahwa hasil clusterisasi berada dalam kategori cukup baik, di mana semakin rendah nilai DBI, semakin baik kualitas pemisahan antar cluster.

Perhitungan Elbow Method

Submitted: **19/07/2025**; Revised: **19/08/2025**; Accepted: **19/09/2025**; Published: **30/04/2026**

Dalam penelitian ini digunakan *elbow method* untuk menentukan jumlah cluster yang optimal dalam proses pengelompokan UMKM. Cara kerja *Elbow Method* yaitu dengan melakukan iterasi pada nilai K mulai dari K=1 hingga K=n untuk memperoleh titik "siku" (*elbow point*) pada grafik, yang menunjukkan nilai K terbaik. Titik tersebut merepresentasikan jumlah cluster optimal berdasarkan penurunan nilai inertia yang signifikan sebelum grafik mulai melandai



Gambar 1. Grafik Hasil Perhitungan Elbow Method

Berdasarkan grafik pada Gambar 1 keterkaitan antara nilai distortion score (inertia) dan jumlah cluster (k) pada rentang nilai 1 hingga 9. Penurunan nilai distortion yang signifikan terlihat dari k = 1 hingga k = 3, kemudian grafik menunjukkan pola yang mulai mendatar setelah k = 3. Titik siku (elbow point) pada grafik ditandai dengan garis putus-putus vertikal pada k = 3, dengan nilai distortion sebesar 16.083,114. Titik ini menjadi indikasi bahwa pemilihan tiga cluster merupakan jumlah yang paling optimal, karena penambahan jumlah cluster setelah titik ini tidak lagi menghasilkan penurunan inertia yang berarti. Oleh karena itu, jumlah cluster yang ideal untuk pengelompokan data UMKM di Kota Bekasi ditentukan sebanyak tiga cluster. Hasil ini digunakan sebagai dasar dalam proses pemodelan menggunakan algoritma K-Means Clustering, dengan tujuan membentuk kelompok UMKM berdasarkan kesamaan karakteristik dalam hal modal usaha, omzet, dan lokasi kecamatan.

Perhitungan Algoritma *K-Means Clustering*

Setelah jumlah cluster optimal ditentukan sebanyak tiga, tahap awal dalam proses pemodelan menggunakan algoritma K-Means adalah penentuan centroid awal. Penetapan ini dilakukan dengan memilih secara acak 3 data dari dataset UMKM yang telah melalui proses normalisasi pada variabel modal usaha, omzet, dan kode kecamatan. Pemilihan data acak ini bertujuan untuk menginisialisasi posisi awal centroid sebagai titik pusat dari masing-masing cluster.

Tabel 1. Menentukan Centroid awal

No	Nama NIK	*Centroid*
1	Nuryono Basyori 3275062002700020	0
2	Uswatun Hasanah 3212314502840001	1
3	Fitria Ningsih 3275044707830029	2

Centroid yang dipilih secara acak dari dataset UMKM digunakan sebagai titik pusat awal (initial centroid) pada iterasi pertama dalam proses perhitungan algoritma K-Means secara manual melalui Microsoft Excel. Ketiga centroid awal (C_0, C_1, dan C_2) merepresentasikan pusat dari masing-masing cluster yang akan dibentuk. Pemilihan centroid ini didasarkan pada tiga variabel utama yang telah dinormalisasi sebelumnya, yaitu modal usaha, omzet, dan kode kecamatan.

Tahap selanjutnya adalah menghitung jarak setiap data terhadap ketiga centroid tersebut menggunakan rumus Euclidean Distance, dimulai dari data pertama hingga seluruh data sampel sebanyak 138 entri. Perhitungan ini dilakukan secara berulang agar setiap data dapat diklasifikasikan ke dalam cluster dengan jarak terdekat terhadap salah satu centroid.

Setelah seluruh data teralokasi ke dalam cluster masing-masing, dilakukan pembaruan posisi centroid dengan menghitung rata-rata nilai variabel dari setiap anggota dalam tiap cluster. Proses ini diulang hingga diperoleh kondisi konvergen, yaitu ketika tidak terdapat lagi perubahan dalam keanggotaan cluster pada iterasi selanjutnya. Hasil perhitungan manual ini menjadi acuan dalam mengevaluasi dan membandingkan hasil pemodelan K-Means menggunakan keseluruhan data UMKM Kota Bekasi secara komputasional.

Data ke-1 dengan "Nama NIK Sampurna Ginting 3275035011680052" dengan *centroid* 0 (C0).

$$d_0(p,q) = \sqrt{(0{,}227-(-0{,}227))^2+(-0{,}196-(-0{,}196))^2+(0-(-0{,}5))^2} = 0.456$$

Jarak data ke-1 dengan "Nama NIK Sampurna Ginting 3275035011680052" dengan centroid 1 ($C1$).

$$d_1(p,q) = \sqrt{(0{,}227-4{,}093)^2+(-0{,}196-24{,}294)^2+(0-(-0{,}166))^2} = 614{,}776$$

Jarak data ke-1 dengan "Nama NIK Sampurna Ginting 3275035011680052" dengan centroid 2 ($C2$).

$$d_2(p,q) = \sqrt{(0{,}227-14{,}328)^2+(-0{,}196-0{,}981)^2+(0-0{,}666)^2} = 200{,}680$$

Centroid yang telah dipilih secara acak dari dataset UMKM Kota Bekasi digunakan sebagai titik pusat awal (centroid) pada iterasi pertama dalam proses perhitungan manual algoritma K-Means menggunakan Microsoft Excel. Masing-masing centroid awal (C_0, C_1, dan C_2) merepresentasikan pusat dari ketiga cluster yang akan dibentuk. Pemilihan centroid dilakukan berdasarkan nilai dari tiga variabel utama yang telah dinormalisasi sebelumnya, yaitu modal usaha, omzet, dan kode kecamatan. Tahap berikutnya adalah melakukan perhitungan jarak antara setiap data terhadap ketiga centroid awal dengan menggunakan rumus Euclidean Distance, yang dimulai dari data ke-1 hingga data ke-138. Sebagai contoh, pada data ke-1 (UMKM dengan NIK 3275035011680052), diperoleh jarak terhadap: C_0 = 0,456, C_1 = 614,776, C_2 = 200,604, Berdasarkan nilai jarak tersebut, data ke-1 memiliki jarak terkecil terhadap C_0, sehingga pada iterasi pertama data tersebut dimasukkan ke dalam Cluster 0. Prosedur ini diulang untuk seluruh data dalam sampel. Setelah semua jarak dihitung dan data dialokasikan ke cluster berdasarkan jarak terdekat ke centroid, langkah selanjutnya adalah menghitung ulang nilai centroid baru. Nilai centroid baru ditentukan dengan mengambil rata-rata dari seluruh anggota cluster sementara yang terbentuk pada iterasi pertama. Centroid baru ini kemudian digunakan dalam iterasi selanjutnya, dan proses yang sama diulang kembali.

Iterasi akan terus berlangsung hingga posisi centroid tidak lagi berubah secara signifikan dari satu iterasi ke iterasi berikutnya, yang menandakan bahwa algoritma telah mencapai kondisi konvergen. Dalam penelitian ini, konvergensi tercapai pada iterasi ke-3, yang sekaligus menjadi dasar dalam mengevaluasi hasil pengelompokan secara komputasi terhadap keseluruhan data UMKM sebanyak 7.923 entri.

$$c^0 = \left(\frac{x_{1,1}+x_{1,2}+\ldots+x_{i,n}}{n_k}\right), \left(\frac{x_{2,1}+x_{2,2}+\ldots+x_{2,n}}{n_k}\right), \left(\frac{x_{3,1}+x_{3,2}+\ldots+x_{3,n}}{n_k}\right)$$

- $c^0 = \left(\frac{0{,}227+(-0{,}409)+\cdots+0{,}227}{122}\right), \left(\frac{(-0{,}196)+(-0{,}137)\ldots+(-0{,}122)}{122}\right), \left(\frac{0+0+\cdots+0{,}16}{122}\right.$
  = **{(0,437), (0,592), (0,244)}**

$$c^1 = \left(\frac{x_{1,1}+x_{1,2}+\ldots+x_{i,n}}{n_k}\right), \left(\frac{x_{2,1}+x_{2,2}+\ldots+x_{2,n}}{n_k}\right), \left(\frac{x_{3,1}+x_{3,2}+\ldots+x_{3,n}}{n_k}\right)$$

- $c^1 = \left(\frac{4{,}093+8{,}642+\cdots+22{,}289}{6}\right), \left(\frac{24{,}294+24{,}294+\cdots+26{,}748}{6}\right), \left(\frac{(-0{,}166)+(-0{,}666)+\cdots+(-0333}{6}\right)$
  = **{(8,870), (22,249), (0,027)}**

$$c^2 = \left(\frac{x_{1,1}+x_{1,2}+\ldots+x_{i,n}}{n_k}\right), \left(\frac{x_{2,1}+x_{2,2}+\ldots+x_{2,n}}{n_k}\right), \left(\frac{x_{3,1}+x_{3,2}+\ldots+x_{3,n}}{n_k}\right)$$

- $c^2 = \left(\frac{14{,}328+10{,}917+\cdots+13{,}191}{10}\right), \left(\frac{0{,}981+12{,}024+\cdots+1{,}521}{10}\right), \left(\frac{0{,}666+(-0{,}333)+\cdots}{10}\right.$
  = **{(15,579), (3,051), (0,4)}**

Proses perhitungan dalam algoritma K-Means Clustering dilakukan secara iteratif hingga mencapai titik konvergensi, yaitu kondisi di mana posisi centroid tidak mengalami perubahan berarti antar iterasi. Keadaan ini menandakan bahwa struktur cluster telah stabil, dan tidak terdapat perpindahan anggota data antar cluster.

Dalam penelitian ini, hasil perhitungan manual terhadap 138 sampel data UMKM

Submitted: **19/07/2025**; Revised: **19/08/2025**; Accepted: **19/09/2025**; Published: **30/04/2026**

menunjukkan bahwa konvergensi tercapai pada iterasi ketiga. Artinya, pada iterasi ke-3, seluruh centroid telah mencapai posisi tetap dan keanggotaan data dalam masing-masing cluster tidak lagi berubah.

Output akhir dari proses clusterisasi pada iterasi ke-3 disajikan dalam Tabel 2, yang menampilkan pengelompokan UMKM berdasarkan kemiripan nilai pada variabel modal, omzet, dan kode kecamatan. Hasil ini digunakan sebagai dasar untuk mengevaluasi kualitas cluster yang terbentuk dan sebagai pembanding terhadap hasil komputasi yang dilakukan menggunakan keseluruhan data UMKM Kota Bekasi.

	Nama NIK	Modal	Omzet	kecamatan_encoded	cluster	kategori
70	Aby Dahana 123332111	10000000	5000000	0	0	UMKM Rendah
73	Edi Prayitno 3275071301850005	25000000	4000000	0	0	UMKM Rendah
75	Iis Yulianah 3275074306920017	50000000	40000000	0	0	UMKM Rendah
78	SITI AMINAH 3275076008860010	15000000	4000000	0	0	UMKM Rendah
81	Indah 2375056002880011	100000000	30000000	0	0	UMKM Rendah
...	...	...	...	...	...	...
7916	Anggria Jaya Putri 3172026108840005	120000000	3000000	7	0	UMKM Rendah
7918	Wahyu Ajiningrat 3216061309800016	300000000	20000000	7	2	UMKM Potensial
7919	Ida Fitriya 3216066810730008	150000000	25000000	7	0	UMKM Rendah
7920	Komalasari 3275024808740036	5000000	2000000	7	0	UMKM Rendah
7921	Sahrul Fauzi 3216090405970008	15000000	5000000	7	0	UMKM Rendah

1378 rows × 6 columns

Gambar 2. Hasil Clusterisasi UMKM

Hasil pengelompokan terhadap 1.378 data UMKM di Kota Bekasi menggunakan algoritma K-Means Clustering menghasilkan tiga cluster utama berdasarkan variabel modal, omzet, dan lokasi kecamatan. Masing-masing cluster diberi label sesuai karakteristik dominan yang terbentuk:

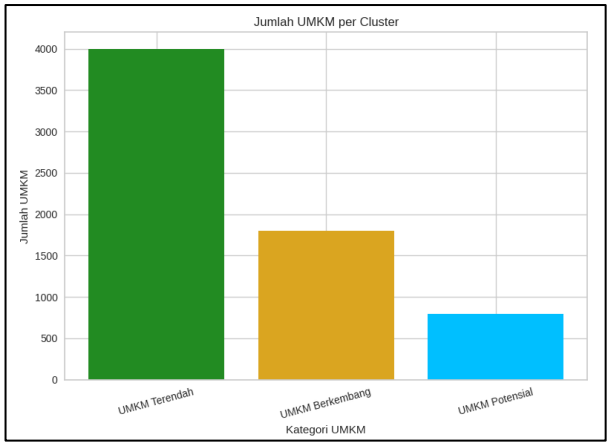
- Cluster 0 (UMKM Rendah): berisi UMKM dengan modal dan omzet rendah. Contohnya UMKM milik *Aby Dahana* dengan modal Rp10 juta dan omzet Rp5 juta.
- Cluster 1 (UMKM Berkembang): terdiri dari UMKM dengan modal menengah dan omzet stabil, berpotensi naik kelas jika didukung pembinaan.
- Cluster 2 (UMKM Potensial): mencakup UMKM dengan modal dan omzet tinggi, seperti *Wahyu Ajiningrat* yang memiliki modal Rp30 juta dan omzet Rp20 juta.

Klasifikasi ini digunakan untuk mendukung analisis segmentasi dan menjadi dasar dalam menyusun strategi pengembangan UMKM yang lebih terarah dan berbasis data.

Model Hasil Clusterisasi

Hasil pemodelan yang telah dilakukan sebelumnya kemudian dianalisis lebih lanjut untuk menentukan karakteristik dari masing-masing cluster, yang selanjutnya dikategorikan sebagai UMKM Potensial, UMKM Berkembang, dan UMKM Rendah. Penentuan kategori ini didasarkan pada nilai rata-rata modal, omzet, dan lokasi kecamatan dari setiap cluster yang terbentuk.

Analisis ini bertujuan untuk memberikan interpretasi yang lebih aplikatif terhadap hasil pengelompokan, sehingga dapat digunakan sebagai dasar dalam perumusan strategi pengembangan UMKM secara lebih tepat sasaran. Dengan memahami kondisi ekonomi dari tiap cluster, pembinaan dan intervensi kebijakan dapat disesuaikan sesuai kebutuhan kelompok.



Gambar 3 hasil UMKM per Cluster

Gambar 3 menyajikan hasil akhir dari pengelompokan data UMKM menggunakan algoritma K-Means Clustering, lengkap dengan atribut cluster dan kategori yang telah diberikan berdasarkan kemiripan karakteristik antar pelaku usaha.

Langkah akhir dalam proses pemodelan ini adalah menyajikan hasil pengelompokan dalam bentuk diagram batang untuk menampilkan jumlah pelaku UMKM pada masing-masing cluster hasil segmentasi algoritma K-Means Clustering.Visualisasi ini menunjukkan distribusi UMKM berdasarkan kemiripan karakteristik modal, omzet, dan lokasi, yang telah dikategorikan menjadi UMKM Rendah, UMKM Berkembang, dan UMKM Potensial.

Sebagaimana ditunjukkan pada Gambar 3, sebanyak 754 UMKM masuk dalam kategori Rendah, 466 UMKM dalam kategori Berkembang, dan 158 UMKM tergolong Potensial.Informasi ini memberikan gambaran menyeluruh tentang kondisi UMKM di Kota Bekasi dan dapat dimanfaatkan sebagai dasar

dalam perencanaan pembinaan dan pengembangan UMKM secara terarah.

Evaluasi Model

Proses terakhir dari penelitian adalah tahap evaluasi, yang memiliki tujuan untuk mengukur kualitas model clustering. *Davies-Bouldin Index* (DBI), metrik yang mengukur tingkat kohesi (kedekatan data dalam satu cluster) dan separasi (jarak antar cluster), digunakan untuk menilai evaluasi. Nilai DBI yang lebih rendah menunjukkan bahwa kualitas pemisahan cluster yang terbentuk lebih baik. Tahapan proses perhitungan DBI yaitu:

a. Menghitung *Sum of Square Within Cluster* (SSW) Langkah pertama dalam evaluasi adalah menghitung Sum of Square Within Cluster (SSW), yang digunakan untuk mengukur tingkat kemiripan atau kedekatan antar data dalam satu cluster terhadap centroid-nya. Nilai ini mencerminkan tingkat kohesi dari cluster tersebut. Perhitungan dilakukan menggunakan persamaan (6).

$$SSW = \frac{1}{m}\sum_{i=1}^{k} d\,(x_i y_i) \quad (6)$$

Langkah selanjutnya adalah menentukan nilai SSW pada masing-masing cluster dengan cara menghitung rata-rata jarak dari tiap anggota cluster yang ada.

SSW *cluster 0*

$SSW_0 = \frac{0{,}852+1{,}144+\cdots+0{,}624+0{,}979}{122} = 1{,}542$

SSW *cluster 1*

$SSW_1 = \frac{5{,}199+2{,}171+\cdots+4{,}955+14{,}157}{6} = 7{,}205$

SSW *cluster 2*

$SSW_2 = \frac{2{,}433+10{,}138+\cdots+3{,}350+2{,}985}{10} = 4{,}770$

b. Menghitung *Sum of Square Between Cluster (SSB)* untuk mengetahui seberapa jauh *centroid* antar *cluster* tersebar atau dapat disebut juga sebagai nilai separasi, dengan menggunakan persamaan (7).

$$SSB_{ij} = d(y_i\, y_j) \quad (7)$$

SSB cluster 0

$SSB_0 = 0$

SSB *cluster 1dengan cluster 0*

$SSB_{1,0} = \sqrt{(0{,}438-8{,}87)^2+(0{,}593-22{,}25)^2+(0{,}245-0{,}028)^2}$

$SSB_{1,0} = 23{,}241$

SSB *cluster 1 dengan cluster 2*

$SSB_{1,2} = \sqrt{(15{,}58-8{,}87)^2+(3{,}051-22{,}25)^2+(0{,}4-0{,}028)^2}$

$SSB_{1,2} = 20{,}340$

SSB *cluster 2 dengan cluster 0*

$SSB_{2,0} = \sqrt{(0{,}438-15{,}58)^2+(0{,}593-3{,}051)^2+(0{,}245-0{,}4)^2}$

$SSB_{2,0} = 15{,}342$

Tabel 1 merupakan data hasil perhitungan nilai SSB untuk ketiga cluster.

Tabel 1. Hasil Perhitungan SSB

SSB	C0	C1	C2
C0	0	23,241	15,341
C1	23,241	0	20,340
C2	15,341	20,340	0

c. Menghitung *Sum of Square Between Cluster (SSB)* untuk mengetahui seberapa jauh *centroid* antar *cluster* tersebar atau dapat disebut juga sebagai nilai separasi, dengan menggunakan persamaan (8).

$$SSB_{ij} = d(y_i\, y_j) \quad (8)$$

SSB cluster 0

$SSB_0 = 0$

SSB *cluster 1dengan cluster 0*

$SSB_{1,0} = \sqrt{(0{,}438-8{,}87)^2+(0{,}593-22{,}25)^2+(0{,}245-0{,}028)^2}$

$SSB_{1,0} = 23{,}241$

SSB *cluster 1 dengan cluster 2*

$SSB_{1,2} = \sqrt{(15{,}58-8{,}87)^2+(3{,}051-22{,}25)^2+(0{,}4-0{,}028)^2}$

$SSB_{1,2} = 20{,}340$

SSB *cluster 2 dengan cluster 0*

$SSB_{2,0} = \sqrt{(0{,}438-15{,}58)^2+(0{,}593-3{,}051)^2+(0{,}245-0{,}4)^2}$

$SSB_{2,0} = 15{,}342$

Tabel 2 merupakan data hasil perhitungan nilai SSB untuk ketiga cluster.

Tabel 2. Hasil Perhitungan SSB

SSB	C0	C1	C2
C0	0	23,24148	15,34128
C1	23,24148	0	20,34009
C2	15,34128	20,34009	0

Submitted: **19/07/2025**; Revised: **19/08/2025**; Accepted: **19/09/2025**; Published: **30/04/2026**

KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, jumlah cluster optimal ditentukan menggunakan metode Elbow, yang menunjukkan titik siku pada K = 3. Proses pengelompokan dilakukan melalui dua pendekatan, yaitu perhitungan manual terhadap 138 data sampel dan komputasi otomatis menggunakan Python pada seluruh dataset sebanyak 7.923 data UMKM. Kedua pendekatan menghasilkan hasil yang konsisten, di mana kondisi konvergen tercapai pada iterasi ke-3 untuk data manual dan iterasi ke-10 untuk keseluruhan data.Distribusi anggota cluster dari hasil komputasi adalah sebagai berikut:Cluster 0: 754 UMKM, Cluster 1: 466 UMKM, Cluster 2: 158 UMKM.Pelabelan cluster dilakukan berdasarkan karakteristik dominan masing-masing kelompok, yakni:Cluster 0: UMKM Rendah, Cluster 1: UMKM Berkembang, Cluster 2: UMKM Potensial Evaluasi kualitas cluster dilakukan menggunakan metode Davies-Bouldin Index (DBI) dan menghasilkan skor sebesar 0,5940, yang berada dalam kategori cukup baik karena mendekati nilai nol. Hal ini menunjukkan bahwa hasil segmentasi memiliki pemisahan cluster yang layak dan dapat digunakan sebagai dasar dalam pengambilan keputusan, khususnya dalam penyusunan strategi pengembangan, pembinaan, dan pengalokasian bantuan bagi pelaku UMKM di Kota Bekasi.

DAFTAR PUSTAKA

[1] W. Sudrajat, I. Cholid, dan J. Petrus, "Penerapan Algoritma K-Means Clustering untuk Pengelompokan UMKM Menggunakan Rapidminer," J. JUPITER, vol. 14, no. 1, hal. 27–36, 2022.

[2] J. H. HANGHANG TONG, JIAN PEI, Data Mining Conceptual and Techniques Fourth Edition. 2023.

[3] N. Wahyunisari dan R. Kurniawan, "PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN DATA PENJUALAN PAKAIAN (STUDI KASUS : UMKM LIMA MEDIA KUNINGAN)," vol. 8, no. 2, 2024.

[4] N. Ameliana, N. Suarna, dan W. Prihartono, "Analisis Data Mining Pengelompokkan Umkm Menggunakan Algoritma K-Means Clustering Di Provinsi Jawa Barat," JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 3, hal. 3261–3268, 2024, doi: 10.36040/jati.v8i3.9655.

[5] Sari, R., Helmy, N., & Alexander, A. D. (2024). Determining Sales Patterns Using the Apriori Algorithm: A Case Study of Unlocked Cafe's Website Applications. PIKSEL: Penelitian Ilmu Komputer Sistem Embedded and Logic, 12(1), 189-200.

[6] I. Sukmadewanti, R. Arifudin, dan E. Sugiharti, "Use of K-Means Clustering and Analytical Methods Hierarchy Process in Determining the Type of MSME Financing in Semarang City," Sci. J. Informatics, vol. 5, no. 2, hal. 2407–7658, 2018, [Daring]. Tersedia pada: http://journal.unnes.ac.id/nju/index.php/sji

[7] N. T. H. E. Challenges, "CLASSIFICATION IN IMBALANCED Table Of Content," 2023.

[8] M. A. Nasir Muhammad, DATA MINING. 2020.

[9] M. S. Dr. Suyanto, S.T., Data Mining untuk Klasifikasi dan Clusterisasi data. 2017.

[10] Y. Sopyan, A. D. Lesmana, dan C. Juliane, "Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan," Build. Informatics, Technol. Sci., vol. 4, no. 3, hal. 1464–1470, 2022, doi: 10.47065/bits.v4i3.2697.

[10] M. A. S. F. S. T., "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin," J. Evolusi, vol. 10, no. 1, hal. 1–9, 2021, [Daring]. Tersedia pada: https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/download/10428/4839