

Optimalisasi Pemetaan UMKM Di Wilayah Dki Jakarta Melalui K-Means Clustering dan PCA

Muh. Junaidi¹, Pita Penengah², Danda Bagaskoro^{3*}, Muhammad Ilham Afandi⁴

Program Studi Sistem Informasi, Universitas Nusantara PGRI Kediri^{1,2,3,4}

muhjunaidi183@gmail.com¹, pitapenengah@gmail.com², bagaskorodanda@gmail.com³,
ne9hams12@gmail.com⁴

**Corresponding author: Danda Bagaskoro*

Usaha mikro, kecil, dan menengah (UMKM) merupakan pilar penting dalam perekonomian Indonesia, namun distribusi dan karakteristik UMKM di berbagai wilayah DKI Jakarta masih belum terpetakan secara optimal. Penelitian ini bertujuan untuk melakukan segmentasi wilayah berdasarkan jumlah dan jenis UMKM dengan menerapkan algoritma K-Means Clustering menggunakan Python. Proses analisis dimulai dari integrasi data UMKM, praproses (penghapusan kolom tidak relevan, penanganan missing value), transformasi data menjadi pivot table, hingga stan'darisasi fitur numerik. Penentuan jumlah kluster optimal dilakukan dengan metode Elbow, sementara kualitas klusterisasi dievaluasi menggunakan Silhouette Score dan Davies-Bouldin Index. Hasil klusterisasi menunjukkan adanya tiga segmen utama wilayah: rendah, sedang, dan tinggi berdasarkan jumlah UMKM. Visualisasi menggunakan diagram lingkaran dan Principal Component Analysis (PCA) memperjelas distribusi dan pemisahan antar kluster. Nilai Silhouette Score sebesar 0,867 dan Davies-Bouldin Score sebesar 0,366 menandakan kualitas klusterisasi yang sangat baik. Temuan ini diharapkan dapat menjadi dasar bagi pemerintah dan pemangku kepentingan dalam merumuskan kebijakan pengembangan UMKM yang lebih tepat sasaran dan berbasis data.

Kata Kunci : *UMKM, K-Means Clustering, DKI Jakarta*

A. PENDAHULUAN

Ini Usaha Mikro, Kecil, dan Menengah (UMKM) memegang peranan penting dalam perekonomian Indonesia, baik sebagai penyerap tenaga kerja maupun sebagai penggerak ekonomi daerah. Namun, dinamika pertumbuhan UMKM di berbagai wilayah sering kali tidak merata, dipengaruhi oleh faktor geografis, jenis usaha, serta akses terhadap sumber daya[1]. Untuk memahami pola persebaran dan karakteristik UMKM di suatu wilayah, diperlukan analisis data yang komprehensif dan berbasis teknologi. Salah satu pendekatan yang semakin banyak digunakan adalah analisis kluster berbasis algoritma K-Means dengan dukungan bahasa pemrograman Python, yang menawarkan kemudahan dalam pengolahan data, visualisasi, serta interpretasi hasil secara sistematis. Urgensi penelitian ini terletak pada kebutuhan pemerintah daerah dan pemangku kepentingan untuk memperoleh gambaran segmentasi wilayah berdasarkan potensi UMKM, sehingga intervensi kebijakan dapat dilakukan secara lebih terarah dan efektif[2].

Permasalahan utama yang dihadapi adalah belum optimalnya pemetaan dan pengelompokan wilayah berdasarkan karakteristik dan jumlah UMKM per jenis usaha. Data yang tersedia sering kali belum diolah secara maksimal, sehingga potensi wilayah tertentu tidak teridentifikasi dengan baik. Selain itu, masih terdapat kendala dalam penanganan data yang tidak lengkap (missing value) dan kebutuhan akan validasi hasil segmentasi yang akurat[3].

Rumusan masalah yang diangkat dalam penelitian ini adalah bagaimana proses analisis kluster menggunakan algoritma K-Means berbasis Python dapat mengelompokkan wilayah berdasarkan jenis jumlah UMKM, serta bagaimana validasi dan visualisasi hasil klusterisasi dapat memberikan gambaran yang jelas bagi pengambilan keputusan[4]

Penelitian ini bertujuan untuk mengimplementasikan analisis kluster K-Means pada data UMKM menggunakan Python, mulai dari tahap praproses data, transformasi, penanganan missing value, hingga evaluasi dan visualisasi hasil klusterisasi. Dengan demikian, diharapkan

dapat diperoleh segmentasi wilayah yang lebih informatif dan mendukung perumusan kebijakan pengembangan UMKM yang berbasis data[5].

Hasil riset ini diharapkan memberikan kontribusi nyata berupa metode analisis data UMKM yang terstruktur dan dapat direplikasi, sekaligus menyediakan visualisasi segmentasi wilayah yang mudah dipahami. Penelitian ini juga dapat menjadi referensi bagi akademisi, pemerintah, maupun pelaku usaha dalam melakukan pemetaan potensi wilayah dan perencanaan strategi pengembangan UMKM yang lebih efektif[6].

Kerangka berpikir riset dimulai dengan tahap import dan integrasi data UMKM dari Google Drive ke Google Colab, dilanjutkan dengan pra-proses data seperti penghapusan kolom tidak relevan dan penanganan missing value menggunakan median. Data kemudian ditransformasikan ke dalam bentuk pivot table untuk agregasi jumlah usaha per jenis di setiap wilayah. Selanjutnya, dilakukan standarisasi data dan penentuan jumlah kluster optimal menggunakan metode elbow, diikuti dengan penerapan algoritma K-Means untuk segmentasi wilayah[7]. Hasil klusterisasi divalidasi menggunakan silhouette score dan davies-bouldin score, serta divisualisasikan dengan teknik PCA untuk memudahkan interpretasi. Seluruh proses ini diharapkan menghasilkan segmentasi wilayah UMKM yang valid, informatif, dan aplikatif dalam mendukung pengambilan keputusan berbasis data[8].

Tinjauan Pustaka

B. LANDASAN TEORI

1. Usaha Mikro, Kecil, dan Menengah (UMKM)

UMKM merupakan sektor usaha yang berperan penting dalam mendorong pertumbuhan ekonomi, menciptakan lapangan kerja, dan mengurangi tingkat kemiskinan di Indonesia. Perkembangan UMKM yang merata dinilai dapat memperkuat struktur perekonomian nasional secara berkelanjutan. UMKM dinilai produktif karena sifatnya yang fleksibel dan menjangkau berbagai lapisan masyarakat. Oleh karena itu, perhatian terhadap pola penyebaran dan jumlah UMKM di tiap wilayah sangat penting untuk perencanaan pembangunan ekonomi daerah[4].

2. K-Means Clustering

K-Means Clustering adalah salah satu metode unsupervised learning yang digunakan untuk mengelompokkan data ke dalam beberapa kluster berdasarkan kemiripan karakteristik. Proses ini dilakukan dengan menentukan pusat kluster (centroid), lalu mengelompokkan data berdasarkan jarak terdekat ke centroid tersebut. Proses ini berlangsung secara iteratif hingga pusat kluster stabil atau tidak berubah lagi.

Metode ini banyak digunakan dalam analisis segmentasi karena kesederhanaannya, kemampuannya menangani data numerik, serta efisiensinya dalam mengelola dataset berskala besar.

3. Proses Knowledge Discovery in Database (KDD)

Implementasi K-Means dalam penelitian ini mengikuti tahapan metodologi KDD, yang meliputi:

1. **Data Selection:** Pemilihan data yang relevan, misalnya jumlah UMKM per wilayah.
2. **Preprocessing:** Pembersihan data dari noise, duplikasi, dan atribut yang tidak diperlukan untuk memastikan data siap dianalisis.
3. **Transformation:** Mengubah format data agar sesuai dengan kebutuhan model, seperti membuat pivot table dan agregasi jumlah usaha.
4. **Data Mining:** Penerapan algoritma K-Means untuk membentuk kluster.
5. **Evaluation:** Menilai kualitas kluster menggunakan metrik seperti *Silhouette Coefficient* dan *Davies-Bouldin Index*. Nilai *Silhouette Coefficient* yang tinggi (mendekati 1) dan nilai DBI yang rendah menunjukkan kluster yang baik.

4. Evaluasi Kualitas Kluster

Evaluasi hasil klusterisasi sangat penting untuk memastikan bahwa segmentasi yang terbentuk memiliki struktur yang baik:

1. **Silhouette Coefficient:** Mengukur konsistensi internal dalam kluster. Nilai antara 0,71 hingga 1 menunjukkan struktur yang kuat.

2. **Davies-Bouldin Index (DBI):** Mengukur tingkat pemisahan antar kluster. Nilai semakin mendekati nol menandakan pemisahan kluster yang semakin optimal.

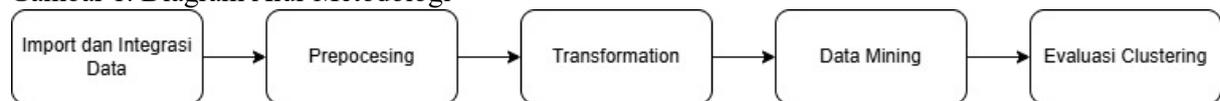
5. Visualisasi Kluster

Visualisasi seperti *Elbow Method* untuk menentukan jumlah kluster optimal, serta scatter plot dari hasil *Principal Component Analysis (PCA)*, digunakan untuk memudahkan interpretasi dan validasi hasil segmentasi secara visual. Pie chart atau diagram lingkaran juga bermanfaat untuk menampilkan proporsi wilayah dalam tiap kluster

C. METODE PENELITIAN

K-Means Clustering adalah salah satu algoritma pengelompokan data non-hirarki yang bertujuan untuk membagi data ke dalam sejumlah kelompok (cluster) berdasarkan kemiripan karakteristik antar data[9]. Algoritma ini bekerja dengan meminimalkan jarak antar data dalam satu cluster dan memaksimalkan jarak antar cluster yang berbeda. Setiap cluster direpresentasikan oleh sebuah titik pusat (centroid), yang merupakan rata-rata dari seluruh data anggota cluster tersebut. Setiap tahapan digambarkan dalam sebuah alur metodologi penelitian seperti berikut[10]:

Gambar 1. Diagram Alur Metodologi



1. Import dan Integrasi Data

Tahap ini merupakan langkah awal dalam proses segmentasi wilayah UMKM di DKI Jakarta menggunakan algoritma K-Means. Pada tahap ini, data UMKM diimpor dari Google Drive ke Google Colab, sehingga memungkinkan proses analisis dilakukan secara online dan efisien dengan dukungan Python.

2. Preprocessing

Preprocessing adalah proses pembersihan dan penyederhanaan teks. Dalam proses pengambilan data, data yang didapatkan pasti berupa unstructured data, yang di mana isi dalam data tersebut masih terdapat noise. Maka dari itu, tujuan dari preprocessing ini untuk menghilangkan noise yang ada. Selain itu, proses pembersihan data lainnya diantaranya menghilangkan duplikasi data, mengoreksi kesalahan pada data, dan meninjau data yang inkonsisten[11].

3. Transformation

Tahap transformasi adalah proses pengembangan dataset sesuai dengan kebutuhan padatah modeling atau tahap data mining, karena sebagian metode dalam data mining membutuhkan ketentuan khusus sebelum diolah dalam data mining.

4. Data Mining

Tahap data mining merupakan proses di mana pengolahan data, proses mencari pola atau informasi berdasarkan metode ataupun teknik yang akan digunakan. Pada pemilihan metode yang sesuai digunakan sangat bergantung kepada tujuan dan proses KDD keseluruhan[12]

a) Diagram Elbow

Metode Elbow digunakan untuk menentukan jumlah cluster optimal dengan cara memplot nilai Within Cluster Sum of Squares (WCSS) terhadap jumlah cluster.

Titik "elbow" pada grafik menunjukkan jumlah cluster yang memberikan penurunan signifikan pada WCSS dan merupakan pilihan terbaik untuk k [13]

b) Diagram Lingkaran (pie cart)

Diagram lingkaran digunakan untuk memvisualisasikan proporsi atau distribusi data dalam setiap kluster hasil segmentasi wilayah UMKM. Dalam konteks penelitian ini, diagram lingkaran menampilkan persentase wilayah yang masuk ke masing-masing kategori kluster, yaitu "rendah", "sedang", dan "tinggi" berdasarkan jumlah UMKM di setiap wilayah.

c) Diagram PCA

Diagram PCA adalah scatter plot yang menampilkan hasil reduksi dimensi data menggunakan teknik Principal Component Analysis. PCA memproyeksikan data berdimensi banyak ke dalam dua dimensi utama, sehingga memudahkan visualisasi hasil klusterisasi[14].

5. Evaluasi Clustering

Evaluasi hasil clustering dapat dilakukan dengan metrik seperti:

1. **Silhouette Score:** Mengukur seberapa mirip data dalam cluster dengan data di cluster lain.
2. **Davies-Bouldin Index:** Mengukur rasio jarak intra-cluster dan inter-cluster. Visualisasi hasil clustering juga membantu dalam evaluasi, misalnya dengan diagram lingkaran untuk proporsi cluster dan plot PCA untuk melihat distribusi cluster dalam ruang dua dimensi[15]

D. HASIL DAN PEMBAHASAN

1. Import dan Integrasi Data

Proses analisis dimulai dengan mengimpor library utama seperti pandas dan numpy untuk manipulasi data, serta matplotlib dan seaborn untuk visualisasi. Data UMKM diimpor dari Google Drive ke Google Colab, memastikan akses dataset yang efisien dan terintegrasi

Gambar 2. hasil load data

periode_data	triwulan	wilayah	jenis_usaha	jumlah_usaha	nilai_pendapatan	
0	201903	1	Kab. Adm. Kep. Seribu	Kuliner	863	841485000
1	201903	1	Kab. Adm. Kep. Seribu	Lainnya	285	241942000
2	201903	1	Kab. Adm. Kep. Seribu	Fashion	69	25750000
3	201903	1	Kab. Adm. Kep. Seribu	Craft	22	10300000
4	201903	1	Kota Adm. Jakarta Barat	Craft	526	NaN
...
571	202306	2	Kota Adm. Jakarta Selatan	Craft	908	3855601904
572	202306	2	Kota Adm. Jakarta Timur	Kuliner	11917	89106260098
573	202306	2	Kota Adm. Jakarta Timur	Fashion	2149	10192288999
574	202306	2	Kota Adm. Jakarta Timur	Lainnya	1258	7123768040
575	202306	2	Kota Adm. Jakarta Timur	Craft	704	3265999700

576 rows x 6 columns

2. Preprocessing

Tahap praproses dilakukan dengan menghapus kolom yang tidak relevan, seperti "nilai_pendapatan" dan "triwulan", untuk menyederhanakan data. Informasi tipe data dan jumlah nilai non-null diperiksa menggunakan `df.info()`, memastikan struktur data siap untuk analisis lebih lanjut.

Gambar 3. Hasil preprocessing

	periode_data	wilayah	jenis_usaha	jumlah_usaha
0	201903	Kab. Adm. Kep. Seribu	Kuliner	863
1	201903	Kab. Adm. Kep. Seribu	Lainnya	285
2	201903	Kab. Adm. Kep. Seribu	Fashion	69
3	201903	Kab. Adm. Kep. Seribu	Craft	22
4	201903	Kota Adm. Jakarta Barat	Craft	526
...
571	202306	Kota Adm. Jakarta Selatan	Craft	908
572	202306	Kota Adm. Jakarta Timur	Kuliner	11917
573	202306	Kota Adm. Jakarta Timur	Fashion	2149
574	202306	Kota Adm. Jakarta Timur	Lainnya	1258
575	202306	Kota Adm. Jakarta Timur	Craft	704

576 rows x 4 columns

Pemeriksaan nilai kosong dilakukan untuk memastikan tidak ada missing value yang tersisa setelah praproses, sehingga data siap digunakan untuk proses klusterisasi.

Gambar 4. hasil pengecekan missing value

```
periode_data    0
wilayah         0
jenis_usaha     0
jumlah_usaha   0
dtype: int64
```

gambar ini menunjukkan hasil pemeriksaan missing value pada data utama: yakni semua kolom (periode_data, wilayah, jenis_usaha, jumlah_usaha) memiliki nilai 0, artinya tidak ada missing value di seluruh kolom.

Data sudah bersih dan siap untuk proses analisis lebih lanjut. Tidak adanya missing value memastikan hasil klusterisasi lebih akurat dan tidak bias akibat data yang tidak lengkap.

3. Transformasi

Data diubah ke format pivot table, di mana setiap jenis usaha (Kuliner, Fashion, Craft, Lainnya) menjadi kolom, dan indeksnya adalah wilayah. Selanjutnya, dibuat kolom baru 'jumlah_total' sebagai agregasi jumlah usaha per wilayah, dan urutan wilayah diselaraskan dengan data asli.

Gambar 5. hasil transformation

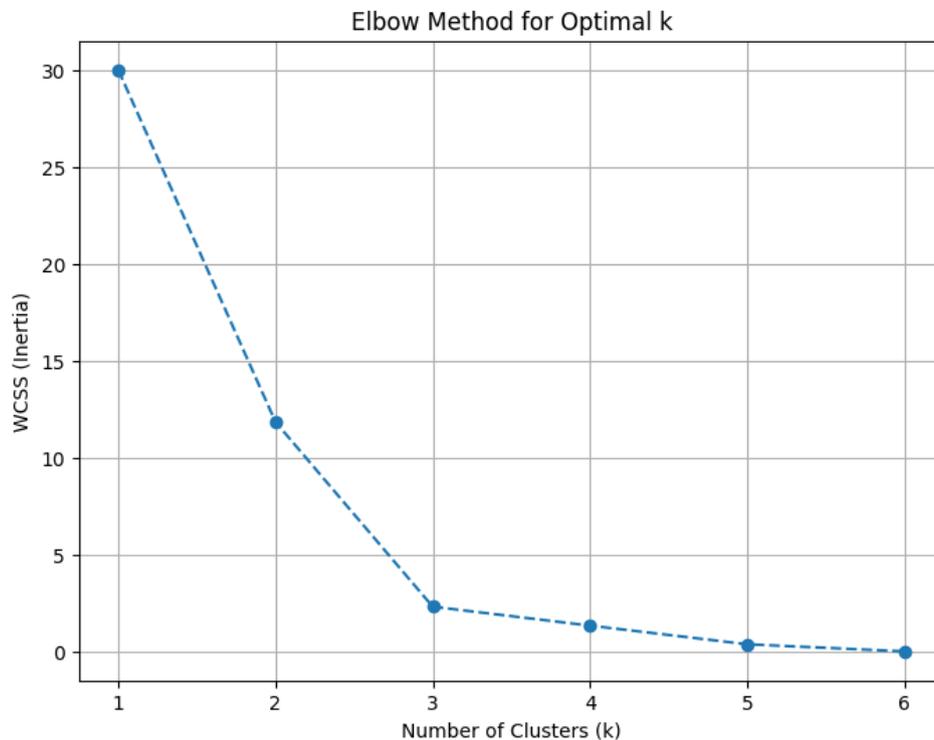
jenis_usaha	wilayah	Craft	Fashion	Kuliner	Lainnya	jumlah_total
0	Kab. Adm. Kep. Seribu	538	1690	20880	6866	29974
1	Kota Adm. Jakarta Barat	12806	30145	201379	13142	257472
2	Kota Adm. Jakarta Pusat	16886	27836	221571	22191	288484
3	Kota Adm. Jakarta Selatan	22035	46594	376388	35590	480607
4	Kota Adm. Jakarta Timur	17247	52075	289191	30525	389038
5	Kota Adm. Jakarta Utara	8508	27918	175628	17922	229976

4. Data Mining

a) Penentuan Jumlah Kluster menggunakan Diagram Elbow

Penentuan jumlah kluster optimal dilakukan menggunakan metode Elbow, dengan menghitung nilai Within-Cluster Sum of Squares (WCSS) untuk berbagai nilai k. Grafik Elbow menunjukkan titik optimal (elbow point) yang menandakan jumlah kluster terbaik untuk segmentasi wilayah UMKM.

Gambar 6. diagram elbow



Sumbu X menunjukkan jumlah kluster, sedangkan sumbu Y menunjukkan nilai WCSS (Within-Cluster Sum of Squares). Pada grafik, terlihat penurunan tajam WCSS dari k=1 ke k=2 dan dari k=2 ke k=3, lalu penurunan mulai melandai setelah k=3. Titik siku (“elbow”) yang terbentuk pada k=3 menandakan bahwa jumlah kluster optimal adalah 3. Menambah kluster lebih dari 3 tidak lagi memberikan pengurangan WCSS yang signifikan, sehingga secara statistik 3 kluster sudah cukup untuk segmentasi data UMKM.

b) Klusterisasi K-Means

Algoritma K-Means diterapkan pada data yang telah distandarisasi. Hasil klusterisasi disimpan dalam kolom 'Kluster', dan setiap kluster diberi label kategori deskriptif seperti "rendah", "sedang", dan "tinggi" sesuai dengan karakteristik jumlah UMKM di setiap wilayah.

Gambar 7. hasil klustering

jenis_usaha	wilayah	Kluster	Kategori
0	Kab. Adm. Kep. Seribu	0	rendah
1	Kota Adm. Jakarta Barat	1	sedang
2	Kota Adm. Jakarta Pusat	0	rendah
3	Kota Adm. Jakarta Selatan	1	sedang
4	Kota Adm. Jakarta Timur	2	tinggi
5	Kota Adm. Jakarta Utara	0	rendah

Gambar ini menampilkan hasil klasterisasi untuk setiap wilayah di Jakarta berdasarkan kategori jumlah UMKM. Setiap wilayah dikelompokkan ke dalam klaster dengan label numerik (0, 1, 2) dan diberikan kategori deskriptif (“rendah”, “sedang”, “tinggi”). hasil:

- Kab. Adm. Kep. Seribu, Jakarta Pusat, dan Jakarta Utara masuk klaster 0 (rendah)
- Jakarta Barat dan Jakarta Selatan masuk klaster 1 (sedang)
- Jakarta Timur masuk klaster 2 (tinggi)

Ini menunjukkan bahwa wilayah-wilayah di Jakarta dapat dipetakan berdasarkan potensi dan jumlah UMKM-nya, sehingga strategi pengembangan dapat lebih spesifik sesuai segmen.

c) Visualisasi Hasil Klasterisasi

Proporsi masing-masing klaster divisualisasikan menggunakan diagram lingkaran, memberikan gambaran distribusi wilayah dalam setiap segmen. Selain itu, Principal Component Analysis (PCA) digunakan untuk mereduksi dimensi data menjadi dua komponen utama. Scatter plot hasil PCA menunjukkan pemisahan klaster yang cukup jelas, meskipun terdapat sedikit tumpang tindih antar segmen, sehingga memudahkan interpretasi **visual**.

Gambar 8. diagram pie cart

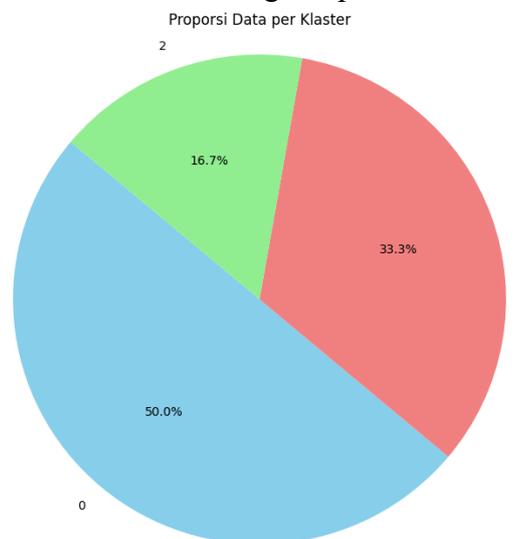
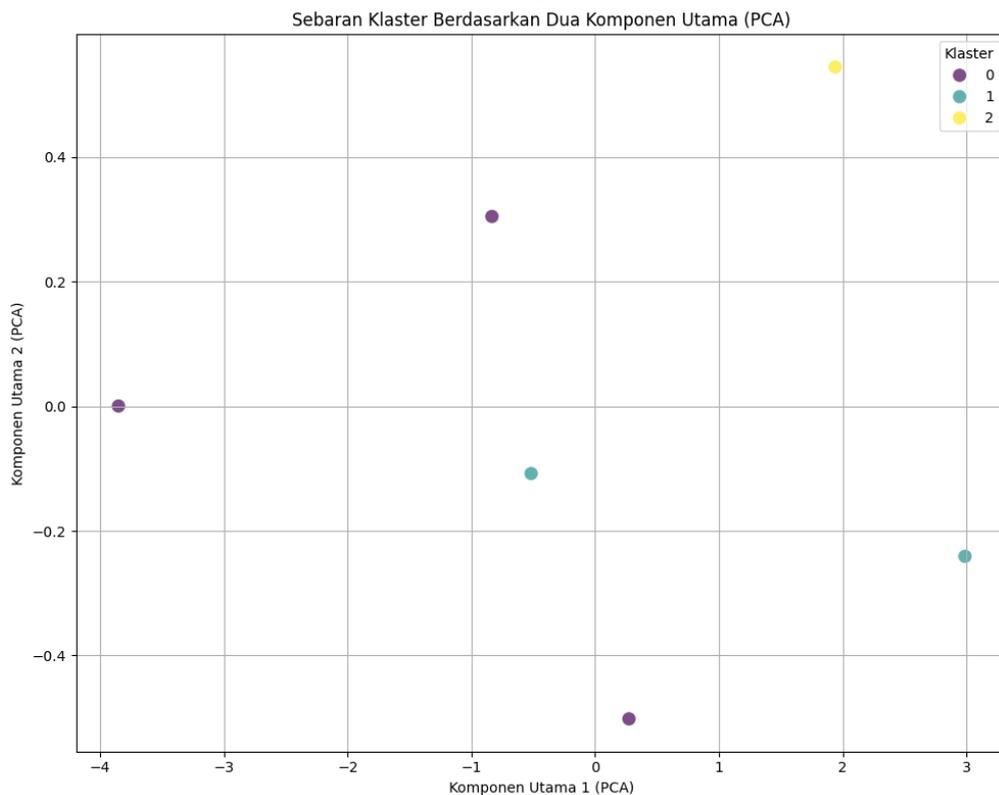


Diagram lingkaran ini menunjukkan distribusi proporsi wilayah pada masing-masing klaster:

1. Klaster 0 (rendah): 50% wilayah
2. Klaster 1 (sedang): 33,3% wilayah
3. Klaster 2 (tinggi): 16,7% wilayah

Artinya, separuh wilayah di Jakarta tergolong dalam kategori UMKM rendah, sepertiga sedang, dan sisanya tinggi. Visualisasi ini memudahkan pengambil keputusan untuk mengetahui sebaran segmentasi wilayah secara cepat dan intuitif.

Gambar 9. diagram PCA



Penjelasan:

Scatter plot hasil PCA memproyeksikan data ke dalam dua dimensi utama, sehingga memudahkan visualisasi pemisahan kluster. Setiap titik mewakili satu wilayah, dan warna menunjukkan kluster hasil K-Means:

1. Kluster 0, 1, dan 2 tampak terpisah dengan baik, meskipun ada sedikit tumpang tindih.
2. Visualisasi ini menegaskan bahwa segmentasi kluster yang dihasilkan cukup baik dan dapat diinterpretasikan secara visual, sehingga pengelompokan wilayah berdasarkan karakteristik UMKM menjadi lebih jelas.

d) Evaluasi Klasterisasi

Evaluasi kualitas klasterisasi dilakukan dengan menghitung Silhouette Score dan Davies-Bouldin Index. Silhouette Score yang tinggi menunjukkan bahwa kluster yang terbentuk memiliki jarak antar kluster yang baik dan kompak, sedangkan Davies-Bouldin Index yang rendah mengindikasikan kluster yang terpisah dengan baik.

Gambar 10. hasil sihoutte score

`np.float64(0.8671317535061608)`

Silhouette Score = 0.867

Nilai ini sangat tinggi (mendekati 1), menandakan bahwa objek dalam satu kluster sangat mirip satu sama lain dan sangat berbeda dengan objek di kluster lain. Artinya, hasil klasterisasi sangat baik dan pemisahan antar kluster sangat jelas.

np.float64(0.36575595939196326)

Davies-Bouldin Score = 0.366

Nilai ini sangat rendah (semakin kecil semakin baik), menunjukkan bahwa kluster yang terbentuk saling terpisah dengan baik dan memiliki dispersi internal yang kecil. Ini menegaskan bahwa kualitas klusterisasi sangat baik.

e) Interpretasi dan Implikasi Hasil

Hasil segmentasi wilayah UMKM yang diperoleh dapat digunakan untuk merumuskan strategi pengembangan wilayah secara lebih spesifik. Wilayah dengan kategori UMKM "tinggi" dapat menjadi prioritas pengembangan atau intervensi, sedangkan wilayah dengan kategori "rendah" dapat difokuskan pada program pemberdayaan. Visualisasi dan evaluasi klusterisasi memberikan dasar yang kuat untuk pengambilan keputusan berbasis data.

E. Kesimpulan dan Saran

Penelitian ini berhasil mengimplementasikan analisis kluster menggunakan algoritma K-Means berbasis Python untuk segmentasi wilayah UMKM di DKI Jakarta. Melalui tahapan praproses data, transformasi, penentuan jumlah kluster optimal dengan metode Elbow, serta evaluasi menggunakan Silhouette Score dan Davies-Bouldin Index, diperoleh segmentasi wilayah menjadi tiga kluster utama: rendah, sedang, dan tinggi berdasarkan jumlah dan jenis UMKM. Hasil evaluasi menunjukkan kualitas klusterisasi yang sangat baik, dengan Silhouette Score sebesar 0,867 dan Davies-Bouldin Score sebesar 0,366. Visualisasi menggunakan diagram lingkaran dan PCA memperjelas distribusi serta pemisahan antar kluster. Segmentasi ini memberikan gambaran yang jelas mengenai potensi dan persebaran UMKM di tiap wilayah, sehingga dapat menjadi dasar yang kuat bagi pengambilan keputusan dan perumusan kebijakan pengembangan UMKM yang lebih terarah dan berbasis data.

Penelitian selanjutnya disarankan untuk memperkaya data dengan menambahkan variabel lain yang relevan, seperti akses pembiayaan, tingkat digitalisasi, atau faktor sosial ekonomi, agar segmentasi wilayah UMKM menjadi lebih komprehensif

DAFTAR PUSTAKA

- [1] Putri Salsabila Indrawan Lubis and Rofila Salsabila, "Peran UMKM (Usaha Mikro, Kecil, Dan Menengah) Dalam Meningkatkan Pembangunan Ekonomi Di Indonesia," *MUQADDIMAH J. Ekon. Manajemen, Akunt. dan Bisnis*, vol. 2, no. 2, pp. 91–110, 2024, doi: 10.59246/muqaddimah.v2i2.716.
- [2] D. Marcelina, A. Kurnia, and T. Terttiaavini, "Analisis Kluster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 293–301, 2023, doi: 10.57152/malcom.v3i2.952.
- [3] A. D. T. Putra *et al.*, "Pemetaan Usaha Mikro Kecil Menengah (UMKM) Desa Glesungrejo, Baturetno, Wonogiri Berbasis Web Map," *J. Pengabd. Masy. Bangsa*, vol. 2, no. 1, pp. 116–123, 2024, doi: 10.59837/jpmba.v2i1.779.
- [4] N. Syifa and R. N. Fahmi, "Implementasi Metode K-Means Clustering dalam Analisis Persebaran UMKM di Jawa Barat," *JOINS (Journal Inf. Syst.)*, vol. 6, no. 2, pp. 211–220, 2021, doi: 10.33633/joins.v6i2.5310.
- [5] 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.
- [6] A. Azzam, A. Irma Purnamasari, and I. Ali, "Implementasi Algoritma K-Means Clustering Untuk Analisis Persebaran Umkm Di Jawa Barat," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 3, pp. 3062–3070, 2024, doi: 10.36040/jati.v8i3.8450.
- [7] Haris Kurniawan, Sarjon Defit, and Sumijan, "Data Mining Menggunakan Metode K-Means

- Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal,” *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 2, pp. 80–89, 2020, doi: 10.52158/jacost.v1i2.102.
- [8] N. Wakhidah, “Clustering Menggunakan K-Means Algorithm (K-Means Algorithm Clustering),” *Fak. Teknol. Inf.*, vol. 21, no. 1, pp. 70–80, 2014.
- [9] F. Nasari and S. Darma, “Seminar Nasional Teknologi Informasi dan Multimedia 2015 PENERAPAN K-MEANS CLUSTERING PADA DATA PENERIMAAN MAHASISWA BARU (STUDI KASUS : UNIVERSITAS POTENSI UTAMA),” pp. 6–8, 2015.
- [10] B. M. Metisen and H. L. Sari, “Analisis clustering menggunakan metode K-Means dalam pengelompokan penjualan produk pada Swalayan Fadhila,” *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.
- [11] A. Ristyawan, A. Nugroho, and T. K. Amarya, “Optimasi Preprocessing Model Random Forest Untuk Prediksi Stroke,” vol. 12, no. 1, 2025.
- [12] M. Bakri, “Penerapan Data Mining untuk Clustering Kualitas Batu Bara dalam Proses Pembakaran di PLTU Sebalang Menggunakan Metode K-Means,” *J. Teknoinfo*, vol. 11, no. 1, p. 6, 2017, doi: 10.33365/jti.v11i1.3.
- [13] D. A. I. C. Dewi and D. A. K. Pramita, “Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali,” *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.
- [14] E. Muningsih, N. Hasan, and G. B. Sulistyono, “Penerapan Metode Principle Component Analysis (PCA) untuk Clustering Data Kunjungan Wisatawan Mancanegara ke Indonesia,” *Biaglala Inform.*, vol. 8, no. 1, pp. 58–62, 2020, doi: 10.31294/bi.v8i1.8470.
- [15] M. Rafi Nahjan, Nono Heryana, and Apriade Voutama, “Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 101–104, 2023, doi: 10.36040/jati.v7i1.6094.