

PENGELOMPOKAN UMKM BERDASARKAN KATEGORI USAHA DAN SEBARAN WILAYAH DI JAWA BARAT DENGAN *CLUSTERING K-MEANS*

Nana Casmana Ade Wikarta¹, Sofi Defiyanti²

^{1,2}Informatika, Universitas Singaperbangsa Karawang; Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat

Keywords:

UMKM;
K-Means;
PCA;
Clustering.

Correspondent Email:

2110631170127@student.unsika.ac.id



JITET is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Abstrak. Usaha Mikro, Kecil, dan Menengah (UMKM) merupakan sektor strategis dalam perekonomian daerah, termasuk di Provinsi Jawa Barat. Pada tahun 2023, jumlah UMKM di Indonesia mencapai 66 juta unit dan menyumbang 61% terhadap Produk Domestik Bruto (PDB) nasional. Pada penelitian ini bertujuan mengelompokkan kabupaten/kota di Jawa Barat berdasarkan kategori usaha dan proyeksi jumlah UMKM menggunakan metode *K-Means Clustering*, dengan bantuan *Principal Component Analysis* (PCA) untuk reduksi dimensi. Data bersumber dari portal Open Data Jabar dan mencakup proyeksi jumlah UMKM per kabupaten/kota dari tahun 2018 hingga 2023. Setiap baris data merepresentasikan satu kabupaten/kota per tahun, dengan sepuluh kategori usaha sebagai fitur utama. Hasil klasterisasi membentuk tiga *cluster*: *Cluster 0* (wilayah dengan rata-rata jumlah UMKM tertinggi), *Cluster 1* (terendah), dan *Cluster 2* (moderat). Komposisi jumlah anggota tiap *cluster* yang terbentuk adalah sebagai berikut: *Cluster 0* sebanyak 30 anggota, *Cluster 1* sebanyak 65 anggota, dan *Cluster 2* sebanyak 67 anggota. Evaluasi dengan *silhouette coefficient* menunjukkan nilai 0,58 yang menandakan pemisahan cluster cukup baik. Hasil ini diharapkan dapat menjadi referensi dalam merumuskan kebijakan pengembangan UMKM di tingkat daerah.

Abstract. Micro, Small, and Medium Enterprises (MSMEs) are a strategic sector in regional economic development, including in West Java Province. In 2023, the number of MSMEs in Indonesia reached 66 million units, contributing 61% to the national Gross Domestic Product (GDP). This study aims to cluster districts/cities in West Java based on business categories and projected MSME numbers using the *K-Means Clustering* method, with the assistance of *Principal Component Analysis* (PCA) for dimensionality reduction. The data was sourced from the Open Data Jabar portal and includes projected MSME data per district/city from 2018 to 2023. Each row in the dataset represents one district/city per year, with ten business categories as the main features. The clustering results formed three clusters: *Cluster 0* (regions with the highest average number of MSMEs), *Cluster 1* (lowest), and *Cluster 2* (moderate). The composition of cluster members is as follows: *Cluster 0* with 30 members, *Cluster 1* with 65 members, and *Cluster 2* with 67 members. Evaluation using the *silhouette coefficient* yielded a value of 0.58, indicating a fairly good cluster separation. These results are expected to serve as a reference for formulating regional MSME development policies.

1. PENDAHULUAN

Pembangunan ekonomi merupakan salah satu pilar utama dalam meningkatkan kesejahteraan masyarakat Indonesia. Dalam konteks ini, Usaha Mikro, Kecil, dan Menengah (UMKM) memainkan peran strategis dalam memperkuat struktur ekonomi nasional. UMKM didefinisikan sebagai usaha produktif yang dimiliki dan dikelola oleh individu atau kelompok dengan kriteria yang ditetapkan dalam UU No. 20 Tahun 2008. Selain menyediakan barang dan jasa, UMKM juga berperan penting dalam menciptakan lapangan kerja, mengurangi tingkat pengangguran, serta meningkatkan daya beli masyarakat [1]. Berdasarkan data dari Kementerian Koperasi dan UKM, terdapat sekitar 64,2 juta pelaku UMKM di Indonesia yang menyumbang 61,07% terhadap Produk Domestik Bruto (PDB) nasional dan menyerap 97% dari total angkatan kerja.

Jawa Barat sebagai salah satu provinsi dengan potensi ekonomi terbesar di Indonesia memiliki jumlah UMKM yang besar dan terus mengalami peningkatan. Data dari Badan Pusat Statistik menunjukkan bahwa jumlah UMKM di Jawa Barat meningkat dari 536.207 unit pada tahun 2018 menjadi 667.795 unit pada tahun 2022. UMKM di provinsi ini mencakup berbagai sektor usaha seperti industri kreatif, kuliner, fashion, dan jasa. Keragaman kategori usaha serta sebaran wilayah yang luas menunjukkan pentingnya adanya pendekatan berbasis data untuk memahami karakteristik dan potensi UMKM di setiap kabupaten/kota secara lebih spesifik.

Beberapa penelitian sebelumnya telah menerapkan metode *clustering*, khususnya algoritma *K-Means*, dalam berbagai konteks seperti evaluasi pembelajaran [2], perencanaan obat [3], hingga daerah rawan kekerasan anak [4]. Hasil penelitian-penelitian tersebut menunjukkan bahwa algoritma *K-Means* efektif dalam mengelompokkan data berdasarkan kemiripan karakteristik, dan dapat dievaluasi menggunakan metrik seperti *Silhouette Score*, *Sum of Squared Error (SSE)*, serta *David Bouldin Index (DBI)*. Selain itu, penggunaan *Principal Component Analysis (PCA)* sebagai teknik reduksi dimensi terbukti mampu meningkatkan performa *clustering* dengan mempertahankan informasi utama dari data [5] [6].

Penelitian ini bertujuan mengelompokkan UMKM di Jawa Barat berdasarkan kategori usaha dan sebaran wilayah menggunakan PCA dan algoritma *K-Means*. Visualisasi hasil klusterisasi melalui QGIS mendukung interpretasi spasial untuk perumusan kebijakan yang lebih tepat sasaran. Berbeda dari penelitian sebelumnya yang umumnya berfokus pada bidang sosial lainnya, studi ini mengisi kesenjangan dalam pemanfaatan data mining untuk pemberdayaan UMKM di tingkat provinsi.

2. TINJAUAN PUSTAKA

2.1 Data Mining

Data mining merupakan proses pengumpulan informasi penting dari dataset yang besar [7]. Secara sederhana, *data mining* merupakan proses menggali atau mengekstraksi pengetahuan dari kumpulan data berskala besar guna menghasilkan informasi yang bernilai dan dapat dimanfaatkan untuk pengembangan lebih lanjut [8].

2.2 Knowledge Discovery in Databases

Knowledge Discovery in Databases adalah serangkaian proses terstruktur yang memiliki tujuan untuk mengidentifikasi pengetahuan baru dan bermanfaat dari data yang tersedia [9]. Berikut adalah tahapan dalam proses *Knowledge Discovery in Databases* [10]:

1. *Data Selection*
2. *Data Preprocessing*
3. *Data Transformation*
4. *Data Mining*
5. *Evaluation*

2.3 Clustering

Clustering adalah salah satu teknik dalam *unsupervised learning*, di mana prosesnya tidak memerlukan pelatihan model atau fase pembelajaran seperti dalam *supervised learning* [11]. Selanjutnya, tujuan utama *clustering* adalah mengelompokkan data atau objek berdasarkan kesamaan sehingga setiap kelompok berisi data yang memiliki karakteristik yang sangat mirip. Melalui pendekatan ini, dapat dilihat pola, struktur, atau hubungan tersembunyi dalam data yang sebelumnya tidak teridentifikasi. Proses

klasterisasi dapat membantu dalam pengambilan keputusan, segmentasi pasar, atau penemuan informasi penting yang bisa memberikan nilai lebih dalam berbagai bidang analisis data [12].

2.4 K-Means

Algoritma *K-Means* merupakan salah satu metode *clustering* yang bertujuan untuk membagi sekumpulan data ke dalam sejumlah *cluster* tertentu. Dalam proses ini, data yang memiliki kemiripan karakteristik akan dikelompokkan ke dalam satu *cluster* yang sama, sedangkan data yang memiliki perbedaan karakteristik akan ditempatkan pada *cluster* yang berbeda. Pendekatan ini digunakan untuk mengidentifikasi struktur atau pola alami dalam data tanpa memerlukan label kelas [13]. Algoritma *k-means* juga termasuk dalam kategori algoritma *clustering non-hirarkis*. Prosesnya dimulai dengan menentukan sejumlah *cluster* di awal, kemudian dilakukan pengelompokan data secara iteratif hingga tidak ada lagi perubahan dalam pembagian *cluster* tersebut [14]. Berikut adalah rumus yang digunakan pada tahap iterasi [15]:

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

Keterangan:

V_{ij} : titik pusat rata – rata *cluster* ke i untuk variabel ke j

n_i : banyaknya data pada *cluster* ke i

i, k : indeks dari *cluster*

j : indeks dari *variable*

X_{Kj} : nilai data ke k *variable* j pada *cluster* tersebut

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

Pada tahap ini, setiap objek diukur jaraknya terhadap titik *centroid* menggunakan persamaan di atas.

Keterangan:

De : jarak antara objek dan *centroid*

i : jumlah objek

(x_i, y_i) : koordinat objek

(s_i, t_i) : koordinat *centroid*

Setelah jarak setiap objek ke *centroid* dihitung, objek dikelompokkan ke *centroid* terdekat. Proses ini diulang terus hingga posisi *centroid* mencapai keadaan paling optimal.

2.5 Elbow Method

Metode elbow merupakan salah satu teknik yang umum digunakan dalam analisis *cluster* untuk menentukan jumlah *cluster* optimal dengan prinsip utama dari metode ini adalah mengidentifikasi titik di mana penambahan jumlah klaster memberikan dampak yang semakin kecil pada penurunan nilai *within-cluster sum of squares* (WCSS), yang menggambarkan jarak antara data dalam satu *cluster* [16]. Dalam grafik, titik tersebut terlihat sebagai “siku” pada kurva, yang menjadi indikator jumlah *cluster* yang optimal.

2.6 Silhouette Coefficient

Silhouette coefficient adalah metode untuk mengevaluasi atau memvalidasi algoritma *clustering* yang digunakan untuk mengukur kualitas *cluster* yang dihasilkan [17]. Metode ini menggambarkan seberapa baik data dikelompokkan ke dalam *cluster* tertentu.

Berikut adalah kriteria subjektif pada pengukuran pengelompokan berdasarkan *silhouette coefficient* [18], dapat dilihat pada Tabel 1 berikut:

Tabel 1. Kriteria Pengukuran *Silhouette Coefficient*

Nilai SC	Kriteria
0,71 – 1,00	Struktur kuat
0,51 – 0,70	Struktur baik
0,26 – 0,50	Struktur lemah
≤ 25	Struktur buruk

2.7 Principal Component Analysis

Principal Component Analysis (PCA) merupakan suatu metode statistik yang digunakan untuk mereduksi dimensi data. Teknik ini bekerja dengan mentransformasi variabel-variabel awal ke dalam sejumlah variabel baru yang disebut sebagai *principal components*, di mana masing-masing komponen utama tersebut bersifat tidak saling berkorelasi (independen) satu sama lain. PCA sering digunakan dalam analisis data untuk mempermudah visualisasi, mengurangi kompleksitas, dan meningkatkan efisiensi pemrosesan data tanpa kehilangan informasi penting [19].

Berikut adalah langkah – langkah untuk memperoleh komponen utama dari suatu data, dilakukan tahapan berikut [6]:

1. Data terlebih dahulu dinormalisasi agar setiap variabel memiliki skala yang sebanding, biasanya menggunakan metode *Z-score*.

2. Hitung matriks kovarians guna mengetahui hubungan antarvariabel dalam data.
3. Perhitungan Nilai Eigen dan Vektor Eigen. *Eigenvalue* diperoleh dari penyelesaian:

$$|\lambda I - R| = 0 \quad (3)$$

dan *Eigenvector* dihitung menggunakan:

$$R\vec{v} = \lambda\vec{v} \quad (4)$$

4. Pilih komponen berdasarkan nilai *eigenvalue* yang ≥ 1 , untuk mempertahankan variabilitas data yang signifikan.

5. Data diubah ke bentuk baru berdasarkan kombinasi linier dari komponen utama:

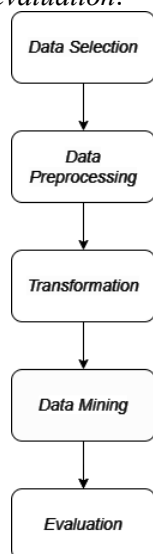
$$PC_{at} = \vec{v}_{1a}Z_1 + \vec{v}_{2a}Z_2 + \dots + \vec{v}_{ap}Z_p \quad (5)$$

6. Membentuk komponen dalam matriks korelasi yang menggambarkan tingkat hubungan antara variabel dengan skor komponen yang terbentuk menggunakan rumus:

$$rx_p, PC_t = \vec{v}_{1a}\sqrt{\lambda_t} \quad (6)$$

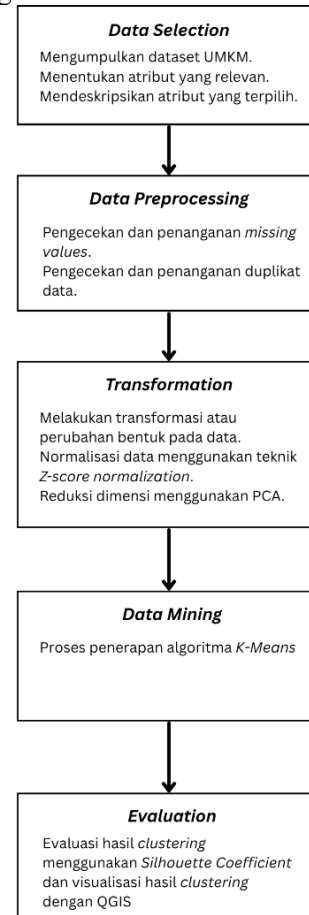
3. METODE PENELITIAN

Penelitian ini menerapkan pendekatan *Knowledge Discovery in Database* (KDD), yaitu proses sistematis untuk mengubah data menjadi informasi dengan mengidentifikasi pola – pola tertentu. Pendekatan ini dirancang untuk menggali informasi berharga yang tersembunyi dalam suatu dataset dan sebelumnya tidak diketahui potensinya. Tahapan dalam proses ini mencakup *data selection*, *data preprocessing*, *transformation*, *data mining*, dan *evaluation*.



Gambar 1. Diagram Alur Knowledge Discovery in Database

Untuk melakukan pengelompokan UMKM berdasarkan kategori usaha dan sebaran wilayah di Jawa Barat, setiap tahapan disesuaikan dengan langkah-langkah pada model KDD. Alur rancangan penelitian dapat dilihat pada gambar berikut:



Gambar 2. Diagram Alur Rancangan Penelitian

4. HASIL DAN PEMBAHASAN

4.1 Data Selection

Data yang digunakan pada penelitian ini, yaitu data UMKM kabupaten/kota di Jawa Barat dengan rentang waktu 2018 – 2023 yang diperoleh dari *website* OpenDataJabar dan terdiri dari 1620 *entry data* mentah. Data tersebut terdiri dari kolom ID, kode_provinsi, nama_provinsi, kode_kabupaten_kota, nama_kabupaten_kota, kategori_usaha, proyeksi_jumlah_umkm, satuan, dan tahun.

Sebagai kelanjutan dari pemilihan data, pada tabel 2 menampilkan dan mendeskripsikan atribut – atribut yang dipilih untuk analisis lebih lanjut. Tabel 2 memberikan gambaran tentang data yang digunakan dalam klasterisasi UMKM berdasarkan kategori usaha

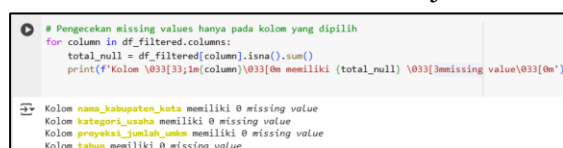
di Jawa Barat, dengan hanya memilih atribut – atribut ini, data yang digunakan menjadi lebih fokus dan relevan untuk tujuan penelitian, serta mempermudah proses analisis lebih lanjut pada tahap *preprocessing*.

Tabel 2. Deskripsi Atribut

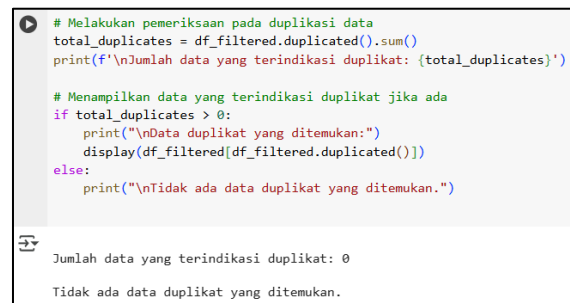
Atribut	Tipe Data	Deskripsi
Nama_kabupaten_kota	Kategorikal	Atribut ini menunjukkan lokasi geografis dari UMKM yang tercatat dalam dataset, yaitu kabupaten atau kota di wilayah Jawa Barat.
Kategori_usaha	Kategorikal	Atribut ini mencakup jenis usaha yang dijalankan oleh UMKM di masing-masing kabupaten/kota.
Proyeksi_jumlah_umkm	Numerikal	Atribut ini menunjukkan jumlah UMKM yang terdaftar dalam setiap kategori usaha di setiap kabupaten/kota pada tahun tertentu.
Tahun	Numerikal	Atribut ini mencatat tahun data UMKM yang tercatat.

4.2 Data Preprocessing

Pada tahap *data preprocessing*, dilakukan serangkaian proses pemeriksaan terhadap *missing values* dan data duplikat guna memastikan kualitas dan konsistensi data sebelum dilakukan analisis lebih lanjut.



Gambar 3. Pengecekan Missing Values



Gambar 4. Pengecekan Duplikat Data

Hasilnya menunjukkan tidak ada nilai kosong maupun duplikasi pada atribut utama, sehingga data dinyatakan bersih dan siap untuk tahap transformasi dan reduksi dimensi menggunakan PCA sebelum diterapkan algoritma *K-Means*.

4.3 Data Transformation

Pada tahap awal *data transformation*, dilakukan penyesuaian bentuk pada data untuk memudahkan proses *data mining*. Nantinya, setiap kategori usaha akan menjadi atribut baru (total 10 kolom kategori usaha) dan nilai dalam setiap sel kategori usahanya diambil dari kolom proyeksi_jumlah_umkm. Proses ini bertujuan untuk melihat *cluster* yang terbentuk di setiap wilayah pada tahun tertentu.

nama_kabupaten_kota	tahun	AKSESORIS	BATIK	BORDIR	CRAFT	FASHION	JASA/LATMANYA	KONSEKSI	KULNER	MAKANAN	MINUMAN
KABUPATEN BANDUNG	2018	927	527	132	33118	32323	45041	19738	142541	104256	15041
KABUPATEN BANDUNG BARAT	2018	410	410	59	14851	14300	19926	8732	63059	46122	8506
KABUPATEN BEKASI	2018	606	606	87	21659	21139	29456	12909	93221	68183	12649
KABUPATEN BOGOR	2018	984	984	141	35159	34315	47816	20955	151325	110681	20533
KABUPATEN CIAMIS	2018	367	367	52	13098	12764	17813	7805	56374	41233	7649
KOTA CIMAHI	2023	202	202	29	7203	7030	9796	4293	31001	22674	4206
KOTA DEPOK	2023	575	575	82	20553	20059	27952	12249	88458	64700	12003
KOTA BOGOR	2023	306	306	44	10936	10673	14873	6518	47068	34426	6387
KOTA SUKABUMI	2023	142	142	20	5060	4939	6882	3016	21780	15930	2905
KOTA TASIKMALAYA	2023	323	323	46	11532	11255	15683	6873	49632	36302	6735

Gambar 2. Hasil Transformasi Data

Proses berikutnya, yaitu proses standarisasi menggunakan *StandardScaler*, hal ini dilakukan untuk mengubah skala data agar dapat dibandingkan secara lebih adil antar kategori.

kategori_usaha	nama_kabupaten_kota	tahun	AKSESORIS	BATIK	BORDIR	CRAFT	FASHION	JASA/LATMANYA	KONSEKSI	KULNER	MAKANAN	MINUMAN
0	KABUPATEN BANDUNG	2018	1.314622	1.314622	1.304334	1.315608	1.315697	1.315643	1.315544	1.315850	1.315621	1.315614
1	KABUPATEN BANDUNG BARAT	2018	-0.381464	-0.381464	-0.372485	-0.380769	-0.380788	-0.380824	-0.380806	-0.380810	-0.380860	-0.380860
2	KABUPATEN BEKASI	2018	0.261559	0.261559	0.279679	0.262947	0.262933	0.262931	0.262962	0.262961	0.262960	0.262967
3	KABUPATEN BOGOR	2018	1.501618	1.501618	1.510165	1.503101	1.503068	1.503064	1.503122	1.503113	1.503112	1.503115
4	KABUPATEN CIAMIS	2018	-0.522532	-0.522532	-0.533275	-0.523494	-0.523459	-0.523512	-0.523495	-0.523488	-0.523478	-0.523535
5	KOTA CIMAHI	2023	-1.060326	-1.060326	-1.061088	-1.060328	-1.060328	-1.060324	-1.060312	-1.060341	-1.060356	-1.060312
6	KOTA DEPOK	2023	0.159840	0.159840	0.159838	0.161346	0.161281	0.161341	0.161270	0.161301	0.161322	0.161381
7	KOTA BOGOR	2023	-0.722650	-0.722650	-0.717036	-0.722102	-0.722151	-0.722100	-0.722070	-0.722112	-0.722116	-0.722042
8	KOTA SUKABUMI	2023	-1.260674	-1.260674	-1.260319	-1.261891	-1.261848	-1.261864	-1.261838	-1.261851	-1.261855	-1.261884
9	KOTA TASIKMALAYA	2023	-0.666879	-0.666879	-0.671036	-0.667332	-0.667372	-0.667387	-0.667394	-0.667387	-0.667372	-0.667332

Gambar 6. Hasil Standarisasi Data

Hasil dari proses standarisasi data ini menunjukkan nilai-nilai yang telah diubah ke dalam skala yang seragam (*mean* = 0 dan *standard deviation* = 1). Setiap kategori usaha (seperti aksesoris, batik, bordir, dan lainnya) di

setiap kabupaten/kota pada tahun 2018 hingga 2023 telah diproses dengan menggunakan metode standarisasi.

Langkah selanjutnya merupakan proses diterapkannya *Principal Component Analysis* untuk mereduksi dimensi data proyeksi jumlah UMKM yang tersedia dalam dataset. Tujuan utama dari penggunaan PCA adalah untuk mengidentifikasi komponen utama yang menyaring informasi penting dari data yang memiliki dimensi tinggi, sehingga lebih mudah dianalisis dan divisualisasikan.

```
from sklearn.decomposition import PCA
import pandas as pd

# Lowercase kolom pivoted_data
pivoted_data.columns = pivoted_data.columns.str.lower()

# Definisi kategori usaha dari kolom ke-3 sampai akhir (kolom 2: karena index 0 = nama_kabupaten_kota, 1 = tahun)
kategori_umkm = pivoted_data.columns[2:]

# Inisialisasi PCA dengan 2 komponen
pca = PCA(n_components=2)

# Terapkan PCA ke kolom kategori UMKM
pca_result = pca.fit_transform(pivoted_data[kategori_umkm])

# Simpan hasil PCA dalam DataFrame
pca_df = pd.DataFrame(pca_result, columns=['pc1', 'pc2'])

# Gabungkan dengan data awal
data_all_pca = pd.concat([pivoted_data[['nama_kabupaten_kota', 'tahun']], pca_df, pivoted_data[kategori_umkm]], axis=1)
```

Gambar 7. Proses Penerapan PCA

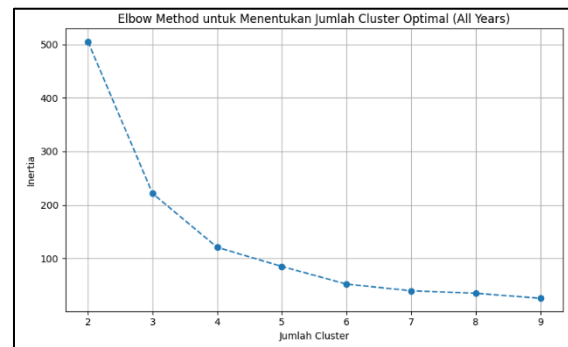
Gambar 8 merupakan hasil dari proses penerapan *Principal Component Analysis* (PCA) pada data proyeksi jumlah UMKM untuk setiap kabupaten/kota di Provinsi Jawa Barat pada tahun 2018 hingga 2023. Hasil PCA ini menunjukkan dua komponen utama (pc1 dan pc2) yang mewakili variasi data dalam dataset.

	nama_kabupaten_kota	tahun	pc1	pc2
0	KABUPATEN BANDUNG	2018	4.156129	-0.010457
1	KABUPATEN BANDUNG BARAT	2018	-1.202014	0.008038
2	KABUPATEN BEKASI	2018	0.833082	0.007644
3	KABUPATEN BOGOR	2018	4.754812	0.007903
4	KABUPATEN CIAMIS	2018	-1.657935	-0.009494
...	...	...	...	...
157	KOTA CIMAHI	2023	-3.366112	0.002997
158	KOTA DEPOK	2023	0.507452	-0.004877
159	KOTA BOGOR	2023	-2.282226	0.004913
160	KOTA SUKABUMI	2023	-3.991650	-0.006407
161	KOTA TASIKMALAYA	2023	-2.111257	-0.003660

Gambar 8. Hasil Penerapan PCA

4.4 Data Mining

Langkah pertama dalam tahapan *data mining*, yaitu mencari nilai K optimal menggunakan *elbow method*. Berdasarkan hasil dari *elbow method* pada gambar 9, jumlah *cluster* yang optimal = 3 *clusters* untuk pengelompokan UMKM di Jawa Barat.



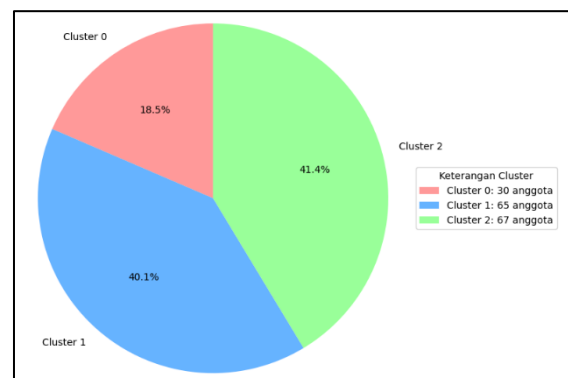
Gambar 9. Hasil Elbow Method

Selanjutnya, dilakukan penerapan algoritma *K-Means* berdasarkan dua komponen utama hasil penerapan *Principal Component Analysis* (PCA), yaitu pc1 dan pc2 yang merepresentasikan variansi terbesar dari data kategori usaha UMKM pada setiap kabupaten/kota untuk setiap tahun. Berdasarkan Gambar 10, dapat diketahui bahwa:

Cluster 0 terdiri dari 30 anggota atau sekitar 18.5% dari total data.

Cluster 1 terdiri dari 65 anggota atau sekitar 40.1% dari total data.

Cluster 2 terdiri dari 67 anggota atau sekitar 41.4% dari total data.



Gambar 10. Hasil Clustering K-Means

Tabel 3 menunjukkan karakteristik masing-masing *cluster* hasil pengelompokan wilayah di Jawa Barat berdasarkan data UMKM tahun 2018 – 2023.

Tabel 3. Karakteristik Cluster

Cluster	Deskripsi
0	Cluster dengan rata-rata UMKM tertinggi. Menggambarkan wilayah dengan aktivitas UMKM sangat tinggi di berbagai kategori. Walaupun total UMKM tidak sebanyak cluster 2, tetapi rata – rata keseluruhan seluruh kategori usaha dalam cluster 0 ini memperoleh nilai tertinggi.

1	Cluster dengan rata-rata UMKM terendah. Wilayah dengan perkembangan UMKM yang lambat atau terbatas.
2	Cluster dengan pertumbuhan UMKM cukup kuat, dilihat dari jumlah total UMKM mencapai 17.302.351. Menunjukkan aktivitas UMKM yang sedang berkembang.

Tabel 4 memperlihatkan hasil pengelompokan Kabupaten/Kota di Jawa Barat tahun 2018 ke dalam tiga *cluster*, yakni *cluster* 0 (3 wilayah), *cluster* 1 (13 wilayah), dan *cluster* 2 (11 wilayah).

Tabel 4. Hasil *Cluster* Tahun 2018

Cluster	Nama Kabupaten/Kota
0	Kabupaten Bandung, Kabupaten Bogor, Kota Bandung
1	Kabupaten Bandung Barat, Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Pangandaran, Kabupaten Majalengka, Kabupaten Purwakarta, Kabupaten Sumedang, Kota Banjar, Kota Bogor, Kota Cimahi, Kota Cirebon, Kota Sukabumi, Kota Tasikmalaya.
2	Kabupaten Bekasi, Kabupaten Cianjur, Kabupaten Cirebon, Kabupaten Garut, Kabupaten Indramayu, Kabupaten Karawang, Kabupaten Subang, Kabupaten Sukabumi, Kabupaten Tasikmalaya, Kota Bekasi, Kota Depok

Tabel 5 memperlihatkan hasil pengelompokan Kabupaten/Kota di Jawa Barat tahun 2019 ke dalam tiga *cluster*, yakni *cluster* 0 (3 wilayah), *cluster* 1 (11 wilayah), dan *cluster* 2 (13 wilayah).

Tabel 5. Hasil *Cluster* Tahun 2019

Cluster	Nama Kabupaten/Kota
0	Kabupaten Bandung, Kabupaten Bogor, Kota Bandung
1	Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Pangandaran, Kabupaten Purwakarta, Kabupaten Sumedang, Kota Banjar, Kota Bogor, Kota Cimahi, Kota Cirebon, Kota Sukabumi, Kota Tasikmalaya.
2	Kabupaten Bandung Barat, Kabupaten Bekasi, Kabupaten Cianjur, Kabupaten Cirebon, Kabupaten Garut, Kabupaten Indramayu, Kabupaten Karawang, Kabupaten Majalengka, Kabupaten Subang, Kabupaten Sukabumi, Kabupaten Tasikmalaya, Kota Bekasi, Kota Depok

Tabel 6 memperlihatkan hasil pengelompokan Kabupaten/Kota di Jawa Barat tahun 2020 ke dalam tiga *cluster*, yakni *cluster*

0 (3 wilayah), *cluster* 1 (11 wilayah), dan *cluster* 2 (13 wilayah).

Tabel 6. Hasil *Cluster* Tahun 2020

Cluster	Nama Kabupaten/Kota
0	Kabupaten Bandung, Kabupaten Bogor, Kota Bandung
1	Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Pangandaran, Kabupaten Purwakarta, Kabupaten Sumedang, Kota Banjar, Kota Bogor, Kota Cimahi, Kota Cirebon, Kota Sukabumi, Kota Tasikmalaya.
2	Kabupaten Bandung Barat, Kabupaten Bekasi, Kabupaten Cianjur, Kabupaten Cirebon, Kabupaten Garut, Kabupaten Indramayu, Kabupaten Karawang, Kabupaten Majalengka, Kabupaten Subang, Kabupaten Sukabumi, Kabupaten Tasikmalaya, Kota Bekasi, Kota Depok

Tabel 7 memperlihatkan hasil pengelompokan Kabupaten/Kota di Jawa Barat tahun 2021 ke dalam tiga *cluster*, yakni *cluster* 0 (5 wilayah), *cluster* 1 (10 wilayah), dan *cluster* 2 (12 wilayah).

Tabel 7. Hasil *Cluster* Tahun 2021

Cluster	Nama Kabupaten/Kota
0	Kabupaten Bandung, Kabupaten Bogor, Kota Bandung, Kabupaten Garut, Kabupaten Sukabumi
1	Kabupaten Kuningan, Kabupaten Pangandaran, Kabupaten Purwakarta, Kabupaten Sumedang, Kota Banjar, Kota Bogor, Kota Cimahi, Kota Cirebon, Kota Sukabumi, Kota Tasikmalaya.
2	Kabupaten Ciamis, Kabupaten Bandung Barat, Kabupaten Bekasi, Kabupaten Cianjur, Kabupaten Cirebon, Kabupaten Indramayu, Kabupaten Karawang, Kabupaten Majalengka, Kabupaten Subang, Kabupaten Tasikmalaya, Kota Bekasi, Kota Depok

Tabel 8 memperlihatkan hasil pengelompokan Kabupaten/Kota di Jawa Barat tahun 2022 ke dalam tiga *cluster*, yakni *cluster* 0 (7 wilayah), *cluster* 1 (10 wilayah), dan *cluster* 2 (10 wilayah).

Tabel 8. Hasil *Cluster* Tahun 2022

Cluster	Nama Kabupaten/Kota
0	Kabupaten Bandung, Kabupaten Bogor, Kota Bandung, Kabupaten Garut, Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Cirebon
1	Kabupaten Kuningan, Kabupaten Pangandaran, Kabupaten Purwakarta, Kabupaten Sumedang, Kota Banjar, Kota

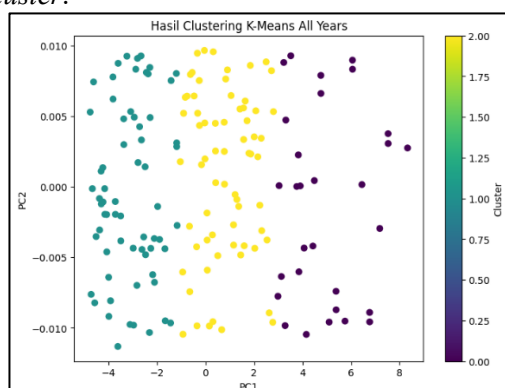
	Bogor, Kota Cimahi, Kota Cirebon, Kota Sukabumi, Kota Tasikmalaya.
2	Kabupaten Ciamis, Kabupaten Bandung Barat, Kabupaten Bekasi, Kabupaten Indramayu, Kabupaten Karawang, Kabupaten Majalengka, Kabupaten Subang, Kabupaten Tasikmalaya, Kota Bekasi, Kota Depok

Tabel 9 memperlihatkan hasil pengelompokan Kabupaten/Kota di Jawa Barat tahun 2023 ke dalam tiga *cluster*, yakni *cluster* 0 (9 wilayah), *cluster* 1 (10 wilayah), dan *cluster* 2 (8 wilayah).

Tabel 9. Hasil *Cluster* Tahun 2023

Cluster	Nama Kabupaten/Kota
0	Kabupaten Bandung, Kabupaten Bogor, Kota Bandung, Kabupaten Garut, Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Cirebon, Kabupaten Karawang, Kabupaten Bekasi
1	Kabupaten Kuningan, Kabupaten Pangandaran, Kabupaten Purwakarta, Kabupaten Sumedang, Kota Banjar, Kota Bogor, Kota Cimahi, Kota Cirebon, Kota Sukabumi, Kota Tasikmalaya.
2	Kabupaten Ciamis, Kabupaten Bandung Barat, Kabupaten Indramayu, Kabupaten Majalengka, Kabupaten Subang, Kabupaten Tasikmalaya, Kota Bekasi, Kota Depok

Gambar 11 memperlihatkan visualisasi hasil *clustering* menggunakan algoritma *K-Means* setelah reduksi dimensi dengan PCA. Tiga kluster teridentifikasi dengan jelas pada *scatter plot*, masing-masing dengan distribusi berbeda pada komponen utama (PC1 dan PC2). *Cluster* 0 ditandai dengan warna kuning dan terletak di bagian bawah grafik, *Cluster* 1 berwarna hijau di bagian tengah, dan *Cluster* 2 berwarna ungu di kanan atas. Pemisahan yang tegas ini menunjukkan adanya perbedaan karakteristik yang signifikan di antara ketiga *cluster*.



Gambar 11. Visualisasi *Scatter Plot* Hasil *Clustering K-Means*

4.5 Evaluation

Pada tahap evaluasi dalam metodologi *Knowledge Discovery in Database* (KDD), kualitas hasil *clustering* dinilai menggunakan metode *silhouette coefficient*. Berdasarkan Gambar 12, diperoleh nilai *silhouette score* sebesar 0,58. Nilai tersebut mengindikasikan bahwa pengelompokan yang dihasilkan berada dalam kategori baik, karena mendekati nilai maksimum yaitu 1.

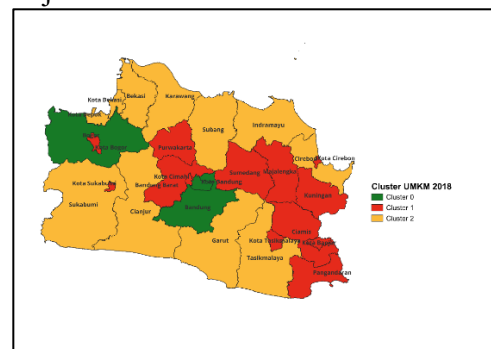
```
[ ] # Hitung Silhouette Score berdasarkan fitur numerik dan label cluster
silhouette_avg = silhouette_score(data_all_pca[['pc1', 'pc2']], data_all_pca['cluster'])

# Tampilkan hasil
print(f"Silhouette Score untuk 3 cluster: {silhouette_avg:.4f}")

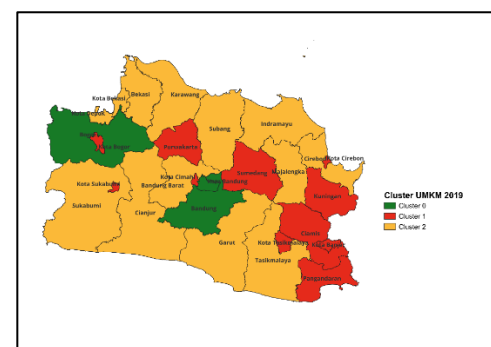
Silhouette Score untuk 3 cluster: 0.5828
```

Gambar 12. Hasil *Evaluasi* dengan *Silhouette Coefficient*

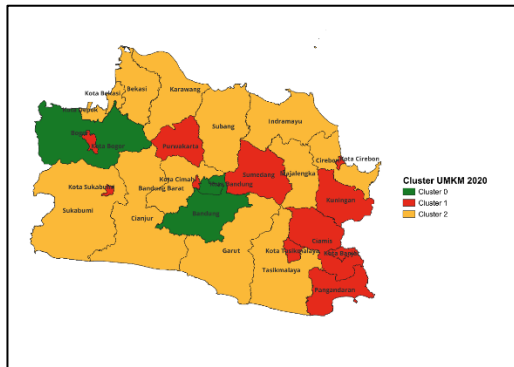
Peta-peta berikut menyajikan visualisasi hasil pengelompokan data UMKM di Provinsi Jawa Barat dari tahun 2018 hingga 2023. Visualisasi ini dihasilkan menggunakan perangkat lunak QGIS dan menggambarkan distribusi spasial kluster UMKM di setiap kabupaten/kota. Melalui peta ini, pola perkembangan dan perbedaan karakteristik UMKM antar wilayah dapat diamati secara lebih jelas dari tahun ke tahun.



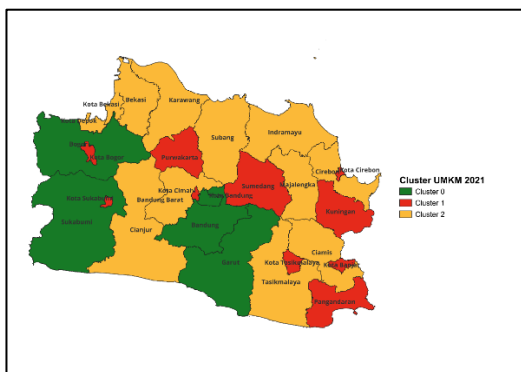
Gambar 13. Peta Distribusi UMKM Jawa Barat Tahun 2018



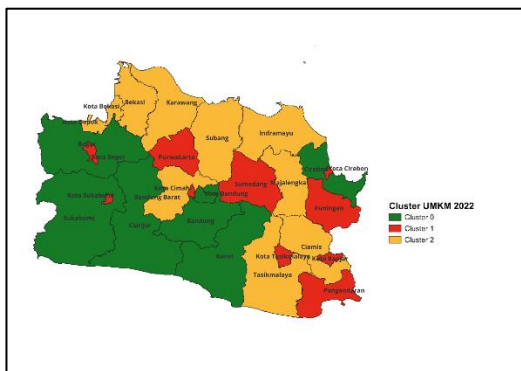
Gambar 14. Peta Distribusi UMKM Jawa Barat Tahun 2019



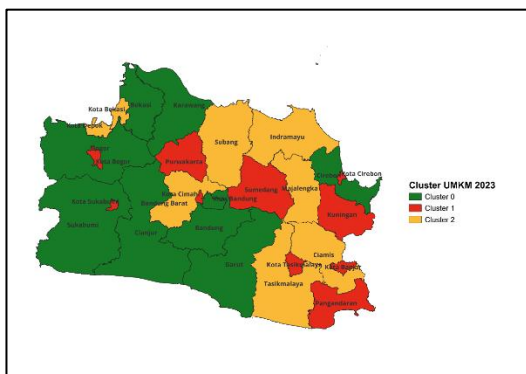
Gambar 15. Peta Distribusi UMKM Jawa Barat Tahun 2020



Gambar 16. Peta Distribusi UMKM Jawa Barat Tahun 2021



Gambar 17. Peta Distribusi UMKM Jawa Barat Tahun 2022



Gambar 18. Peta Distribusi UMKM Jawa Barat Tahun 2023

Berdasarkan hasil *clustering* data UMKM di Provinsi Jawa Barat selama periode 2018 hingga 2023, terlihat adanya dinamika yang cukup signifikan dalam komposisi dan pertumbuhan masing-masing *cluster*. *Cluster* 0, yang merepresentasikan wilayah dengan rata-rata UMKM tertinggi di berbagai kategori usaha, mengalami peningkatan jumlah anggota dari 3 wilayah pada tahun 2018 menjadi 9 wilayah pada tahun 2023. Kenaikan ini mencerminkan adanya penguatan kualitas UMKM di sejumlah daerah.

Sementara itu, *cluster* 1 yang menggambarkan wilayah dengan rata-rata UMKM terendah dan pertumbuhan yang cenderung lambat atau terbatas, relatif stabil dalam jumlah wilayah dari tahun ke tahun, dengan kisaran 10 hingga 13 daerah. Stabilitas ini dapat diartikan sebagai tantangan dalam pengembangan UMKM di wilayah-wilayah tersebut, yang membutuhkan perhatian khusus dari sisi kebijakan maupun pendampingan usaha.

Adapun *cluster* 2, yang diidentifikasi sebagai *cluster* dengan jumlah total UMKM terbesar dan mencerminkan wilayah dengan pertumbuhan UMKM yang cukup kuat, juga menunjukkan tren yang fluktuatif. Pada awal periode (2018–2020), *cluster* ini mendominasi jumlah wilayah, namun mulai tahun 2021 hingga 2023, sebagian wilayah mulai berpindah ke *cluster* 0.

Secara keseluruhan, tren selama enam tahun terakhir menunjukkan adanya perbaikan kualitas UMKM di berbagai wilayah Jawa Barat, ditandai dengan bertambahnya jumlah kabupaten/kota yang masuk ke dalam *cluster* 0. Sementara itu, wilayah dalam *cluster* 1 memerlukan intervensi lebih lanjut agar dapat mengejar ketertinggalan dan bergeser ke *cluster* dengan pertumbuhan yang lebih tinggi di masa mendatang.

5. KESIMPULAN

- Penelitian ini berhasil melakukan pengelompokan UMKM di kabupaten/kota Jawa Barat menggunakan algoritma *K-Means clustering* berdasarkan kategori usaha dan sebaran wilayah. Data UMKM dikelompokkan ke dalam tiga *cluster* optimal yang mencerminkan karakteristik dan perkembangan UMKM di masing-masing wilayah. Teknik *Principal*

Component Analysis (PCA) digunakan untuk mengurangi dimensi data agar proses *clustering* menjadi lebih efektif dan hasilnya lebih mudah diinterpretasikan.

- b. Hasil evaluasi dengan metode *silhouette coefficient* memberikan nilai sebesar 0.58 yang menunjukkan bahwa pemisahan *cluster* cukup baik dan layak.
- c. Hasil visualisasi pemetaan segmentasi UMKM di Kota/Kabupaten Jawa Barat selama periode 2018-2023 dengan QGIS, dapat disimpulkan bahwa terdapat pola pengelompokan wilayah yang relatif stabil, namun dengan beberapa perubahan komposisi *cluster* dari tahun ke tahun. Dengan melihat perubahan komposisi wilayah dalam setiap *cluster* dari tahun ke tahun, dapat diidentifikasi daerah-daerah yang mengalami peningkatan aktivitas UMKM, terutama wilayah yang beralih ke *cluster* 0 dengan rata-rata UMKM tertinggi. Temuan ini memberikan dasar yang kuat bagi pemerintah daerah dan provinsi dalam merumuskan kebijakan pengembangan UMKM yang lebih terfokus dan tepat sasaran.

DAFTAR PUSTAKA

- [1] B. Arifitama and A. Syahputra, "Analisis Data Mining Pada Klasterisasi UMKM Dengan Menggunakan Algoritma K-Means," *J. Ind. Kreat. dan Inform.*, vol. 2, no. 2, pp. 66–72, 2022.
- [2] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *J. Ilm. Inform. Dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024, [Online]. Available: <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- [3] T. A. Aria, M. Julkarnain, and F. Hamdani, "Penerapan Algoritma K-Means Clustering Untuk Data Obat," vol. 4, no. 1, pp. 649–657, 2023, doi: 10.30865/klik.v4i1.1117.
- [4] H. Sa'diah, U. Enri, and T. Nur Padilah, "Penerapan Algoritma K-Means Dalam Segmentasi Daerah Rawan Kekerasan Anak Di Jawa Barat," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 2, pp. 1351–1357, 2023, doi: 10.36040/jati.v7i2.6838.
- [5] M. Kandias, N. C. Wibowo, and A. R. E. Najaf, "PENERAPAN ALGORITMA K-MEANS UNTUK SEGMENTASI DAERAH DI JAWA TIMUR BERDASARKAN INDIKATOR KESEJAHTERAAN MASYARAKAT," vol. 2, pp. 21–52, 2024.
- [6] I. A. Rosyada and D. T. Utari, "Penerapan Principal Component Analysis untuk Reduksi Variabel pada Algoritma K-Means Clustering," *Jambura J. Probab. Stat.*, vol. 5, no. 1, pp. 6–13, 2024, doi: 10.37905/jjps.v5i1.18733.
- [7] R. S. Wahono, *Data Mining Data mining*, vol. 2, no. January 2013. 2023. [Online]. Available: https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part
- [8] N. Yulia, R. Saragih, and I. Ambarita, "Data Mining Pengelompokan Anak Stunting Berdasarkan Usia, Penyebab dan Pekerjaan Orang Tua Dengan Menggunakan Metode Clustering (Studi Kasus: Dinas Kesehatan Kabupaten Langkat)," *Semin. Nas. Inform. (SENATIKA) Prosiding Senat. 2021*, p. 12, 2021, [Online]. Available: <http://www.ejournal.pelitaindonesia.ac.id/ojs32/index.php/SENATIKA/article/view/1174/661>
- [9] R. Bahtiar, "Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor," *J. Inform. MULTI*, vol. 1, no. 3, pp. 203–214, 2023, [Online]. Available: <https://jurnal.publikasitecno.id/index.php/jim203>
- [10] F. Alghifari and D. Juardi, "Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes," *J. Ilm. Inform.*, vol. 9, no. 02, pp. 75–81, 2021, doi: 10.33884/jif.v9i02.3755.
- [11] Y. Elda, S. Defit, Y. Yunus, and R. Syaljumairi, "Klasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan K-Means," *J. Inf. dan Teknol.*, vol. 3, pp. 103–108, 2021, doi: 10.37034/jidt.v3i3.130.
- [12] M. R. Nugroho, I. E. Hendrawan, and P. P. Purwantoro, "Penerapan Algoritma K-Means Untuk Klasterisasi Data Obat Pada Rumah Sakit ASRI," *Nuansa Inform.*, vol. 16, no. 1, pp. 125–133, 2022, doi: 10.25134/nuansa.v16i1.5294.
- [13] R. Fauziah and A. I. Purnamasari, "Implementasi Algoritma K-Means pada Kasus Kekerasan Anak dan Perempuan Berdasarkan Usia," *Hello World J. Ilmu Komput.*, vol. 2, no. 1, pp. 34–41, 2023, doi: 10.56211/helloworld.v2i1.232.
- [14] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi: 10.33365/jtk.v15i2.1162.
- [15] M. Amelia, A. Faqih, A. R. Rinaldi, and K.

- Sosial, “PENERAPAN METODE K-MEANS CLUSTERING DALAM PEMETAAN KEMISKINAN KABUPATEN / KOTA DI TEPAT,” vol. 13, no. 2, 2025.
- [16] I. N. Abrar, A. Abdullah, and S. Sucipto, “Liver Disease Classification Using the Elbow Method to Determine Optimal K in the K-Nearest Neighbor (K-NN) Algorithm,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 2, pp. 218–228, 2023, doi: 10.32736/sisfokom.v12i2.1643.
- [17] Y. Pitaloka Anggriani, A. Arif, and F. Febriansyah, “Implementasi Algoritma K-Means Clustering Dalam Menentukan Blok Tanaman Sawit Produktif Pada Pt Arta Prigel,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 1820–1825, 2024, doi: 10.36040/jati.v8i2.9225.
- [18] 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.
- [19] A. Septianingsih, “Analisis K-Means Clustering Pada Pemetaan Provinsi Indonesia Berdasarkan Indikator Rumah Layak Huni,” *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 3, no. 1, pp. 224–241, 2022, doi: 10.46306/lb.v3i1.116.