

Implementasi K-means untuk Clustering pada Industri Non-Agro di Jawa Tengah

¹⁾ **Rafif Femas Erwanto**

Program Studi Sistem Informasi, Universitas Pembangunan Jaya, Indonesia
E-Mail: rafif.femaservanto@student.upj.ac.id

²⁾ **Safrizal Abdurahman**

Program Studi Sistem Informasi, Universitas Pembangunan Jaya, Indonesia
E-Mail: safrizal.abdurahman@upj.ac.id

ABSTRACT

This study aims to segment the non-agro industry in Central Java Province using the K-Means Clustering method. The approach used refers to the Knowledge Discovery in Database (KDD) stages, which include data selection, preprocessing, transformation, data mining, and interpretation/evaluation. Data were obtained from the Central Java Open Data portal, with variables used including district/city, industry type, business scale, marketing reach, and energy use. The preprocessing stage was carried out through data cleaning, integration of several regional files, and normalization using StandardScaler. The data mining process was carried out using the K-Means algorithm and the determination of the optimal number of clusters using the Elbow Method. The results of the study indicate the formation of several clusters that describe market segmentation patterns in the non-agro industry based on similarities in operational characteristics between regions in Central Java.

Keyword : K-means, Clustering, KDD, Market Segmentation, Central Java

PENDAHULUAN

Industri non-agro merupakan salah satu subsektor yang berperan penting dalam mendukung struktur perekonomian nasional. Berbeda dengan sektor agro yang bergantung pada hasil pertanian dan faktor musiman, industri non-agro cenderung lebih stabil dan lebih fleksibel terhadap dinamika pasar. Dalam menghadapi persaingan yang semakin ketat, pelaku industri non-agro perlu memahami karakteristik pasar dan kondisi industri di wilayahnya. Salah satu pendekatan yang dapat digunakan untuk memahami karakteristik tersebut adalah melalui analisis data untuk mengelompokkan wilayah berdasarkan potensi industrinya[1]. Melalui segmentasi pasar yang tepat, strategi pengembangan industri dapat disusun secara lebih efektif dan terarah. Data tersebut dapat menjadi salah satu indikator untuk mengukur tingkat kemajuan ekonomi daerah serta potensi daya saing wilayah dalam pengembangan basis industri modern yang berorientasi ekspansi.

Berdasarkan data industri non-agro Jawa Tengah, setiap kabupaten/kota memiliki karakteristik dan potensi industri yang berbeda. Ada daerah yang didominasi industri berskala kecil/rumahan, dan yang sudah berkembang dalam bentuk skala menengah, dan bahkan ada daerah dengan orientasi pemasaran sudah menembus luar provinsi hingga ekspor. Perbedaan karakteristik ini menandakan adanya segmentasi alami di lapangan yang sebenarnya dapat dipetakan secara ilmiah agar pemerintah

daerah maupun instansi terkait dapat menyusun kebijakan pengembangan yang lebih terarah.

Untuk mendukung proses segmentasi tersebut, dibutuhkan metode analisis data yang mampu mengelompokkan objek berdasarkan kemiripan karakteristik. Salah satu metode yang umum digunakan adalah algoritma K-Means Clustering. K-Means merupakan algoritma data mining yang berfungsi untuk mengelompokkan data ke dalam beberapa kluster berdasarkan tingkat kesamaan antar data. Pendekatan serupa juga dilakukan dalam penelitian sebelumnya yang menerapkan metode *K-Means* untuk mengelompokkan kabupaten/kota di Provinsi Jawa Tengah berdasarkan indikator Indeks Pembangunan Manusia (IPM) sebagai dasar analisis wilayah[2]. Selain itu, penelitian lain juga menunjukkan bahwa algoritma K-Means efektif dalam mengelompokkan wilayah berdasarkan karakteristik kompleks, seperti pengelompokan daerah rawan bencana di Jawa Tengah, yang memudahkan pengambilan keputusan dan alokasi sumber daya secara lebih terarah[3]. Dengan penerapan metode ini, data industri non-agro di Jawa Tengah dapat diklasifikasikan ke dalam beberapa segmen atau kelompok yang merepresentasikan karakteristik pasar yang berbeda-beda.

Pada penelitian ini akan digunakan variabel terkait kabupaten/kota, jenis industri, skala usaha, energi, dan pemasaran. Hasil cluster diharapkan mampu menghasilkan segmentasi pasar yang dapat menjadi dasar pengambilan keputusan bagi pihak pemerintah daerah, pelaku industri, maupun investor. Hasil dari proses clustering diharapkan mampu

memberikan gambaran yang lebih jelas mengenai kondisi industri non-agro di Jawa Tengah, serta membantu dalam merancang strategi pengembangan industri yang lebih tepat sasaran, sehingga dapat menjadi dasar dalam rekomendasi perencanaan, strategi penguatan kompetensi industri daerah, serta perbandingan kapasitas antar wilayah untuk keperluan evaluasi pembangunan sektor industri.

BAHAN DAN METODE

Penelitian ini menggunakan pendekatan kuantitatif deskriptif, dimana pemanfaatan algoritma K-Means difokuskan untuk melakukan segmentasi dan pemetaan pola karakteristik industri non-agro di Provinsi Jawa Tengah.

1) Sumber dan Jenis Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Portal Open Data Pemerintah Provinsi Jawa Tengah (<https://data.jatengprov.go.id/>) yang di simpan dalam format (Xls) Excel. Berikut detail dari dataset tersebut memuat informasi terkait :

Tabel 1. Deskripsi Data Keseluruhan
 [sumber : Ervanto & Abdurahman, 2025]

Variabel	Keterangan
Tahun	Representasi wilayah administratif usaha
Kabkota	Representasi wilayah administratif usaha
Bu	Badan usaha
Perusahaan	Nama perusahaan
Pemilik	Nama Pemilik Usaha
Alamatusaha	Lokasi usaha
Desa	Desa
Kecamatan	Kecamatan
Kbli	Kode usaha
Komoditi	Kategori/sub-sektor industri
Jenis	Kategori / sub-sektor industri

Jml_laki	Jumlah pekerja berjenis kelamin laki-laki
Jml_perempuan	Jumlah pekerja berjenis kelamin perempuan
Penolong	-
Pemasaran	Cakupan pemasaran usaha (lokal/regional/nasional)
Pemasaran2	-
Energi	Jenis energi yang digunakan untuk operasional
Skala	Skala usaha (kecil/menengah/besar)
Sumber_file	File berasal dari file Excel mana
Kode	-

2) Teknik pengolahan data

Pengolahan data pada penelitian ini menggunakan pendekatan Knowledge Discovery in Database (KDD) yang terdiri dari lima tahapan utama. Pendekatan KDD dipilih karena memberikan kerangka sistematis untuk pengolahan data sehingga dapat mengekstraksi informasi yang relevan dan berguna dari dataset besar[4]. Lima tahapan utama tersebut adalah:

a) Data Selection

Pada tahap ini, dilakukan proses seleksi variabel (*data selection*) dari dataset awal yang memuat sejumlah atribut terkait industri non-agro di Provinsi Jawa Tengah. Proses ini bertujuan untuk memastikan bahwa hanya variabel-variabel yang memiliki relevansi tinggi terhadap pembentukan pola karakteristik usaha yang digunakan dalam analisis. Dari keseluruhan variabel yang tersedia, dipilih lima variabel utama yang dianggap paling representatif dalam menggambarkan aktivitas dan kinerja sektor industri non-agro di Jawa Tengah. Dengan demikian, tahap seleksi data ini menjadi fondasi penting dalam menjamin relevansi dan validitas hasil segmentasi pasar yang dihasilkan pada tahap berikutnya.

b) Preprocessing Data

Tahap data preprocessing merupakan langkah krusial yang dilakukan untuk

memastikan bahwa data yang digunakan dalam penelitian memiliki kualitas yang baik, konsisten, dan siap untuk diolah pada tahap analisis berikutnya. Proses ini mencakup beberapa kegiatan utama seperti data integration dan data cleaning, yang secara bersama-sama bertujuan untuk meningkatkan keandalan data serta mengurangi potensi kesalahan dalam proses clustering.

Kegiatan data cleaning mencakup pemeriksaan dan penanganan nilai yang hilang (missing values), pengecekan format kolom, penghapusan atribut yang tidak relevan. Selanjutnya, dilakukan data integration dengan menggabungkan beberapa file data yang berasal dari sumber yang sama, yaitu portal Open Data Jateng, namun terpisah berdasarkan kabupaten atau kota di Provinsi Jawa Tengah. Proses penggabungan tersebut menghasilkan satu dataset terpadu yang merepresentasikan keseluruhan industri non-agro di wilayah Jawa Tengah, sehingga siap untuk digunakan pada tahap data transformation dan clustering menggunakan algoritma K-Means.

c) Data Transformation

Dalam tahap transformasi data dilakukan proses konversi dan penyesuaian format data agar dapat diolah secara optimal dalam implementasi kmeans. Data yang semula masih mengandung nilai kategorikal terlebih dahulu diubah menjadi bentuk numerik melalui metode Label Encoding, dengan tujuan agar setiap kategori pada variabel dapat direpresentasikan dalam bentuk angka tanpa mengubah makna kategorinya. Proses ini penting karena algoritma K-Means hanya dapat memproses data numerik dalam perhitungan jarak antar titik data.

Setelahnya adalah tahapan normalisasi data menggunakan metode StandardScaler agar seluruh variabel memiliki distribusi nilai yang sebanding. Normalisasi dilakukan untuk menghindari dominasi variabel tertentu yang memiliki skala lebih besar dibandingkan variabel lainnya, yang dapat menyebabkan bias dalam hasil pengelompokan. Data yang digunakan dalam analisis K-Means distandarisasi untuk menyeimbangkan skala variabel, sehingga setiap variabel memberikan kontribusi yang sebanding dalam pengelompokan cluster[5]. Tahap transformasi ini memastikan bahwa data berada dalam kondisi yang stabil dan terstandarisasi sehingga hasil clustering yang diperoleh dapat merepresentasikan

karakteristik industri non-agro di Jawa Tengah secara lebih akurat.

d) Data Mining

Tahap ini merupakan proses penerapan dari algoritma K-means, di mana data yang telah melalui tahap preprocessing lalu akan dikelompokkan berdasarkan tingkat kemiripan karakteristik antar objek. Algoritma K-Means digunakan karena dikenal luas sebagai metode clustering yang efektif dan sering digunakan untuk menganalisis sebuah data serta mengelompokkan objek berdasarkan kesamaan karakteristiknya[6]. Proses ini diawali dengan menentukan nilai K atau jumlah kluster yang akan dibentuk. Penentuan jumlah kluster dilakukan melalui metode Elbow (Elbow Method), serupa dengan pendekatan yang digunakan dalam penelitian sebelumnya untuk mengelompokkan kabupaten/kota berdasarkan kepadatan penduduk[7].

Setiap kluster yang terbentuk memiliki centroid, yaitu titik rata-rata dari semua data dalam kluster tersebut. Centroid ini merepresentasikan posisi pusat dari karakteristik tiap kluster, sehingga data yang memiliki kemiripan lebih dekat dengan centroid akan masuk ke dalam kluster tersebut[8]. Posisi centroid menggambarkan nilai rata-rata dari seluruh data yang termasuk dalam satu kelompok, sehingga dapat dijadikan representasi karakteristik umum dari cluster tersebut. Hasil akhir dari tahap ini berupa pembentukan sejumlah kluster yang merepresentasikan kelompok wilayah industri non-agro di Jawa Tengah dengan karakteristik serupa, yang selanjutnya digunakan untuk analisis segmentasi pasar dan evaluasi performa model.

e) Interpretation / Evaluation

Tahap akhir ini dilakukan dengan menginterpretasikan hasil clustering untuk memperoleh pemahaman mendalam mengenai pola segmentasi industri non-agro di Provinsi Jawa Tengah berdasarkan kesamaan karakteristik antar wilayah. Interpretasi hasil kluster bertujuan untuk mengidentifikasi perbedaan antar kelompok yang terbentuk, sehingga setiap kluster dapat dijelaskan dari sisi potensi, skala usaha, maupun tingkat kontribusi ekonominya. Analisis dilakukan dengan meninjau nilai rata-rata setiap variabel pada masing-masing kluster untuk menentukan ciri khas utama yang membedakan satu kluster dengan kluster lainnya.

HASIL DAN PEMBAHASAN

1. Data selection

Pada tahap ini peneliti menerapkan proses pemilihan data (selection) dengan mengambil dataset industri non-agro yang tersedia pada Portal Data Terbuka Pemerintah Provinsi Jawa Tengah. Dari keseluruhan atribut pada dataset sumber, peneliti hanya memilih atribut yang relevan dengan kebutuhan proses clustering.

Atribut yang diseleksi dan digunakan pada penelitian ini terdiri dari:

Tabel 2. Deskripsi Data yang Digunakan [sumber : Ervanto & Abdurahman, 2025]

Variabel	Keterangan	Tipe Data
Kabkota	Representasi wilayah administratif usaha	Object
Jenis	Kategori / sub-sektor industri	Object
Skala	Skala usaha (kecil/menengah/besar)	Object
Pemasaran	Cakupan pemasaran usaha (lokal/regional/nasional)	Object
Energi	Jenis energi yang digunakan untuk operasional	Object

Kelima atribut tersebut dipilih karena dianggap mampu mewakili karakteristik industri non-agro dan menjadi dasar pembentukan pola kelompok pada tahap data mining. Data di luar atribut tersebut tidak digunakan karena tidak memberikan kontribusi terhadap pembentukan cluster.

2. Data Pre-processing

Tahapan pre-processing yang telah dijelaskan secara konseptual pada bagian bahan dan metode akhirnya diimplementasikan pada dataset industri non-agro menggunakan Jupyter Notebook. Proses ini bertujuan untuk memastikan kualitas data agar siap digunakan pada tahap analisis lebih lanjut dengan algoritma K-Means clustering. Secara umum, pre-processing dilakukan melalui dua langkah utama, yaitu data integration dan data cleansing. Kedua proses ini berperan penting dalam menghasilkan dataset yang bersih, terstruktur, dan bebas dari ketidakkonsistenan yang dapat memengaruhi hasil analisis.

2.1 Data integration

Pada tahapan ini dilakukan untuk menggabungkan beberapa sumber data menjadi satu kesatuan dataset yang utuh. Data awal diperoleh dari portal

Open Data Jateng, di mana informasi industri non-agro disajikan dalam bentuk terpisah berdasarkan kabupaten/kota. Oleh karena itu, langkah pertama yang dilakukan adalah penggabungan (integration) beberapa file Excel menjadi satu dataset terpadu agar seluruh wilayah di Provinsi Jawa Tengah dapat dianalisis secara menyeluruh.

```
# Menggabungkan beberapa file Excel
menjadi satu dataset
df_list = []
for filename in all_files:
    data = pd.read_excel(filename)
    data['sumber_file'] = filename.split("\")[1]
    df_list.append(data)
# Menggabungkan seluruh data menjadi
satu DataFrame
df = pd.concat(df_list,
ignore_index=True)
```

```
=== Contoh Data Gabungan ===
Unnamed: 0 tahun kabkota bu perusahaan \
0 1,0 2019 KABUPATEN CILACAP Perorangan PEMBUATAN BATA MERAH
1 2,0 2019 KABUPATEN CILACAP perorangan PEMBUATAN BATA MERAH
2 3,0 2019 KABUPATEN CILACAP Perorangan LAS/PASANG BAJA RINGAN
3 4,0 2019 KABUPATEN CILACAP Perorangan LAS/PASANG BAJA RINGAN
4 5,0 2019 KABUPATEN CILACAP perorangan PEMBUATAN BATA MERAH

pemilik alamatusaha desa kecamatan kbli ... jenis \
0 Muslih NaN Ketanggung Sampang 23921.0 ... Bata merah
1 Sarkum NaN Ketanggung Sampang 23921.0 ... Bata merah
2 Suparno NaN Ketanggung Sampang 25111.0 ... Baja ringan
3 Sarino NaN Ketanggung Sampang 25111.0 ... Baja ringan
4 Achmad Matori NaN Ketanggung Sampang 23921.0 ... Bata merah

jml_laki jml_perempuan penolong pemasaran pemasaran2 energi skala \
0 0 0 0 Bata merah Lokal Tidak Ada Listrik Kecil
1 0 0 0 Bata merah Lokal Tidak Ada Listrik Kecil
2 0 0 0 Baja ringan Lokal Tidak Ada Listrik Kecil
3 0 0 0 Baja ringan Lokal Tidak Ada Listrik Kecil
4 0 0 0 Bata merah Lokal Tidak Ada Listrik Kecil

sumber_file kode
0 CILACAP.xlsx NaN
1 CILACAP.xlsx NaN
2 CILACAP.xlsx NaN
3 CILACAP.xlsx NaN
4 CILACAP.xlsx NaN
```

Gambar 1. Gabungan Data Set [sumber: Ervanto & Abdurahman, 2025]

Proses integrasi dilakukan dengan menggabungkan seluruh file Excel tersebut menjadi satu dataset komprehensif menggunakan Python (pandas) di dalam Jupyter Notebook. Proses ini memastikan bahwa seluruh entitas industri dari 35 kabupaten/kota di Jawa Tengah dapat dianalisis secara bersamaan tanpa kehilangan informasi.

2.2 Data cleansing

Pada tahap ini dilakukan pemeriksaan terhadap data untuk memastikan tidak terdapat *missing value*, duplikasi, maupun ketidakkonsistenan format. Proses pembersihan mencakup penyesuaian nama kabupaten/kota, perbaikan penulisan jenis industri, serta penghapusan kolom yang tidak relevan terhadap analisis. Data dengan nilai kosong pada variabel penting tetap dipertahankan apabila tidak

memengaruhi hasil secara signifikan, agar struktur data tetap merepresentasikan kondisi sebenarnya. Hasil akhir dari proses ini menghasilkan dataset yang siap digunakan untuk tahap transformasi selanjutnya.

```
df_clean = df.copy()
for c in ['kabkota', 'jenis', 'skala',
'pemasaran', 'energi']:
    df_clean[c] = (
        df_clean[c].astype(str)
        .str.strip()
        .str.lower()
        .replace({'-': np.nan, "': np.nan, 'nan':
np.nan})
    )
# Menampilkan hasil data setelah
pembersihan
df_clean.head()
```

	kabkota	jenis	skala	pemasaran	energi
0	kabupaten cilacap	bata merah	kecil	lokal	listrik
1	kabupaten cilacap	bata merah	kecil	lokal	listrik
2	kabupaten cilacap	baja ringan	kecil	lokal	listrik
3	kabupaten cilacap	baja ringan	kecil	lokal	listrik
4	kabupaten cilacap	bata merah	kecil	lokal	listrik
5	kabupaten cilacap	pakaian	kecil	lokal	listrik
6	kabupaten cilacap	pakaian	kecil	lokal	listrik
7	kabupaten cilacap	pakaian	kecil	lokal	listrik
8	kabupaten cilacap	semen	kecil	lokal	listrik
9	kabupaten cilacap	bata	kecil	lokal	listrik

Gambar 2. Data Setelah Cleansing
 [sumber : Ervanto & Abdurahman, 2025]

3. Data transformation

Pada tahap ini, proses transformasi data dilakukan untuk mengubah variabel kategorikal menjadi numerik agar dapat diolah oleh algoritma K-Means. Proses pengubahan dilakukan menggunakan teknik *Label Encoding* terhadap seluruh variabel kategorikal yang digunakan, yaitu kabupaten/kota, jenis industri, skala, pemasaran, dan energi. Setelah proses *encoding* selesai, data kemudian dinormalisasi menggunakan *StandardScaler* agar setiap fitur memiliki skala yang sebanding dan tidak menimbulkan bias dalam perhitungan jarak antar data pada proses *clustering*.

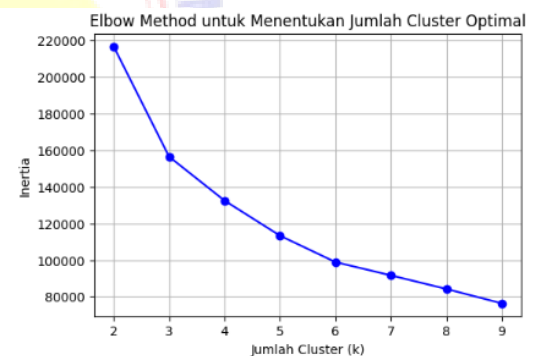
	kabkota	jenis	skala	pemasaran	energi
0	5	449	5	381	101
1	5	449	5	381	101
2	5	268	5	381	101
3	5	268	5	381	101
4	5	449	5	381	101

4. Data Mining

Pada tahap ini dilakukan penerapan algoritma *K-Means Clustering* untuk membentuk kelompok industri non-agro berdasarkan kemiripan atribut yang dimiliki. Proses ini melibatkan pemilihan nilai K melalui pengujian beberapa alternatif nilai, hingga diperoleh K = 4 sebagai hasil yang paling representatif. Model K-Means kemudian di *training* menggunakan data hasil data transformasi, sehingga setiap

Gambar 3. Data Setelah Encode
 [sumber : Ervanto & Abdurahman, 2025]

data industri tergabung ke dalam salah satu dari empat klaster yang terbentuk. Sebelum menentukan jumlah klaster, dilakukan pengujian nilai K menggunakan *Elbow Method* untuk menemukan jumlah klaster optimal. Berdasarkan hasil perhitungan, nilai inerti menurun tajam hingga K=4 dan mulai melandai setelahnya. Hal ini menunjukkan bahwa K=4 merupakan titik optimal (*elbow point*) yang paling representatif untuk digunakan pada proses klasterisasi. Visualisasi hasil pengujian ditunjukkan pada Gambar 4, di mana tekukan grafik terlihat jelas pada nilai K=4.



Gambar 4. Elbow Method
 [sumber : Ervanto & Abdurahman, 2025]

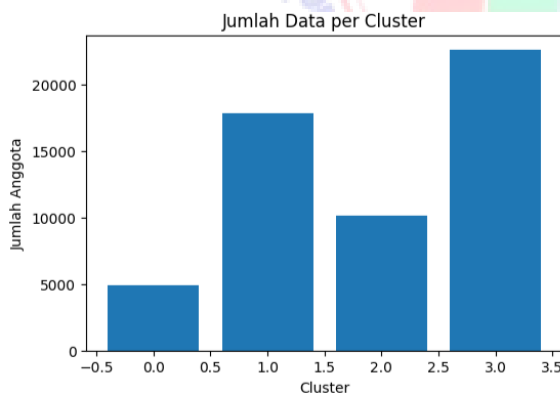
Berdasarkan hasil pemodelan K-Means dengan jumlah klaster optimal K = 4, diperoleh empat kelompok yang memiliki jumlah anggota pada setiap klaster, yaitu Cluster 0 sebanyak 4.928 data, Cluster 1 sebanyak 17.820 data, Cluster 2 sebanyak 10.090 data, dan Cluster 3 sebanyak 22.553 data. Pada kalster tersebut menunjukkan pola industri non-agro dengan karakteristik yang berbeda. Pembagian ini tidak hanya memisahkan berdasarkan kabupaten/kota, tetapi juga berdasarkan kecenderungan jenis usaha yang paling dominan pada masing-masing klaster. Secara ringkas, karakteristik setiap cluster dapat dijelaskan sebagai berikut:

- Cluster 0, didominasi oleh industri konveksi kecil, anyaman bambu,

penjahitan, dan beberapa jasa berbasis rumahan. Pola ini menggambarkan kelompok industri dengan karakter kerajinan lokal dan produksi berskala kecil.

- Cluster 1, memiliki komposisi industri yang lebih bervariasi seperti batu bata, batik, jasa penjahitan, hingga bengkel. Kelompok ini merepresentasikan home industry campuran yang telah memiliki orientasi komersial lebih luas dibanding cluster 0.
- Cluster 2, terdiri terutama dari industri material bangunan tradisional seperti bata merah, genteng, dan tenun. Klaster ini merepresentasikan industri produksi bahan bangunan maupun tekstil tradisional yang masih banyak ditemukan di wilayah pedesaan.
- Cluster 3, memiliki dominasi yang kuat pada industri konveksi dan pakaian jadi dalam jumlah besar, yang menunjukkan bahwa cluster ini cenderung merepresentasikan industri fashion/tekstil dengan skala produksi yang lebih tinggi atau semi-mass production.

Distribusi jumlah anggota pada tiap klaster divisualisasikan menggunakan diagram batang sebagaimana ditampilkan pada gambar.



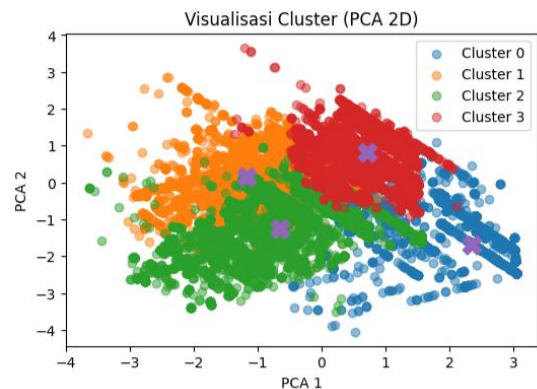
Gambar 5. Bar Chart Jumlah Anggota per Cluster
[sumber : Ervanto & Abdurahman, 2025]

5. Interpretation / Evaluation

Hasil clustering yang telah diperoleh menggunakan algoritma K-Means, dengan tujuan untuk memahami pola segmentasi industri non-agro di Provinsi Jawa Tengah berdasarkan kesamaan karakteristik antar entitas data. Setelah model menghasilkan empat klaster utama, dilakukan analisis terhadap komposisi variabel pada masing-masing klaster untuk mengetahui

karakteristik dominan yang membedakan setiap kelompok.

Secara umum, hasil interpretasi menunjukkan bahwa setiap klaster merepresentasikan segmen industri dengan ciri khas tertentu. Misalnya, Cluster 0 didominasi oleh industri berskala kecil dengan jangkauan pemasaran lokal dan penggunaan energi yang relatif rendah. Sementara itu, Cluster 1 menggambarkan kelompok industri dengan skala menengah dan jangkauan pemasaran regional. Cluster 2 cenderung berisi industri yang memiliki skala lebih besar dengan diversifikasi jenis usaha yang tinggi serta penggunaan energi yang lebih intensif. Adapun Cluster 3 terdiri atas industri yang memiliki persebaran luas antar wilayah dengan variasi kategori jenis usaha yang lebih kompleks.



Gambar 6. Visualisasi Scatter Plot
[sumber : Ervanto & Abdurahman, 2025]

Hasil visualisasi dalam scatter plot di mana setiap klaster tampak membentuk area yang relatif terpisah. Hal ini menunjukkan bahwa proses pengelompokan berhasil memisahkan data berdasarkan tingkat kemiripan atribut secara efektif. Evaluasi terhadap hasil clustering juga dilakukan menggunakan Elbow Method, yang menunjukkan nilai K=4 optimal dalam penelitian, menandakan bahwa jarak antar klaster cukup jelas dan setiap klaster tergolong baik. Pada visualisasi tersebut, setiap titik kecil merepresentasikan satu entitas data industri non-agro, sedangkan tanda "X" berukuran besar di tengah masing-masing kelompok menunjukkan posisi centroid dari setiap klaster hasil algoritma K-Means. Posisi centroid yang relatif tersebar menunjukkan bahwa keempat klaster memiliki pola karakteristik yang cukup berbeda, sehingga proses clustering berhasil memisahkan data dengan baik.

KESIMPULAN

Penelitian ini bertujuan untuk mengimplementasikan algoritma K-Means Clustering dalam melakukan segmentasi industri non-agro di Provinsi Jawa Tengah, guna mengidentifikasi pola kesamaan karakteristik antar kelompok industri. Melalui pendekatan Knowledge Discovery in Database (KDD) yang meliputi tahapan data selection, preprocessing, transformation, data mining, dan interpretation/evaluation, diperoleh hasil analisis yang mampu menggambarkan segmentasi industri secara komprehensif.

Hasil penerapan algoritma K-Means menunjukkan bahwa nilai K optimal adalah 4, dengan pembentukan empat kluster utama yang merepresentasikan perbedaan karakteristik industri di wilayah Jawa Tengah. Masing-masing kluster memiliki ciri khas tersendiri, mulai dari industri berskala kecil berbasis kerajinan dan jasa rumahan (Cluster 0), industri campuran berorientasi komersial (Cluster 1), industri material dan tekstil tradisional di wilayah pedesaan (Cluster 2), hingga industri fashion atau konveksi berskala lebih besar (Cluster 3). Nilai Silhouette Score yang positif mengindikasikan bahwa hasil pengelompokan memiliki tingkat validitas dan pemisahan antar-kluster yang baik.

Secara keseluruhan, penelitian ini menunjukkan bahwa penerapan metode K-Means Clustering efektif dalam memetakan segmentasi industri non-agro di Jawa Tengah. Temuan ini dapat menjadi dasar bagi pemerintah daerah dan pemangku kepentingan industri dalam merancang strategi pengembangan sektor industri yang lebih tepat sasaran, baik dalam hal pemberdayaan usaha kecil, peningkatan kapasitas produksi, maupun pemerataan pertumbuhan ekonomi antar wilayah.

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan ke hadirat Allah SWT atas rahmat dan karunia-Nya sehingga tugas akhir berjudul "*Implementasi K-Means Clustering pada Industri Non-Agro di Jawa Tengah*" dapat diselesaikan dengan baik. Ucapan terima kasih penulis sampaikan kepada dosen pembimbing, seluruh dosen Program Studi Sistem Informasi, orang tua, serta rekan-rekan yang telah memberikan dukungan, bimbingan, dan motivasi selama proses penyusunan tugas akhir ini. Semoga karya ini dapat bermanfaat bagi pembaca dan menjadi referensi bagi penelitian selanjutnya.

DAFTAR PUSTAKA

[1] J. Jtik, J. Teknologi, R. Aryasatya, and V. Lusiana, "Penentuan Klustering Indeks Pembangunan Manusia Provinsi Jawa Tengah dengan Metode K-Means Berbasis Web," vol. 8, no. 1, 2024.

- [2] M. R. Fadillah, M. R. Ardiansyah, R. Junaidi, and F. Fauzi, "Penerapan Clustering dalam Mengelompokkan Kabupaten / Kota Berdasarkan Indikator IPM di Jawa Tengah," vol. 3, no. 3, pp. 674–685, 2025.
- [3] R. S. Hidayat, M. R. Muttaqin, and D. Irmayanti, "PENGELOMPOKAN DAERAH RAWAN BENCANA DI JAWA TENGAH MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING," vol. 8, no. 5, pp. 10035–10042, 2024.
- [4] I. A. Nikmatun, U. Diponegoro, I. Waspada, and U. Diponegoro, "IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR," vol. 10, no. 2, pp. 421–432, 2019.
- [5] P. S. Statistika and U. I. Indonesia, "Implementasi K-Means Cluster Analysis untuk Mengelompokkan Kabupaten / Kota Berdasarkan Data Kemiskinan di Provinsi Jawa Tengah Tahun 2022," vol. 3, no. 2, pp. 317–329, 2023.
- [6] A. F. Nashrullah, R. D. Mahardhika, N. R. Rusdiyanto, and S. Shella, "Klasterisasi Produktifitas Daerah di Jawa Tengah Berdasarkan Ketenagakerjaan Menggunakan K-Means dan Average Linkage," vol. 7, no. 2, pp. 138–149, 2025.
- [7] Y. Asyfani, I. M. Nur, I. F. Amri, and N. Yunanita, "Pengelompokan Kabupaten / Kota di Jawa Tengah Berdasarkan Kepadatan Penduduk Menggunakan Metode Hierarchical Clustering," vol. 2, no. 1, pp. 1–8, 2024.
- [8] M. U. Yanuar, "Metode K-Means Clustering untuk Pengelompokan Kabupaten / Kota dalam Upaya Pengendalian Tingkat Inflasi di Pulau Jawa dan Sumatera K-Means Clustering Method for District / City Grouping in Effort to Control Inflation Rates in Java," vol. 1, pp. 29–42, 2022.