

KLASTERISASI KABUPATEN DAN KOTA DI PROVINSI SUMATERA UTARA BERDASARKAN INDUSTRI KECIL DAN MIKRO MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING

Agnes Irene Silitonga¹, Ali Akbar Lubis², Jufri Darma³, Ferry Indra H Sinaga⁴, Yoakim Simamora⁵

¹ Program Studi Bisnis Digital, Universitas Negeri Medan

² Program Studi Pendidikan Teknologi Informatika dan Komputer, Universitas Negeri Medan

³ Program Studi Pendidikan Akuntansi, Universitas Negeri Medan

⁴ Program Studi Pendidikan Teknik Mesin, Universitas Negeri Medan

⁵ Program Studi Pendidikan Teknik Elektro, Universitas Negeri Medan

agnesirenesilitonga@unimed.ac.id

ABSTRAK

Industri kecil dan mikro memiliki peran strategis dalam perekonomian daerah, terutama dalam meningkatkan kesejahteraan masyarakat dan mendorong pertumbuhan ekonomi. Penelitian ini bertujuan untuk mengklasterisasi kabupaten dan kota di Provinsi Sumatera Utara menggunakan algoritma *K-Means Clustering*. Data yang digunakan meliputi jumlah unit usaha, tenaga kerja, dan pinjaman modal bank di setiap wilayah. Metode *K-Means Clustering* dipilih karena kemampuannya dalam mengklasterisasi data berdasarkan kemiripan karakteristik sehingga dapat memberikan gambaran klasifikasi wilayah dengan potensi industri kecil dan mikro yang serupa. Hasil penelitian menunjukkan bahwa kabupaten dan kota di Provinsi Sumatera Utara dapat diklasterisasikan menjadi tiga klaster yaitu klaster rendah, sedang, dan tinggi. Hasil klasterisasi ini diharapkan dapat menjadi dasar bagi pemerintah daerah dalam merancang kebijakan yang lebih efektif untuk pengembangan industri kecil dan mikro di masing-masing wilayah, seperti alokasi bantuan modal, pelatihan, dan penguatan rantai pasok industri.

Kata kunci: Algoritma *K-Means Clustering*, Industri Kecil dan Mikro, Klasterisasi, *Machine Learning*

1. PENDAHULUAN

Transformasi ekonomi Indonesia yang semakin dinamis menuntut pemahaman yang mendalam terhadap produktivitas industri kecil dan mikro terutama di Provinsi Sumatera Utara yang memiliki potensi ekonomi signifikan. Hal ini dikarenakan industri kecil dan mikro berperan penting terhadap perekonomian Indonesia. Sektor ini tidak hanya menjadi penyumbang besar terhadap pertumbuhan ekonomi, tetapi juga dapat menyerap tenaga kerja dan menurunkan angka kemiskinan di tingkat lokal. Lebih dari 64 juta unit usaha mikro, kecil, dan menengah (UMKM) menyumbang sekitar 60% dari Produk Domestik Bruto (PDB) nasional dan menyerap lebih dari 97% tenaga kerja [1]. Di Provinsi Sumatera Utara, industri kecil dan mikro memainkan peran signifikan dalam mendukung ketahanan ekonomi daerah, terutama di tengah situasi global yang penuh ketidakpastian.

Meskipun memiliki potensi besar, industri kecil dan mikro sering kali belum dioptimalkan secara maksimal. Industri kecil dan mikro menghadapi sejumlah tantangan yang dapat menghambat pertumbuhannya [2]. Tantangan tersebut meliputi akses terhadap sumber daya keuangan, keterbatasan dalam pemasaran produk, serta rendahnya adopsi teknologi modern. Di samping itu, kurangnya informasi tentang segmentasi atau karakteristik spesifik industri kecil dan mikro di setiap wilayah menjadi salah satu hambatan dalam merumuskan kebijakan yang efektif.

Sumatera Utara adalah salah satu provinsi dengan tingkat keanekaragaman ekonomi yang tinggi.

Sektor industri kecil dan mikro di provinsi ini mencakup berbagai bidang, seperti makanan dan minuman, kerajinan tangan, tekstil, serta jasa kreatif [3]. Secara geografis, Sumatera Utara memiliki karakteristik yang heterogen, mulai dari kawasan perkotaan seperti Medan hingga daerah pedesaan di Tapanuli atau Nias. Keberagaman ini menciptakan kompleksitas tambahan dalam pengelolaan industri kecil dan mikro.

Wilayah-wilayah dengan konsentrasi industri kecil dan mikro tinggi di Sumatera Utara sering kali menunjukkan pola ketimpangan dalam hal dukungan infrastruktur, akses pasar, dan kualitas sumber daya manusia [4]. Sebagai contoh, industri kecil dan mikro di kawasan urban cenderung memiliki akses yang lebih baik terhadap teknologi dan pembiayaan dibandingkan industri kecil dan mikro di daerah rural. Fenomena ini membutuhkan pendekatan berbasis data untuk mengidentifikasi kebutuhan dan peluang setiap kelompok industri kecil dan mikro.

Kemajuan teknologi informasi dan analisis data membuka peluang baru dalam pengelolaan industri kecil dan mikro, salah satu kemajuan ini yaitu penggunaan *machine learning*. Salah satu penggunaan *machine learning* yang telah banyak digunakan adalah pengambilan keputusan berbasis data. Dalam konteks klasterisasi industri kecil dan mikro di Provinsi Sumatera Utara, *machine learning* digunakan untuk menemukan pola atau struktur tersembunyi dalam data industri menggunakan pendekatan *unsupervised learning*, yaitu metode pembelajaran tanpa label data.

Salah satu algoritma *machine learning* dalam bidang klasterisasi adalah algoritma *K-Means*

Clustering, yang mana algoritma ini mampu mengelompokkan data ke dalam kelompok – kelompok homogen berdasarkan karakteristik tertentu. *K-Means Clustering* dapat digunakan untuk mengelompokkan data industri kecil dan mikro berdasarkan berbagai faktor, seperti lokasi geografis, jenis produk, dan skala usaha (referensi). Studi serupa telah dilakukan di berbagai negara untuk mendukung pengembangan industri kecil dan mikro. Sebagai contoh, penelitian yang dilakukan di India menggunakan *K-Means Clustering* untuk mengelompokkan UMKM berdasarkan sektor industri dan skala usaha [5]. Hasil penelitian ini membantu pemerintah India dalam merancang kebijakan fiskal yang spesifik untuk setiap kelompok UMKM. Di Indonesia, penelitian oleh [4] menunjukkan bahwa klasterisasi dapat meningkatkan efisiensi program bantuan industri kecil dan mikro hingga 40%.

Dengan algoritma ini, kabupaten / kota di Provinsi Sumatera Utara dapat diklasterisasikan berdasarkan produktivitas industri kecil dan mikro ke dalam tiga klaster yaitu produktivitas tinggi, sedang, dan rendah. Melalui penerapan *K-Means Clustering* di Provinsi Sumatera Utara diharapkan diperoleh informasi kabupaten / kota yang memiliki produktivitas industri kecil dan mikro yang rendah. Sebagai contoh, kabupaten / kota yang memiliki jumlah tenaga kerja yang besar cenderung memiliki jumlah industri kecil dan mikro yang besar juga.

Selain itu, melalui penerapan *K-Means Clustering*, dapat diketahui kabupaten / kota mana saja yang memerlukan akses teknologi berdasarkan klaster produktivitas industri kecil dan menengah. Akses terhadap teknologi sering menjadi faktor pembeda antara industri kecil dan mikro yang berkembang pesat dan yang stagnan [2]. Dengan mengklasterisasikan industri kecil dan mikro, kebijakan pengembangan dapat difokuskan pada peningkatan teknologi di klaster tertentu. Klasterisasi juga dapat mengungkapkan perbedaan kebutuhan antara kelompok industri kecil dan mikro. Sebagai contoh, industri kecil dan mikro di klaster dengan skala kecil mungkin memerlukan pelatihan dasar manajemen keuangan, sedangkan klaster yang lebih maju dapat difokuskan pada pemasaran digital.

Di Sumatera Utara, keberagaman jenis usaha dan kondisi geografis menambah kompleksitas dalam pengelolaan industri kecil dan mikro di tiap kabupaten / kota. Oleh karena itu, penelitian ini bertujuan untuk mengklasterisasikan kabupaten / kota di Provinsi Sumatera Utara berdasarkan jumlah industri kecil dan mikro menggunakan algoritma *K-Means Clustering* sehingga diharapkan dapat memberikan kontribusi yang signifikan dengan mengadaptasi algoritma serupa. Dengan menggunakan data lokal dan variabel yang relevan, hasil klasterisasi dapat membantu dalam merumuskan kebijakan yang lebih terfokus dan berdampak jangka panjang bagi pengembangan industri kecil dan mikro.

2. TINJAUAN PUSTAKA

2.1. Machine Learning

Machine learning atau pembelajaran mesin merupakan cabang kecerdasan buatan yang mempelajari dan membangun model yang dapat belajar dari data atau lingkungan untuk membuat prediksi tentang data yang tidak diketahui atau membuat keputusan berdasarkan informasi lingkungan [6]. *Machine learning* didefinisikan sebagai ilmu komputer yang memberikan kemampuan pada sistem komputer untuk belajar secara otomatis dari pengalaman sebelumnya (data historis), tanpa harus diprogram secara eksplisit [7]. Sistem ini menggunakan algoritma untuk menemukan pola atau tren dalam data sehingga mampu membuat prediksi atau keputusan. *Machine learning* telah diterapkan di berbagai bidang seperti teknologi informasi, kesehatan, bisnis, dan otomotif.

2.2. K-Means Clustering

Dalam dunia *machine learning*, terdapat berbagai teknik yang digunakan untuk mengekstraksi pola dari data. Salah satu teknik yang penting dalam kelompok pembelajaran tidak terawasi (*unsupervised learning*) adalah algoritma *K-Means Clustering* yang merupakan algoritma pembelajaran tidak terawasi yang bertujuan untuk mengelompokkan data menjadi beberapa klaster berdasarkan kemiripan antar data. Algoritma ini mengelompokkan data dengan cara meminimalkan jarak antara titik data dengan titik pusat klaster (*centroid*) [8]. Hasil dari proses ini adalah pembagian dataset ke dalam beberapa kelompok yang memiliki pola atau karakteristik serupa.

Dalam implementasi di dunia nyata, *K-Means Clustering* digunakan secara luas karena kecepatan dan kemampuannya dalam menangani dataset besar. *K-Means Clustering* telah diaplikasikan dalam bidang seperti gizi / *stunting* (Silitonga et al., 2024) [9], *base transceiver station* (Silitonga et al., 2025) [10], penyebaran UMKM (Azzam et al., 2024) [11], dan lainnya.

Implementasi *K-Means Clustering* sering menggunakan bahasa pemrograman *Python* dengan bantuan pustaka seperti *scikit-learn*.

```
from sklearn.cluster import KMeans
import numpy as np

X = np.array([[1,2], [1,4], [1,0],
              [10,2], [10,4], [10,0]])
kmeans = KMeans(n_clusters=2)
kmeans.fit(X)
print(kmeans.cluster_centers_)
```

Machine learning dan *K-Means Clustering* saling berkaitan erat, khususnya dalam kategori *unsupervised learning*. *Machine learning* menyediakan kerangka kerja untuk algoritma *clustering* ini, yaitu kemampuan belajar dari data tanpa adanya label khusus. Sebagai metode *unsupervised*, *K-*

Means Clustering membantu dalam eksplorasi dan identifikasi pola tersembunyi dalam dataset yang kompleks [7].

3. METODE PENELITIAN

Penelitian ini menggunakan algoritma *K-Means Clustering* karena kemampuannya mengklusterisasikan data ke dalam beberapa kluster homogen berdasarkan kesamaan karakteristik (referensi). Berikut langkah – langkah algoritma *K-Means Clustering*:

- Melakukan persiapan data
 - Menyatukan semua variabel (x_1 , x_2 , x_3) dalam satu basis data yang terstruktur.
 - Membersihkan data dari nilai yang tidak valid.
 - Melakukan normalisasi/standarisasi (jika diperlukan) agar semua variabel memiliki skala yang sebanding.
- Menentukan jumlah kluster (k)
 - Dalam penelitian ini, jumlah kluster ditetapkan sebanyak 3 (rendah, sedang, dan tinggi).
 - Penetapan $k=3$ juga dapat diperkuat dengan uji evaluasi kluster metode Elbow atau Silhouette.
- Inisialisasi pusat kluster (centroid)
 - Memilih 3 titik awal (centroid) secara acak dari kumpulan data.
- Pengelompokkan awal
 - Menghitung jarak setiap data (kabupaten/kota) ke masing-masing centroid menggunakan jarak Euclidean
 - Menempatkan data ke kluster dengan jarak terpendek.
- Perhitungan ulang centroid
 - Setelah seluruh data dikelompokkan, centroid baru dihitung sebagai nilai rata-rata posisi data dalam kluster.
- Iterasi dan konvergensi
 - Proses mengelompokkan data dan menghitung ulang centroid diulang hingga tidak ada perpindahan data antar-kluster atau hingga jumlah iterasi maksimum tercapai.
 - Hasil akhir adalah kluster yang stabil (konvergen).
- Interpretasi kluster
 - Menganalisis karakteristik setiap kluster, misalnya kluster dengan nilai x_1 , x_2 , dan x_3 tinggi digolongkan ke dalam kluster tinggi, kluster dengan nilai sedang digolongkan sedang, dan kluster dengan nilai rendah seterusnya.

Adapun objek dan ruang lingkup penelitian adalah kabupaten / kota di Provinsi Sumatera Utara. Jumlah industri kecil dan mikro, jumlah tenaga kerja yang terserap di sektor industri kecil dan mikro, dan jumlah pinjaman modal dari bank atau lembaga keuangan lainnya merupakan variabel yang digunakan dalam penelitian ini dengan data yang bersumber dari

Badan Pusat Statistik Provinsi Sumatera Utara. Teknik analisis data yang dilakukan dalam penelitian ini, diantaranya:

- Analisis deskriptif
Menyajikan profil awal setiap kabupaten dan kota, seperti jumlah industri kecil dan mikro, jumlah tenaga kerja, dan pinjaman modal secara rata-rata.
- Analisis kluster
Menggunakan algoritma *K-Means Clustering*, kemudian memvalidasi hasil dengan mengecek jarak rata-rata intra-kluster dan inter-kluster, serta dapat menggunakan *Silhouette Score* untuk melihat seberapa baik pemisahan kluster.
- Penarikan kesimpulan
Menjelaskan implikasi kebijakan berdasarkan kluster, seperti kabupaten dan kota yang masuk ke dalam kluster tinggi mendapat dukungan kebijakan berbeda dengan kluster rendah.

Berikut proses klusterisasi kabupaten dan kota di Provinsi Sumatera Utara berdasarkan industri kecil dan mikro dengan *K-Means Clustering*:

- Sebelum dilakukan klusterisasi, data dinormalisasi menjadi skala 0–1 (jika diperlukan). Hal ini memastikan setiap variabel (x_1 , x_2 , x_3) memiliki bobot yang seimbang dalam penghitungan jarak. Variabel yang digunakan dalam penelitian ini yaitu banyaknya usaha kecil dan mikro di tiap kabupaten / kota, jumlah tenaga kerja, dan jumlah usaha yang mendapatkan pinjaman modal bank.
- Penentuan jumlah kluster sebanyak 3 kluster ($k=3$). Alasan pemilihan tiga kluster adalah untuk memudahkan interpretasi: (1) kluster rendah, (2) kluster sedang, dan (3) kluster tinggi. Secara empiris, metode Elbow juga dapat digunakan untuk melihat titik tekukan pada grafik SSE (*Sum of Squared Errors*) yang mengindikasikan penambahan kluster tidak lagi memberikan pengurangan kesalahan signifikan.
- Iterasi proses *K-Means Clustering*
 - Penentuan tiga *centroid* awal secara acak dari data.
 - Perhitungan jarak setiap kabupaten/kota ke *centroid*, kemudian dikelompokkan ke kluster dengan jarak terdekat.
 - Perhitungan ulang *centroid* berdasarkan rata – rata dari setiap kluster.

Proses berhenti ketika tidak ada perubahan signifikan pada komposisi anggota kluster atau *centroid* sudah stabil.

4. HASIL DAN PEMBAHASAN

Sebagai langkah awal, data yang digunakan dalam penelitian ini merupakan data 33 kabupaten / kota di Provinsi Sumatera Utara seperti pada Tabel 1.

Tabel 1. Data Usaha Kecil dan Mikro menurut Kabupaten / Kota di Provinsi Sumatera Utara

No.	Kabupaten / Kota	Banyaknya Usaha	Tenaga Kerja	Pinjaman Modal Bank
1	Nias	955	1741	0
2	Mandailing Natal	3342	6435	376

3	Tapanuli Selatan	4635	7650	282
4	Tapanuli Tengah	5250	8498	1434
5	Tapanuli Utara	6681	7974	132
6	Toba	3258	6055	560
7	Labuhan Batu	4614	9748	1363
8	Asahan	8258	23198	1417
9	Simalungun	9051	19415	855
10	Dairi	1859	2872	84
11	Karo	1875	3172	463
12	Deli Serdang	8556	23282	1290
13	Langkat	10001	23615	1754
14	Nias Selatan	1826	3450	25
15	Humbang Hasundutan	2066	2741	202
16	Pakpak Bharat	266	360	28
17	Samosir	6151	7951	1019
18	Serdang Bedagai	6002	14699	1330
19	Batu Bara	4706	11276	1197
20	Padang Lawas Utara	1169	2414	14
21	Padang Lawas	4017	8685	361
22	Labuhan Batu Selatan	908	1794	135
23	Labuhan Batu Utara	2034	3585	158
24	Nias Utara	586	1253	0
25	Nias Barat	336	389	0
26	Sibolga	650	1298	159
27	Tanjung Balai	2592	4064	74
28	Pematang Siantar	3383	5582	290
29	Tebing Tinggi	2718	6611	408
30	Medan	12300	30946	1418
31	Binjai	2785	6559	161
32	Padangsidempuan	1802	3932	337
33	Gunungsitoli	2275	3915	108
Total		126907	265159	17434

Langkah pertama yang dilakukan adalah penentuan jumlah kluster sebanyak 3 kluster yaitu rendah (K1), sedang (K2), tinggi (K3). Penentuan kluster ini dilakukan berdasarkan produktivitas industri kecil dan mikro pada kabupaten / kota di Provinsi Sumatera Utara. Selanjutnya ditentukan titik pusat tiap kluster, yaitu Kabupaten Pakpak Bharat sebagai titik pusat kluster rendah, Kabupaten Toba sebagai titik pusat kluster sedang, dan Kota Medan sebagai titik pusat kluster tinggi. Setelah menentukan titik pusat tiap kluster, jarak setiap data ke titik pusat tiap kluster dihitung dengan menggunakan rumus *Euclidian Distance* seperti pada Persamaan 1.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

dimana:

$(x_i - y_i)$ = selisih nilai atribut ke-i antara dua objek
 $\sum_{i=1}^n (...)$ = semua selisih atribut (komponen) dijumlahkan setelah dikuadratkan

$\sqrt{...}$ = akar kuadrat dari jumlah kuadrat selisih atribut

Perhitungan jarak dari setiap data ke titik pusat kluster dilakukan secara berulang. Proses perhitungan *Euclidian Distance* akan berhenti jika nilai pusat kluster stagnan atau tidak mengalami perubahan. Perhitungan *Euclidian Distance* yang digunakan dalam penelitian ini seperti pada Persamaan 2.

$$d(x, y) =$$

$$\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2} \quad (2)$$

Hasil perhitungan menggunakan *Euclidian Distance* pada iterasi pertama dapat dilihat pada Tabel 2.

Tabel 2. Hasil Perhitungan Iterasi Pertama

No.	Kabupaten / Kota	Jarak ke Centroid		
		K1 (Rendah)	K2 (Sedang)	K3 (Tinggi)
1	Nias	1544	8227.95	31363.22
2	Mandailing Natal	6818	605.60	26117.44
3	Tapanuli Selatan	8503	3643.20	24550.89
4	Tapanuli Tengah	9646	5441.10	23529.03
5	Tapanuli Utara	9957	6628.46	23684.17
6	Toba	6455	0.00	26496.34
7	Labuhan Batu	10432	6400.73	22548.46
8	Asahan	24236	28446.35	8738.95
9	Simalungun	20999	24060.05	11993.20
10	Dairi	2975	5751.82	29982.39

No.	Kabupaten / Kota	Jarak ke Centroid		
		K1 (Rendah)	K2 (Sedang)	K3 (Tinggi)
11	Karo	3269	5331.34	29681.44
12	Deli Serdang	24408	28848.12	8530.58
13	Langkat	25269	30732.03	7690.37
14	Nias Selatan	3461	5013.38	29456.33
15	Humbang Hasundutan	2990	5719.05	30028.91
16	Pakpak Bharat	0	10809.67	32897.61
17	Samosir	9656	5907.52	23806.29
18	Serdang Bedagai	15499	14557.98	17425.20
19	Batu Bara	11842	8592.50	21086.17
20	Padang Lawas Utara	2244	7103.08	30658.53
21	Padang Lawas	9137	4356.41	23775.56
22	Labuhan Batu Selatan	1575	8208.77	31325.12
23	Labuhan Batu Utara	3680	4609.20	29250.69
24	Nias Utara	949	9276.82	31951.57
25	Nias Barat	81	10695.58	32846.28
26	Sibolga	1022	9147.35	31879.64
27	Tanjung Balai	4374	3386.34	28612.82
28	Pematang Siantar	6087	600.38	26909.44
29	Tebing Tinggi	6725	775.24	26173.02
30	Medan	32898	43043.79	0.00
31	Binjai	6693	691.89	26207.65
32	Padangsidempuan	3901	4408.66	29002.29
33	Gunungsitoli	4084	3911.40	28859.86

Berikutnya ditentukan jarak terpendek data terhadap pusat kluster awal yang sebelumnya dipilih secara acak. Penentuan jarak terpendek dengan pusat

kluster dengan memilih nilai yang terkecil. Hasil penentuan jarak terpendek terhadap pusat kluster yang telah ditabulasikan dapat dilihat pada Tabel 3.

Tabel 3. Tabulasi Binner Jarak Terdekat Data Terhadap Kluster

No.	Kabupaten / Kota	Jarak ke Centroid		
		K1 (Rendah)	K2 (Sedang)	K3 (Tinggi)
1	Nias	1	0	0
2	Mandailing Natal	0	1	0
3	Tapanuli Selatan	0	1	0
4	Tapanuli Tengah	0	1	0
5	Tapanuli Utara	0	1	0
6	Toba	0	1	0
7	Labuhan Batu	0	1	0
8	Asahan	0	0	1
9	Simalungun	0	0	1
10	Dairi	1	0	0
11	Karo	1	0	0
12	Deli Serdang	0	0	1
13	Langkat	0	0	1
14	Nias Selatan	1	0	0
15	Humbang Hasundutan	1	0	0
16	Pakpak Bharat	1	0	0
17	Samosir	0	1	0
18	Serdang Bedagai	0	1	0
19	Batu Bara	0	1	0
20	Padang Lawas Utara	1	0	0
21	Padang Lawas	0	1	0
22	Labuhan Batu Selatan	1	0	0
23	Labuhan Batu Utara	1	0	0
24	Nias Utara	1	0	0
25	Nias Barat	1	0	0
26	Sibolga	1	0	0
27	Tanjung Balai	0	1	0
28	Pematang Siantar	0	1	0
29	Tebing Tinggi	0	1	0
30	Medan	0	0	1
31	Binjai	0	1	0

No.	Kabupaten / Kota	Jarak ke Centroid		
		K1 (Rendah)	K2 (Sedang)	K3 (Tinggi)
32	Padangsidempuan	1	0	0
33	Gunungsitoli	0	1	0

Berdasarkan Tabel 3, maka kabupaten / kota dapat dikelompokkan ke dalam kluster rendah, sedang, tinggi seperti pada Tabel 4.

Tabel 4. Hasil Klasterisasi Iterasi Pertama

Kluster	Anggota Kluster	Jumlah Anggota
K1 (Rendah)	1,10,11,14,15,16,20,22,23,24,25,26,32	13
K2 (Sedang)	2,3,4,5,6,7,17,18,19,21,27,28,29,31,33	15
K3 (Tinggi)	8,9,12,13,30	5

Tahap selanjutnya, dilakukan penentuan titik pusat kluster baru dengan menghitung rata – rata tiap kluster untuk setiap variabel. Hasil perhitungan rata – rata dapat dilihat pada Tabel 5.

Berdasarkan Tabel 5, diperoleh titik pusat kluster baru yang menjadi acuan pada iterasi kedua, dengan nilai sebagai berikut:

- $K_1 = [1256,31; 2230,85; 123,46]$
- $K_2 = [4160,60; 7713,14; 606,33]$
- $K_3 = [8297; 20069,20; 1181,40]$

Tabel 5. Titik Pusat Kluster Baru

K ₁	X ₁	1256.31
	X ₂	2230.85
	X ₃	123.46
K ₂	X ₁	4160.60
	X ₂	7713.47
	X ₃	606.33
K ₃	X ₁	8297.00
	X ₂	20069.20
	X ₃	1181.40

Selanjutnya, dilakukan perhitungan untuk iterasi kedua hingga hasil titik pusat kluster tidak mengalami perubahan nilai. Ketika titik pusat kluster tidak mengalami perubahan nilai, maka proses klasterisasi selesai. Pada penelitian ini, proses klasterisasi berhenti pada iterasi keempat karena nilai titik pusat kluster pada iterasi keempat sama dengan nilai titik pusat kluster pada iterasi ketiga. Hasil perhitungan pada iterasi keempat dapat dilihat pada Tabel 6.

Tabel 6. Hasil Perhitungan Iterasi Keempat

No.	Kabupaten / Kota	Jarak ke Centroid		
		K1 (Rendah)	K2 (Sedang)	K3 (Tinggi)
1	Nias	865.37	7440.57	24013.67
2	Mandailing Natal	4420.89	2167.77	18768.67
3	Tapanuli Selatan	6106.30	781.95	17217.11
4	Tapanuli Tengah	7269.28	1132.84	16197.78
5	Tapanuli Utara	7621.98	2342.59	16430.32
6	Toba	4059.94	2521.86	19145.93
7	Labuhan Batu	8051.65	1621.85	15196.05
8	Asahan	21871.86	15413.41	1641.31
9	Simalungun	18605.58	12052.49	4737.90
10	Dairi	604.39	6022.34	22633.76
11	Karo	911.49	5719.76	22328.99
12	Deli Serdang	22039.13	15565.61	1348.48
13	Langkat	22885.16	16345.80	726.54
14	Nias Selatan	1071.81	5530.70	22107.89
15	Humbang Hasundutan	713.39	6046.17	22680.48
16	Pakpak Bharat	2399.37	8975.98	25547.08
17	Samosir	7304.00	1788.35	16514.82
18	Serdang Bedagai	13121.94	6634.73	10069.72
19	Batu Bara	9467.40	3045.91	13730.59
20	Padang Lawas Utara	270.84	6748.78	23309.23
21	Padang Lawas	6747.01	657.12	16427.55
22	Labuhan Batu Selatan	840.38	7404.91	23974.21
23	Labuhan Batu Utara	1280.81	5301.37	21901.26
24	Nias Utara	1472.52	8042.79	24601.81
25	Nias Barat	2342.19	8920.17	25496.00
26	Sibolga	1395.29	7960.86	24528.32
27	Tanjung Balai	1986.75	4644.06	21267.05
28	Pematang Siantar	3690.89	2925.53	19564.57
29	Tebing Tinggi	4355.72	2408.80	18821.76
30	Medan	30518.13	23999.78	7355.62
31	Binjai	4317.58	2439.87	18859.53

No.	Kabupaten / Kota	Jarak ke Centroid		
		K1 (Rendah)	K2 (Sedang)	K3 (Tinggi)
32	Padangsidempuan	1532.87	5096.00	21650.42
33	Gunungsitoli	1686.49	4906.20	21511.78

Selanjutnya dilakukan penentuan jarak terpendek dengan pusat klaster dengan memilih nilai yang terkecil, hasil yang telah ditabulasikan dapat dilihat pada Tabel 7.

Tabel 7. Tabulasi Binner Jarak Terdekat Data Terhadap Klaster

No.	Kabupaten / Kota	Jarak ke Centroid		
		K1 (Rendah)	K2 (Sedang)	K3 (Tinggi)
1	Nias	1	0	0
2	Mandailing Natal	0	1	0
3	Tapanuli Selatan	0	1	0
4	Tapanuli Tengah	0	1	0
5	Tapanuli Utara	0	1	0
6	Toba	0	1	0
7	Labuhan Batu	0	1	0
8	Asahan	0	0	1
9	Simalungun	0	0	1
10	Dairi	1	0	0
11	Karo	1	0	0
12	Deli Serdang	0	0	1
13	Langkat	0	0	1
14	Nias Selatan	1	0	0
15	Humbang Hasundutan	1	0	0
16	Pakpak Bharat	1	0	0
17	Samosir	0	1	0
18	Serdang Bedagai	0	1	0
19	Batu Bara	0	1	0
20	Padang Lawas Utara	1	0	0
21	Padang Lawas	0	1	0
22	Labuhan Batu Selatan	1	0	0
23	Labuhan Batu Utara	1	0	0
24	Nias Utara	1	0	0
25	Nias Barat	1	0	0
26	Sibolga	1	0	0
27	Tanjung Balai	1	0	0
28	Pematang Siantar	0	1	0
29	Tebing Tinggi	0	1	0
30	Medan	0	0	1
31	Binjai	0	1	0
32	Padangsidempuan	1	0	0
33	Gunungsitoli	1	0	0

Berdasarkan Tabel 7, maka kabupaten / kota dapat dikelompokkan ke dalam klaster rendah, sedang, tinggi seperti pada Tabel 8.

Tabel 8. Hasil Klasterisasi Iterasi Keempat

Klaster	Anggota Klaster	Jumlah Anggota
K1 (Rendah)	1,10,11,14,15,16,20,22,23,24,25,26,27,32,33	15
K2 (Sedang)	2,3,4,5,6,7,17,18,19,21,28,29,31	13
K3 (Tinggi)	8,9,12,13,30	5

Berdasarkan klasterisasi iterasi terakhir diperoleh hasil sebagai berikut:

- Klaster 1 (Rendah) dengan nilai *centroid* $K_1 = [1413.27; 2465.33; 119.13]$: Nias, Dairi, Karo,

Nias Selatan, Humbang Hasundutan, Pakpak Bharat, Padang Lawas Utara, Labuhan Batu Selatan, Labuhan Batu Utara, Nias Utara, Nias Barat, Sibolga, Tanjung Balai, Padangsidempuan, Gunungsitoli.

- Klaster 2 (Sedang) dengan nilai *centroid* $K_2 = [4426.31; 8286.38; 685.62]$: Mandailing Natal, Tapanuli Selatan, Tapanuli Tengah, Tapanuli Utara, Toba, Labuhan Batu, Samosir, Serdang Bedagai, Batu Bara, Padang Lawas, Pematang Siantar, Tebing Tinggi, Binjai
- Klaster 3 (Tinggi) dengan nilai titik pusat *centroid* $K_3 = [9633.20; 24091.20; 1346.8]$: Asahan, Simalungun, Deli Serdang, Langkat, Medan.

Hasil klasterisasi ini mendukung konsep bahwa industri kecil dan mikro di Indonesia memiliki karakteristik heterogen, terutama dipengaruhi akses

modal, jumlah tenaga kerja, dan dukungan infrastruktur (Tambunan, 2019).

Algoritma *K-Means Clustering* terbukti efektif dalam menangkap pola kesamaan (*similarities*) antardaerah, sejalan dengan penelitian-penelitian terdahulu yang mengaplikasikan algoritma ini untuk segmentasi UMKM di berbagai wilayah (Witten et al., 2017).

Dari sisi kebijakan publik, klasterisasi dapat menjadi acuan yang relevan bagi pemerintah daerah dalam menyalurkan bantuan sesuai kebutuhan klaster, yang mana hal ini sejalan dengan penelitian yang dilakukan oleh Supriyadi (2021) bahwa efektivitas program bantuan UMKM akan meningkat jika didahului oleh proses klasterisasi berbasis data.

Dengan demikian, hasil dan pembahasan penelitian ini menegaskan perlunya pengelompokan data industri kecil dan mikro sebelum merumuskan kebijakan. Tiga klaster yang terbentuk (rendah, sedang, dan tinggi) dapat menjadi pedoman awal untuk menetapkan prioritas intervensi, baik dari sisi pelatihan, pemberian modal, hingga pembangunan infrastruktur pendukung. Melalui pendekatan ini, upaya peningkatan kesejahteraan pelaku industri kecil dan mikro di Provinsi Sumatera Utara dapat dilakukan secara tepat sasaran dan berkelanjutan.

5. KESIMPULAN DAN SARAN

Secara keseluruhan, klasterisasi memperlihatkan bahwa pengelompokan berbasis data mampu memberikan gambaran heterogenitas industri kecil dan mikro di Provinsi Sumatera Utara. Perbedaan akses modal, kapasitas tenaga kerja, dan jumlah unit usaha sangat memengaruhi struktur klaster yang terbentuk. Klasterisasi usaha kecil dan mikro secara cermat dapat membantu pemerintah menyalurkan program bantuan dengan lebih efisien dan tepat sasaran. Dari penelitian ini, diperoleh beberapa konsekuensi kebijakan yang dapat diambil oleh pemerintah, diantaranya:

Wilayah klaster rendah (Melakukan perluasan akses pembiayaan melalui skema kredit berbunga rendah, Memberikan pelatihan dasar kewirausahaan dan manajemen keuangan bagi pelaku usaha lokal, Memberikan fasilitas pemasaran dan jaringan distribusi untuk meningkatkan penjualan).

Wilayah klaster sedang (Meningkatkan kapasitas teknologi melalui pelatihan penggunaan mesin atau digitalisasi, Dorongan untuk inovasi produk dan kolaborasi antarusaha di daerah sekitarnya, Stimulus ekspansi pasar, seperti mengikuti pameran dagang dan pelatihan e-commerce).

Wilayah klaster tinggi (Fasilitasi riset dan pengembangan untuk menciptakan produk bernilai tambah tinggi, Penguatan jaringan global (ekspor) dan penyertaan modal ventura, Diversifikasi produk dan perluasan lini bisnis agar industri kecil dan mikro tidak hanya bergantung pada satu sektor.

[12]

Data yang digunakan sebagian besar bersifat agregat dan mungkin belum mencerminkan variasi internal di tingkat desa atau kecamatan. Penggunaan variabel lainnya, seperti tingkat pendidikan dan adopsi teknologi digital, belum disertakan dalam penelitian ini tetapi dapat mempengaruhi performa klasterisasi. Pemilihan $k=3$ didasarkan pada kebutuhan penelitian dan pertimbangan praktis, penggunaan jumlah klaster yang berbeda memungkinkan menghasilkan segmentasi yang lebih mendetail.

DAFTAR PUSTAKA

- [1] Kementerian Perindustrian, "Statistik Industri Kecil dan Mikro 2023", Jakarta: Kementerian Perindustrian, 2023.
- [2] T.T.H. Tambunan, *UMKM di Indonesia: Masalah dan Kebijakan Pemerintah*, Jakarta: Gadjah Mada University Press, 2019.
- [3] Badan Pusat Statistik, "Profil Industri Mikro dan Kecil Provinsi Sumatera Utara 2023", Medan: BPS, 2023.
- [4] A. Supriyadi, "Strategi Pengembangan UMKM Berbasis Klasterisasi Data", *Jurnal Ekonomi dan Bisnis*, vol. 14, no. 3, pp. 200-220, 2021.
- [5] I. H. Witten and E. Al, *Data Mining: Practical Machine Learning Tools and Techniques*, Amsterdam: Morgan Kaufmann, 2017.
- [6] W. Wang, *Principles of Machine Learning*, Springer Nature, 2025.
- [7] E. Alpaydin, *Introduction to Machine Learning*, Cambridge, Massachusetts: The Mit Press, 2020.
- [8] J. Wu, *Advances in K-means Clustering a Data Mining Thinking*, Berlin Springer Berlin, 2014.
- [9] A. I. Silitonga, Z. A. Nabila, C. R. Z. Lubis, N. Safitri, and Haryadi, "Klasterisasi Gizi Buruk Dan Stunting di Provinsi Sumatera Utara Menggunakan K-Means Clustering," *METHODIKA: Jurnal Teknik Informatika dan Sistem Informasi*, vol. 10, no. 2, pp. 13-18, Sep. 2024, doi: <https://doi.org/10.46880/mtk.v10i2.3147>.
- [10] A. I. Silitonga, M. A. Nasution, A. Fitri, K. S. Rizwinie, A. Hidayatullah, and Y. Simamora, "Klasterisasi Penyebaran Base Transceiver Station Menggunakan K-Means Clustering," *Semnas Ristek (Seminar Nasional Riset dan Inovasi Teknologi)*, vol. 9, no. 1, pp. 344-351, Jan. 2025, doi: <https://doi.org/10.30998/semnasristek.v9i1.794>.
- [11] Azzam, A. I. Purnamasari, and I. Ali, "Implementasi Algoritma K-Means Clustering Untuk Analisis Persebaran Umkm Di Jawa Barat," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 3062-3070, May 2024, doi: <https://doi.org/10.36040/jati.v8i3.8450>.