

Contents list available at www.jurnal.unimed.ac.id

CESS
(Journal of Computing Engineering, System and Science)

journal homepage: <https://jurnal.unimed.ac.id/2012/index.php/cess>



**Analisis Pengelompokan UMKM Berdasarkan Kategori Menggunakan
Algoritma K-Means dan K-Medoids**

Analysis of MSME Grouping Based on Categories Using the K-Means and K-Medoids Algorithms

Cweto Bolodiko Ajarwiro¹, Maimunah^{2*}, Pristi Sukmasetya³

^{1,2,3} Teknik Informatika, Universitas Muhammadiyah Magelang

Jl. Mayjend. Bambang Soegeng KM 5 Mertoyudan, Magelang 56172

email: ¹cwetoajarwiro@gmail.com, ²maimunah@unimma.ac.id, ³pristi.sukmasetya@ummql.ac.id

ABSTRAK

Sektor Usaha Mikro Kecil dan Menengah (UMKM) menjadi salah satu faktor utama yang mendorong pertumbuhan ekonomi Indonesia. Jumlah kategori UMKM yang banyak perlu dilakukan pengelompokan agar dapat membantu pemerintah dalam mendukung pengembangan UMKM. Dalam penelitian ini dilakukan pengelompokan UMKM berdasarkan kategori menggunakan algoritma *K-Means* dan *K-Medoids*. Data yang digunakan dalam penelitian ini adalah data UMKM Kota Magelang yang diambil dari PERSDA sebanyak 3491. Pada tahap pengolahan data dilakukan pengecekan tipe data, penanganan data yang hilang, pelabelan dan penjumlahan kategori UMKM. Setelah data diolah maka dilakukan pengelompokan data menggunakan algoritma *K-Means* dan *K-Medoids*. Pengelompokan kategori UMKM menggunakan algoritma *K-Means* dan *K-Medoids* menghasilkan 3 kluster. Pengelompokan menggunakan *K-Means* menghasilkan kluster tinggi sebanyak 1 kategori, kluster sedang 3 kategori, dan kluster rendah 60 kategori. Pengelompokan menggunakan *K-Medoids* menghasilkan kluster tinggi 1 kategori, kluster sedang 2 kategori, dan kluster rendah 61 kategori. Berdasarkan nilai DBI, algoritma *K-Means* mempunyai nilai 0,496 sedangkan algoritma *K-Medoids* bernilai 0,499. Dengan demikian klustering UMKM Kota Magelang menggunakan *K-Means* lebih baik daripada algoritma *K-Medoids*. Melalui pengelompokan UMKM berdasarkan kategori dapat membantu memberikan informasi untuk pengembangan UMKM.

Kata Kunci: *UMKM, Klustering, K-means, K-Medoids.*

ABSTRACT

The Micro, Small and Medium Enterprises (MSME) sector is one of the main factors driving Indonesia's economic growth. The large number of MSME categories needs to be grouped to

**Penulis Korespondensi:*

email: maimunah@unimma.ac.id

assist the government in supporting the development of MSMEs. This research grouped MSMEs based on categories using the K-Means and K-Medoids algorithms. The data used in this research is 3491 Magelang City MSME data taken from PerSADA. Data type checking, handling of missing data, labelling and adding up MSME categories were carried out at the data processing stage. After the data has been processed, it is grouped using the K-Means and K-Medoids algorithms. Grouping the MSME category using the K-Means and K-Medoids algorithms produces 3 clusters. Grouping using K-Means creates a high cluster of 1 category, a medium cluster of 3, and a low cluster of 60 categories. Grouping using K-Medoids produces a high cluster of 1 category, a medium cluster of 2 categories, and a low cluster of 61 categories. Based on the DBI value, the K-Means algorithm has a value of 0.496, while the K-Medoids algorithm has a value of 0.499. Thus, Magelang City MSME clustering using K-Means is better than the K-Medoids algorithm. Grouping MSMEs based on categories can provide information for MSME development.

Keywords: MSME, Clustering, K-Means, K-Medoids.

1. PENDAHULUAN

Sektor Usaha Mikro Kecil dan Menengah (UMKM) menjadi salah satu faktor utama yang mendorong pertumbuhan ekonomi Indonesia[1]. Website PerSADA (<https://satudata.dinkop-umkm.jatengprov.go.id/>) adalah sistem informasi yang menyediakan layanan data terpadu untuk koperasi dan UMKM, dikelola oleh Dinas Koperasi, Usaha Kecil, dan Menengah Provinsi Jawa Tengah. UMKM Kota Magelang tercatat berjumlah 3.491 di PerSADA dan UMKM dibagi menjadi 9 kategori, yaitu Pertanian, Peternakan, Kehutanan, Perikanan; Pertambangan dan Penggalian; Industri Pengolahan, Listrik, Air, Gas; Bangunan; Perdagangan, Hotel, Resto; Pengangkutan dan Komunikasi; Keuangan, Sewa, Jasa Perusahaan; dan Lainnya. PerSADA memiliki salah satu tugas dalam membuat rekapitulasi data dan statistik UMKM di seluruh provinsi Jawa Tengah. Mengingat jumlah UMKM yang cukup signifikan, diperlukan pengelompokan UMKM agar dapat membantu pemerintah terkait untuk penetapan strategi pemasaran yang tepat sebagai prioritas utama untuk mengembangkan pasar [2].

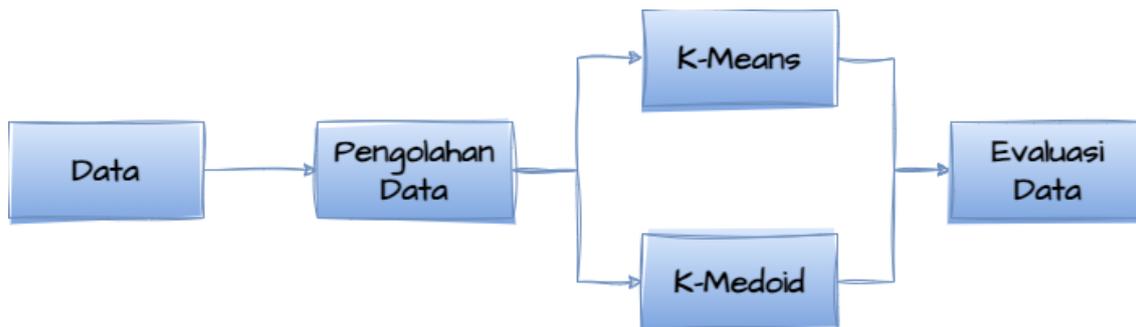
Penelitian yang membahas tentang pengelompokan UMKM sudah banyak dilakukan. Teknik partisi adalah salah satu pendekatan dalam pengelompokan data. Algoritma *K-Means* dan algoritma *K-Medoids* termasuk dalam kategori teknik partisi dalam proses pengelompokan [3], [4]. Penelitian pengelompokan UMKM menggunakan *Rapidminer* dengan algoritma *K-Means Clustering* yang menghasilkan 3 klaster yaitu mikro, kecil dan menengah yang berdasarkan aset dan omset [5]. Penelitian yang membandingkan algoritma *K-Means* dengan *K-Medoids* pada data transaksi bongkar muat yang membutuhkan waktu pengolahan 1 menit untuk *K-Means* dan 1 menit 38 detik untuk *K-Medoids* [6]. Penelitian yang membahas pengelompokan UMKM Kebumen dengan membandingkan algoritma *K-Means* dengan *K-Medoids* yang menghasilkan 3 klaster baik dengan algoritma *K-Means* maupun *K-Medoids* [7].

Berdasarkan penelitian sebelumnya, maka akan dilakukan penelitian dengan menggunakan algoritma *K-Means* dan *K-Medoids* untuk membandingkan dalam pengelompokan data UMKM Kota Magelang. Algoritma *K-Means* dan *K-Medoids* adalah dua metode pengelompokan yang memiliki perbedaan dalam cara mereka menentukan pusat klaster dan dalam hal ketahanan terhadap *outlier* [8]. Maka pada penelitian ini akan

membandingkan kedua algoritma dalam pengelompokan data dan mengukur validitas hasil pengelompokan.

2. METODE

Metode penelitian dilakukan dengan Data yang siap dilakukan Preprocessing Data lalu dikelompokkan dengan 2 metode yaitu *K-Means* dan *K-Medoids* yang hasil dari kedua metode akan dievaluasi dan dibandingkan [9][11].



Gambar 1. Alur metode penelitian

2.1. Data

Metode yang digunakan dalam pengumpulan data penelitian adalah data sekunder [12]. Data UMKM Kota Magelang didapatkan dari website PeRSADA [13]. Data UMKM yang digunakan memiliki atribut NIK, Nama Usaha, Klasifikasi, Kategori, Tenaga Kerja Laki-Laki dan Tenaga Kerja Perempuan dengan jumlah data sebanyak 3.491.

2.2. Pengolahan Data

Proses pengolahan data ini menghasilkan data final yang akan digunakan dalam pemodelan data [13], [14]. Pengolahan data meliputi pengecekan nilai yang hilang (*missing value*), data duplikat, pengecekan tipe data dan *label encoder*. Jika terdapat *missing value* maka nilai yang hilang diganti dengan nilai *modus* karena data yang digunakan termasuk data kategorikal. Selanjutnya data dikelompokkan berdasarkan kategori UMKM dengan atribut yaitu kategori dan jumlah yang menyatakan jumlah UMKM kemudian dilakukan *label encoder*. Hasil dari pengolahan data digunakan sebagai *input* dalam model klustering.

2.3. Metode yang Digunakan

2.3.1. Algoritma K-Means

Berikut ini adalah tahapan-tahapan untuk menghitung metode *K-Means* [15], [16]:

1. Tentukan jumlah K (Klaster) dengan nilai yang lebih kecil dari total jumlah data.
2. Pilih K titik klaster secara acak, di mana setiap titik ini akan menjadi pusat *centroid* dari kelompok (klaster) masing-masing.
3. Hitung jarak dan alokasikan masing-masing data ke *Centroid* atau rata-rata terdekat, dengan rumus:

$$d(x_j, c_j) = \sqrt{\sum_{j=1}^n (x_j - c_j)^2} \quad (1)$$

Keterangan:

d = jarak

n = banyaknya objek

j = dimulai 1 sampai n

x_j = *feature* objek ke j terhadap x

c_j = *Centroid* feature ke j

2.3.2. Algoritma K-Medoids

Berikut ini adalah tahapan-tahapan untuk menghitung metode *K-Medoids* [17], [18]:

1. Menentukan pusat kluster sebanyak k .
2. Mengalokasikan data untuk kluster terdekat dengan rumus ukuran jarak *euclidean distance*.
3. Memilih acak objek pada setiap cluster yang digunakan untuk *medoid* baru.
4. Menghitung jarak masing – masing objek berada pada setiap kluster.
5. Kandidat *medoids* baru dengan menggunakan rumus *euclidian distance*.
6. Menghitung total simpangan (S) menggunakan menghitung nilai total *distance* baru sampai total *distance* lama. Jika $S < 0$, maka tukar objek dengan hasil data klustering untuk membentuk sekumpulan k objek baru untuk *medoids*.
7. Ulangi urutan 3 sampai 5 hingga tidak ada perubahan *medoids*.
8. Sampai mendapatkan kluster beserta masing – masing anggota kluster.

2.4. Evaluasi Hasil

Hasil kluster akan di evaluasi dengan metode yaitu *Davies Bouldin Index* (DBI). DBI adalah ukuran untuk mengevaluasi kinerja pengelompokan [19][21]. Hasil dari DBI akan dibandingkan antara algoritma *K-Means* dengan *K-Medoids*. Tujuan daripada perbandingan ini untuk mengetahui yang mana semakin baik kluster yang diperoleh dari pengelompokan. Rentang nilai DBI dimulai dari nilai optimal (semakin kecil) yaitu nilai DBI yang semakin kecil menunjukkan kualitas clustering yang semakin baik. Nilai mendekati 0 menandakan bahwa kluster bersifat kompak dan terpisah dengan baik [23].

Tabel 1. Rentang Nilai DBI

DBI	Keterangan
0	Kluster bersifat kompak dan terpisah
0 – 0,5	Kluster baik dengan sejumlah pengecualian
0,5 - 1	Kluster dapat diterima, namun perlu diperbaiki

3. HASIL DAN PEMBAHASAN

Data yang digunakan adalah data UMKM yang diperoleh dari situs website resmi PeRSADA [13]. Jumlah data adalah 3941 UMKM dengan atribut NIK, Nama Usaha, Klasifikasi, Kategori, Tenaga Kerja Laki-Laki dan Tenaga Kerja Perempuan seperti pada Tabel 2.

Tabel 2. Dataset

No	NIK	Nama Usaha	Klasifikasi	Kategori	Tenaga Kerja Laki-laki	Tenaga Kerja Perempuan
1	33710167xxxxxx xxx	Linawati	Mikro	Makanan/Minuman	1	1
2	33710163xxxxxx xxx	Snack Diyah	Mikro	Kegiatan Lainnya	0	0
3	33710149xxxxxx xx	Penjahit Mbak Nur	Mikro	Kegiatan Lainnya	0	0
...
34 89	33710202xxxxxx xx	Sudarno kelontong	Kecil	Makanan/Minuman	0	0
34 90	33011158xxxxxx xxx	PawonR N	Kecil	Makanan/Minuman	1	0

Pada tahap pengolahan data dilakukan pengecekan data dan pengolahan data. Pengecekan data dilakukan dengan pengecekan tipe data dan data duplikat. Pada tahap pengolahan data dilakukan pengecekan *missing value*. *Missing value* dalam data sebanyak 13 yang ditampilkan pada Gambar 1 dan ditangani dengan nilai *modus* sehingga tidak ada data *missing value* yang disajikan pada Gambar 2.

No	NIK	Nama Usaha	Klasifikasi	Kategori	Tenaga Kerja Laki Laki	Tenaga Kerja Perempuan
2821	2822 33710267xxxxxxxx	BATIK SYAKRONI	NaN	NaN	0	0
2822	2823 33710241xxxxxxxx	Batik Kinaryo Priyamah	NaN	NaN	0	0
2823	2824 33710265xxxxxxxx	BATIK KEBONPOLO II	Mikro	NaN	0	0
2824	2825 33710104xxxxxxxx	BATIK BONPOLO	Mikro	NaN	3	3
2825	2826 33081015xxxxxxxx	batik kebonpolo	NaN	NaN	2	2
2826	2827 33710111xxxxxxxx	KOMUNITAS BATIK SIGRAK BAWONO	NaN	NaN	0	0
2828	2829 33710268xxxxxxxx	RAHAYU	Mikro	NaN	0	0
2832	2833 33710264xxxxxxxx	pregobatik	NaN	NaN	0	0
2836	2837 33710262xxxxxxxx	batik kinaryo arum	Mikro	NaN	0	0
2844	2845 33710154xxxxxxxx	UKM GRIYA AYU SALON DAN SPA HERBAL	Mikro	NaN	0	4
2850	2851 33710166xxxxxxxx	Kerajinan Alvarizqi	Mikro	NaN	0	1
2851	2852 33710161xxxxxxxx	Keenan Florist Magelang & Qproet	Mikro	NaN	1	1
2854	2855 33710165xxxxxxxx	catring Sego Besek	Mikro	NaN	1	5

Gambar 1. Missing value

No	NIK	Nama Usaha	Klasifikasi	Kategori	Tenaga Kerja Laki Laki	Tenaga Kerja Perempuan	Kategori_encoded
2821	2822	33710267xxxxxxxxxx	BATIK SYAKRONI	Mikro Kegiatan Lainnya	0	0	66
2822	2823	33710241xxxxxxxxxx	Batik Kinaryo Priyamah	Mikro Kegiatan Lainnya	0	0	66
2823	2824	33710265xxxxxxxxxx	BATIK KEBONPOLO II	Mikro Kegiatan Lainnya	0	0	66
2824	2825	33710104xxxxxxxxxx	BATIK BONPOLO	Mikro Kegiatan Lainnya	3	3	66
2825	2826	33081015xxxxxxxxxx	batik kebonpolo	Mikro Kegiatan Lainnya	2	2	66
2826	2827	33710111xxxxxxxxxx	KOMUNITAS BATIK SIGRAK BAWONO	Mikro Kegiatan Lainnya	0	0	66
2828	2829	33710268xxxxxxxxxx	RAHAYU	Mikro Kegiatan Lainnya	0	0	66
2832	2833	33710264xxxxxxxxxx	pregobatik	Mikro Kegiatan Lainnya	0	0	66
2836	2837	33710262xxxxxxxxxx	batik kinaryo arum	Mikro Kegiatan Lainnya	0	0	66
2844	2845	33710154xxxxxxxxxx	UKM GRIYAA YU SALON DAN SPA HERBAL	Mikro Kegiatan Lainnya	0	4	66
2850	2851	33710166xxxxxxxxxx	Kerajinan Alvarizqi	Mikro Kegiatan Lainnya	0	1	66
2851	2852	33710161xxxxxxxxxx	Keenan Florist Magelang & Qproet	Mikro Kegiatan Lainnya	1	1	66
2854	2855	33710165xxxxxxxxxx	catring Sego Besek	Mikro Kegiatan Lainnya	1	5	66

Gambar 2. Penanganan missing value

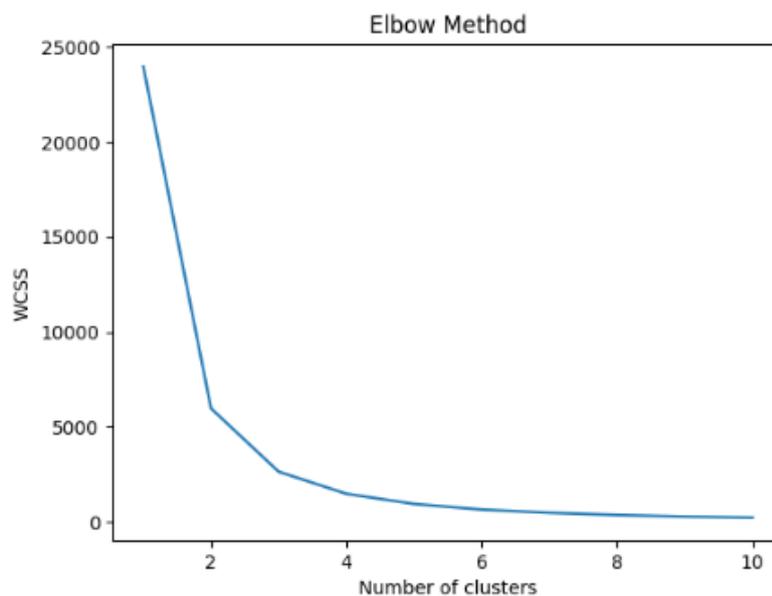
Selanjutnya dilakukan perhitungan jumlah data setiap kategori pada atribut Kategori yang akan menghasilkan dua atribut yaitu kategori dan jumlah seperti pada Gambar 3a.

Kategori	Jumlah	Kategori_encoded	Jumlah
Jual beli burung	1	16	1677
Jual beli dongkrak bekas	1	35	632
Tambal Ban	1	28	524
toko Alquran	1	6	301
jual beli barang bekas	1	8	97
...
Jasa	97	34	1
Industri Pengolahan	301	15	1
Makanan/Minuman	524	27	1
Perdagangan dan Reparasi	632	22	1
Kegiatan Lainnya	1677	12	1

Gambar 3. Hasil perhitungan data berdasarkan kategori (a) dan pelabelan data (b)

Gambar 3a adalah hasil dari perhitungan setiap kategori yang muncul di atribut kategori. Pada kolom kategori merupakan data semua jenis kategori yang ada di dalam data dan kolom jumlah adalah jumlah dari setiap kategori yang ada di dalam data. Kategori yang tercatat dalam data ada sebanyak 66 kategori. Hasil perhitungan data berdasarkan kategori diperoleh bahwa kategori Kegiatan Lainnya mempunyai jumlah yang paling banyak. Tahap pengolahan data selanjutnya adalah melakukan pelabelan untuk data kategori seperti pada Gambar 3b.

Proses awal dalam algoritma *K-Means* adalah menentukan jumlah kluster (*k*). Salah satu metode umum untuk menentukan *k* adalah menggunakan *Elbow Method* dengan mencoba beberapa nilai *k*. Pilih nilai *k* dengan penurunan nilai secara signifikan sehingga membentuk siku-siku pada grafik Gambar 4.



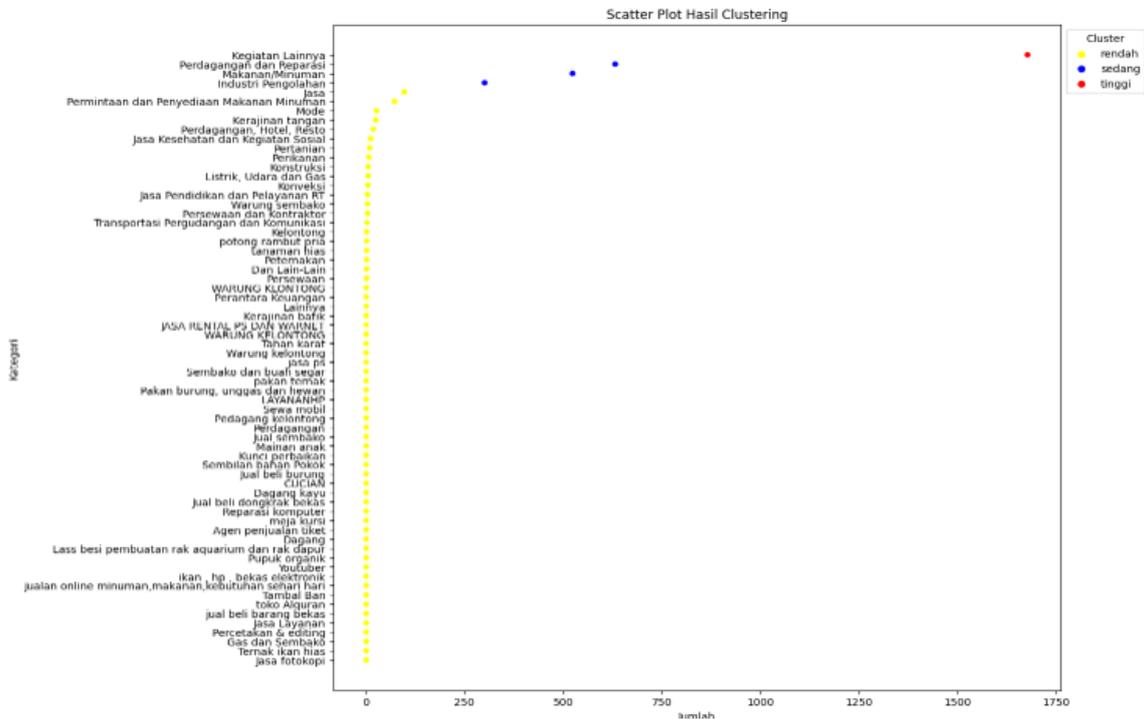
Gambar 4. Elbow method K-Means

Setelah menentukan k, algoritma *K-Means* memulai dengan menentukan *centroid* awal untuk setiap klaster secara acak. Titik data kemudian didistribusikan ke klaster berdasarkan jarak *Euclidean* terdekat ke *centroid* klaster. *Centroid* klaster dihitung sebagai rata-rata dari semua titik data dalam klaster. Klastering dilakukan dengan menggunakan algoritma *K-Means* dan mendapatkan hasil pengelompokannya.

Gambar 5 dan 6 adalah hasil dari klastering metode *K-Means* membagi data UMKM Kota Magelang menjadi tiga klaster. Penamaan klaster adalah tinggi, sedang dan rendah sesuai dengan jumlah UMKM berdasarkan kategori. Pengelompokan data menggabungkan UMKM mikro dan kecil. Klaster pertama hanya ada satu kategori yaitu Kegiatan Lainnya. Kategori Kegiatan Lainnya berisi seperti Jasa, Akomodasi dan Penyediaan Makanan Minuman, Jasa Pendidikan dan Pelayanan RT, Jasa Kesehatan dan Kegiatan Sosial, Konstruksi, dan Perorangan berdasarkan website PER-SADA[13]. Klaster kedua ada tiga kategori didalamnya yaitu Perdagangan dan Reparasi, Makanan/Minuman, Industri Pengolahan, dan Jasa. Klaster ketiga terdapat 60 kategori yang memiliki jumlah paling banyak diantara kedua klaster lainnya.

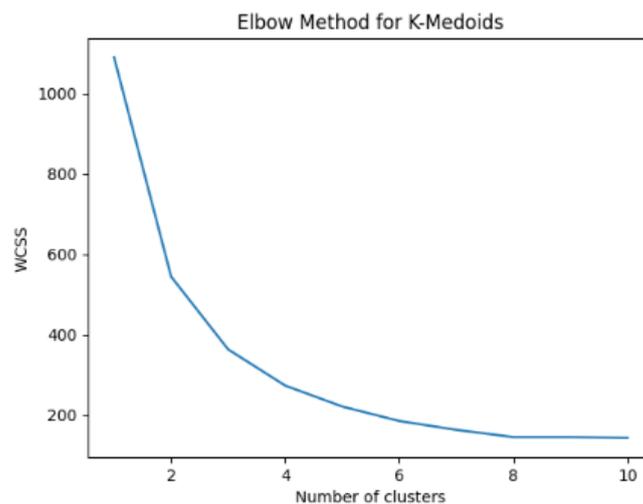
Kategori	Jumlah cluster	
Kegiatan Lainnya	1677	tinggi
Perdagangan dan Reparasi	632	sedang
Makanan/Minuman	524	sedang
Industri Pengolahan	301	sedang
Jasa	97	rendah
...
Jasa Layanan	1	rendah
Percetakan & editing	1	rendah
Gas dan Sembako	1	rendah
Ternak ikan hias	1	rendah
Jasa fotokopi	1	rendah

Gambar 5. Hasil Klastering *K-Means*



Gambar 6. Visualisasi klastering algoritma *K-Means*

Proses awal dalam algoritma *K-Medoids* adalah menentukan jumlah klaster (*k*). Salah satu metode umum untuk menentukan *k* adalah menggunakan *Elbow Method* dengan mencoba beberapa nilai *k*. Pilih nilai *k* dengan nilai penurunan secara signifikan sehingga membentuk siku-siku pada grafik pada Gambar 7.



Gambar 7. *Elbow method K-Medoids*

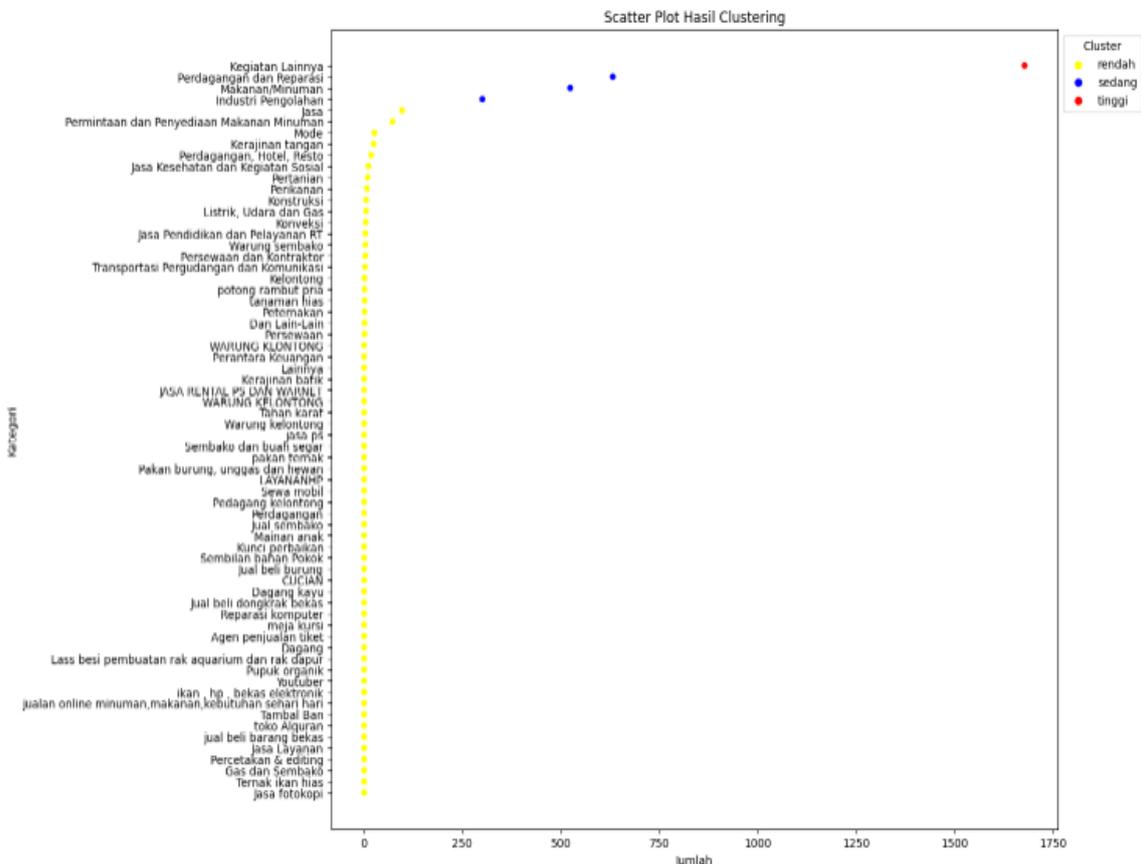
Setelah menentukan *k*, algoritma *K-Medoids* memulai dengan menentukan *medoid* awal untuk setiap klaster secara acak. Titik data kemudian diatribusikan ke klaster berdasarkan jarak *Euclidean* terdekat ke *medoid* klaster. *Medoid* klaster dihitung sebagai objek dalam klaster yang memiliki total jarak terkecil ke objek lain dalam klaster tersebut. Klastering

dilakukan dengan menggunakan algoritma *K-Medoids* dan mendapatkan hasil pengelompokan nya.

Gambar 8 dan 9 adalah hasil dari klastering metode *K-Medoids* membagi data UMKM Kota Magelang menjadi tiga klaster. Penamaan klaster adalah tinggi, sedang dan rendah sesuai dengan jumlah UMKM berdasarkan kategori. Pengelompokan data menggabungkan UMKM mikro dan kecil. Klaster pertama hanya ada satu kategori yaitu Kegiatan Lainnya. Klaster kedua ada tiga kategori didalamnya yaitu Perdagangan dan Reparasi, Makanan/Minuman dan Industri Pengolahan. Klaster ketiga terdapat 61 kategori yang memiliki jumlah paling banyak diantara kedua klaster lainnya.

Kategori	Jumlah	cluster
Kegiatan Lainnya	1677	tinggi
Perdagangan dan Reparasi	632	sedang
Makanan/Minuman	524	sedang
Industri Pengolahan	301	sedang
Jasa	97	rendah
...
Jasa Layanan	1	rendah
Percetakan & editing	1	rendah
Gas dan Sembako	1	rendah
Ternak ikan hias	1	rendah
Jasa fotokopi	1	rendah

Gambar 8. Hasil klastering *K-Medoids*



Gambar 9. Visualisasi klastering *K-Medoids*

3.1. Evaluasi Data

Hasil dari 2 metode dilakukan perbandingan dengan menggunakan *Davies-Bouldin Index* (DBI) untuk penilaiannya. Hasil nilai DBI klustering menggunakan *K-Means* dan *K-Medoids* seperti pada Tabel 3.

Tabel 3. Hasil Nilai DBI

Algoritma	Nilai DBI
<i>K-Means</i>	0,496
<i>K-Medoids</i>	0,499

Berdasarkan evaluasi kinerja pengelompokan menggunakan nilai DBI pada Tabel 3, algoritma *K-Means* mendapatkan nilai sebesar 0,496, sementara algoritma *K-Medoids* memperoleh nilai sebesar 0,499. Terdapat selisih 0,003 antara kedua algoritma, yang menunjukkan bahwa algoritma *K-Means* bekerja lebih efektif dibandingkan dengan algoritma *K-Medoids* dalam konteks klustering.

Perbedaan nilai DBI antara *K-Means* dan *K-Medoids* dapat disebabkan oleh sensitivitas *K-Means* terhadap *outlier*. Jika dataset memiliki *outlier*, *K-Medoids* mungkin memberikan hasil yang lebih baik karena penggunaan *medoid* yang lebih stabil terhadap variabilitas data. Meskipun selisih nilai DBI cukup kecil, penting untuk mempertimbangkan konteks data dan karakteristiknya. Dalam beberapa kasus, perbedaan ini mungkin tidak signifikan, tetapi pada data yang lebih kompleks atau dengan adanya *outlier*, keunggulan *K-Medoids* dalam menangani *outlier* dapat menjadi pertimbangan penting.

Algoritma *K-Means* dan *K-Medoids* adalah dua pendekatan yang berbeda dalam melakukan klustering. Meskipun *K-Means* dan *K-Medoids* memiliki tujuan yang sama, yaitu membentuk kluster yang homogen, terdapat perbedaan dalam cara mereka menentukan pusat kluster (*centroid* atau *medoid*) dan cara mengukur jarak antara data dengan pusat kluster. Namun, keefektifan suatu algoritma dapat sangat tergantung pada distribusi dan sifat spesifik dari data.

Dalam beberapa kasus, *K-Medoids* dapat lebih baik menangani *outlier* karena menggunakan *medoid* sebagai pusat kluster. Oleh karena itu, penting untuk menguji dan membandingkan kinerja keduanya pada dataset spesifik untuk menentukan algoritma yang paling sesuai. Keefektifan algoritma klustering dapat sangat tergantung pada karakteristik spesifik dataset. *K-Means* dan *K-Medoids* memiliki kelebihan dan kekurangan masing-masing, dan pilihan terbaik dapat bervariasi tergantung pada data yang dihadapi. Jika DBI menunjukkan bahwa *K-Means* lebih baik, itu mungkin merupakan pilihan yang baik untuk dataset.

Data yang digunakan dalam penelitian ini masih mempunyai informasi yang belum jelas yaitu untuk kategori Kegiatan Lainnya. Jumlah UMKM untuk kategori kegiatan Lainnya paling banyak dibandingkan kategori UMKM yang lain. Hal ini memberikan informasi yang tidak jelas sehingga identifikasi dan penanganan untuk pengembangan UMKM menjadi sulit dilakukan oleh pemerintah. Dengan demikian perlu peninjauan Kembali terhadap pengelompokan kategori UMKM di website PerSADA sehingga dapat memberikan penjelasan lebih detail tentang kategori UMKM untuk Kegiatan Lainnya.

4. KESIMPULAN

Hasil dari penelitian dari menggunakan algoritma *K-Means* dan *K-Medoids* menghasilkan 3 kluster. Algoritma *K-Means* kluster tinggi sebanyak 1 kategori, kluster sedang 3 kategori, dan kluster rendah 60 kategori. Algoritma *K-Medoids* kluster tinggi 1 kategori, kluster sedang 2 kategori, dan kluster rendah 61 kategori. Berdasarkan nilai DBI algoritma *K-Means* bernilai 0,496 sedangkan algoritma *K-Medoids* bernilai 0,499. Maka berdasarkan hasil DBI algoritma *K-Means* lebih bagus klustering nya daripada algoritma *K-Medoids* pada kasus Klustering UMKM Kota Magelang, karena nilai DBI *K-Means* lebih mendekati 0. Berdasarkan hasil klustering dapat memberikan gambaran yang berguna bagi pihak terkait untuk merancang strategi atau memberikan dukungan khusus terhadap pengembangan UMKM Kota Magelang.

REFERENSI

- [1] G. Wijaya, "Klasifikasi UMKM Menggunakan Algoritma Naive Bayes Berdasarkan Sudah Pernah Mempunyai Atau Mengurus Sertifikat Halal," *JDMSI*, vol. 4, no. 1, p. 36, Mar. 2023, doi: 10.33365/jdmsi.v4i1.2634.
- [2] P. Puntoriza and C. Fibriani, "Analisis Persebaran UMKM Kota Malang Menggunakan Cluster K-means," *J Inform Syst*, vol. 5, no. 1, pp. 86–94, May 2020, doi: 10.33633/joins.v5i1.3469.
- [3] K. P. Sinaga and M.-S. Yang, "Unsupervised K-Means Clustering Algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [4] F. W. Saputri and D. B. Arianto, "Perbandingan Performa Algoritma K-Means, K-Medoidss, Dan DBSCAN Dalam Penggerombolan Provinsi Di Indonesia Berdasarkan Indikator Kesejahteraan Masyarakat," *JTI*, vol. 7, no. 2, pp. 138–151, Aug. 2023, doi: 10.47111/jti.v7i2.9558.
- [5] W. Sudrajat, I. Cholid, and J. Petrus, "Penerapan Algoritma K-Means Clustering untuk Pengelompokan UMKM Menggunakan Rapidminer," vol. 14, no. 1, 2022.
- [6] I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *rmsi*, vol. 5, no. 1, p. 119, Feb. 2019, doi: 10.24014/rmsi.v5i1.7381.
- [7] R. Wahyusari and S. Wardani, "Perbandingan Algoritma K-Means dan Algoritma K-Medoid Untuk Pengelompokan UMKM di Kebumen".
- [8] B. Lund and J. Ma, "A review of cluster analysis techniques and their uses in library and information science research: *k-means* and *K-Medoids* clustering," *PMM*, vol. 22, no. 3, pp. 161–173, Nov. 2021, doi: 10.1108/PMM-05-2021-0026.
- [9] U. R. Gurning and M. Mustakim, "Penerapan Algoritma K-Means dan *K-Medoids* untuk Pengelompokan Data Pasien Covid-19," *bits*, vol. 3, no. 1, pp. 48–55, Jun. 2021, doi: 10.47065/bits.v3i1.1003.
- [10] S. Harini and R. Kusumawati, "Evaluasi Clustering K-Means Dan *K-Medoids* Pada Persebaran Covid-19 Di Indonesia Dengan Metode Davies-Bouldin Index (DBI)," vol. 6, no. 2, 2023.
- [11] S. Y. Riska and L. Farokhah, "Perbandingan Hasil Evaluasi Algoritma K-Means dan *K-Medoids* Berdasarkan Kunjungan Wisatawan Mancanegara ke Indonesia".
- [12] A. P. Y. K. Wardani and N. A. S. Darmawan, "Peran Financial Technology pada UMKM: Peningkatan Literasi Keuangan Berbasis Payment Gateway," *JIAH*, vol. 10, no. 2, p. 170, Aug. 2020, doi: 10.23887/jiah.v10i2.25947.

- [13] Website Portal Satu Data (PeRSADA),” <https://satudata.dinkop-umkm.jatengprov.go.id/>” diakses : 15 Oktober 2023.
- [14] R. Ordila, R. Wahyuni, Y. Irawan, and M. Yulia Sari, “Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma Clustering (Studi Kasus: Poli Klinik PT.Inecda),” *jik*, vol. 9, no. 2, pp. 148–153, Oct. 2020, doi: 10.33060/JIK/2020/Vol9.Iss2.181.
- [15] B. Arifitama and A. Syahputra, “Analisis Data Mining Pada Klasterisasi UMKM Dengan Menggunakan Algoritma K-Means,” vol. 02, 2022.
- [16] S. N. Br Sembiring, H. Winata, and S. Kusnasari, “Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means,” *j. sist. inf. trig. dhar. JURSI TGD*, vol. 1, no. 1, p. 31, Jan. 2022, doi: 10.53513/jursi.v1i1.4784.
- [17] A. Sulistiyawati and E. Supriyanto, “Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan,” *JTK*, vol. 15, no. 2, p. 25, Aug. 2021, doi: 10.33365/jtk.v15i2.1162.
- [18] A. Rohmah Zaidah, C. Indira Septiarani, M. Sholikhatun Nisa, A. Yusuf, and N. Wahyudi, “Komparasi Algoritma K-Means, *K-Medoids*, Agglomerative Clustering Terhadap Genre Spotify,” *JIKOM*, vol. 7, no. 1, pp. 49–54, Apr. 2021, doi: 10.35329/jiik.v7i1.186.
- [19] E. Tasia and M. Afdal, “Comparison Of K-Means And *K-Medoids* Algorithms for Clustering of Flood-Prone Areas in Rokan Hilir District,” 2023.
- [20] A. Idrus, N. Tarihoran, U. Supriatna, A. Tohir, S. Suwarni, and R. Rahim, “Distance Analysis Measuring for Clustering using K-Means and Davies Bouldin Index Algorithm,” *TEM Journal*, pp. 1871–1876, Nov. 2022, doi: 10.18421/TEM114-55.
- [21] Y. Arie Wijaya, D. Achmad Kurniady, E. Setyanto, W. Sanur Tarihoran, D. Rusmana, and R. Rahim, “Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities,” *TEM Journal*, pp. 1099–1103, Aug. 2021, doi: 10.18421/TEM103-13.
- [22] F. Ros, R. Riad, and S. Guillaume, “PDBI: A partitioning Davies-Bouldin index for clustering evaluation,” *Neurocomputing*, vol. 528, pp. 178–199, Apr. 2023, doi: 10.1016/j.neucom.2023.01.043.
- [23] Nanda Shalsadilla, Shantika Martha, and Hendra Perdana, “Penentuan Jumlah Cluster Optimum Menggunakan Davies Bouldin Index dalam Pengelompokan Wilayah Kemiskinan di Indonesia,” *statistika*, vol. 23, no. 1, pp. 63–72, Jun. 2023, doi: 10.29313/statistika.v23i1.1743.