

Perbandingan Algoritma K-Means dan Algoritma K-Medoid Untuk Pengelompokan UMKM di Kebumen

Comparison of the K-Means Algorithm and the K-Medoid Algorithm for Clustering UMKM in Kebumen

Retno Wahyusari^{*1}, Setia Wardani²

¹Program Studi Informatika, Sekolah Tinggi Teknologi Ronggolawe

²Program Studi Informatika, Universitas PGRI Yogyakarta

e-mail: ^{*1}retnowahyusari@gmail.com

Abstrak - Sektor Usaha Mikro Kecil dan Menengah (UMKM) memiliki peran penting sebab merupakan salah satu pendorong pertumbuhan ekonomi di Indonesia. Saat ini Kebumen memiliki total 36.851 jumlah UMKM, dari 26 kecamatan. Melihat banyaknya jumlah UMKM dan jumlah kecamatan, perlu adanya suatu pengelompokan kecamatan yang masih memiliki jumlah UMKM rendah. Dengan pengelompokan dapat membantu pihak terkait untuk fokus pada kecamatan dalam memberikan pendampingan dalam peningkatan jumlah UMKM. Penelitian yang membahas pengelompokan telah banyak dilakukan. Teknik *partition* merupakan salah satu teknik dalam pengelompokan. Algoritma K-Means dan algoritma K-Medoid merupakan teknik *partition* dalam pengelompokan. Algoritma K-Means merupakan salah satu algoritma pengelompokan yang paling kuat dan populer dalam penelitian. Algoritma K-Medoid merupakan algoritma yang sederhana dan efektif dalam melakukan pengelompokan. *Davies-Bouldin Index* (DBI) merupakan metode yang digunakan untuk mengukur validitas *cluster* pada metode pengelompokan. Hasil penelitian terbetuk 3 cluster, baik menggunakan algoritma K-Means dan algoritma K-Medoid. Algoritma K-Means *cluster* 1 sebanyak 21 kecamatan, *cluster* 2 sebanyak 3 kecamatan, dan *cluster* 3 sebanyak 2 kecamatan. Algoritma K-Medoid pada *cluster* 1 sebanyak 19 kecamatan, *cluster* 2 sebanyak 5 kecamatan dan *cluster* 3 sebanyak 2 kecamatan. Berdasarkan nilai DBI algoritma K-Means bernilai 0,324 dan algoritma K-Medoid sebesar 0,536. Maka dapat disimpulkan bahwa algoritma K-Means pada kasus pengelompokan UMKM Kebumen lebih bagus kinerjanya, sebab nilai DBI mendekati nilai 0. Dari hasil pengelompokan dapat membantu dalam memberikan gambaran bagi pihak terkait dalam mendorong atau memberikan pendampingan terhadap kecamatan yang masuk dalam cluster rendah jumlah UMKM nya.

Kata kunci – Cluster, DBI, K-Means, K-Medoid

Abstract - The Usaha Mikro Kecil dan Menengah (UMKM) sector has an important role because it is one of the drivers of economic growth in Indonesia. Currently Kebumen has a total of 36,851 UMKM, from 26 sub-districts. Seeing the large number of UMKM and the number of sub-districts, it is necessary to group sub-districts that still have a low number of UMKM. Grouping can help related parties to focus on sub-districts in providing assistance in increasing the number of UMKM. Research that discusses clustering has been carried out a lot. Partition technique is one technique in grouping. The K-Means algorithm and the K-Medoid algorithm are partition techniques in grouping. The K-Means algorithm is one of the most powerful and popular clustering

algorithms in research. The K-Medoid algorithm is a simple and effective algorithm for grouping. Davies-Bouldin Index (DBI) is a method used to measure the validity of clusters in the clustering method. The research results formed 3 clusters, using both the K-Means algorithm and the K-Medoid algorithm. The K-Means algorithm cluster 1 consists of 21 districts, cluster 2 consists of 3 districts, and cluster 3 consists of 2 districts. The K-Medoid algorithm in cluster 1 consists of 19 districts, cluster 2 consists of 5 districts and cluster 3 consists of 2 districts. Based on the DBI value, the K-Means algorithm is 0.324 and the K-Medoid algorithm is 0.536. So it can be concluded that the K-Means algorithm in the Kebumen UMKM clustering case has better performance, because the DBI value is close to 0. From the grouping results it can help provide an overview for related parties in encouraging or providing assistance to sub-districts that are included in clusters with a low number of UMKM.

Keywords – Cluster, DBI, K-Means, K-Medoid

I. PENDAHULUAN

Sektor Usaha mikro Kecil dan Menengah (UMKM) memiliki peran penting sebab merupakan salah satu pendorong pertumbuhan ekonomi di Indonesia [1][2][3]. Saat ini Kebumen memiliki total 36.851 jumlah UMKM, dari 26 kecamatan. Jenis produk UMKM terdiri dari 6 jenis produk yaitu makanan/minuman dan sembako; perdagangan; Peternakan, Pertanian, Pertambangan, Perkebunan, Perikanan; Konveksi dan Pakaian; jasa; Kerajinan dan Batik.

Melihat banyaknya jumlah UMKM dan jumlah kecamatan, perlu adanya suatu pengelompokan kecamatan yang masih memiliki jumlah UMKM rendah. Dengan pengelompokan dapat membantu pihak terkait untuk fokus pada kecamatan dalam memberikan pendampingan dalam peningkatan jumlah UMKM.

Penelitian yang membahas pengelompokan telah banyak dilakukan. Teknik *partition* merupakan salah satu teknik dalam pengelompokan. Algoritma K-Means dan algoritma K-Medoids merupakan teknik *partition* dalam pengelompokan [4]. Penelitian yang menggunakan Algoritma K-Means dalam mengelompokkan UMKM dilakukan oleh Puntoriza dan Charitas Fibriani [5], Wahyu Sudrajat, dkk [3], Erni Rouza, dkk [6] menghasilkan 3 kelompok UMKM. Algoritma K-Means merupakan salah satu algoritma pengelompokan yang paling kuat dan populer dalam penelitian [7][8].

Berdasarkan penelitian sebelumnya belum membandingkan Algoritma K-Means dengan Algoritma K-Medoid. Algoritma K-Medoids merupakan algoritma yang sederhana dan efektif dalam melakukan pengelompokan [9][10]. Maka pada penelitian ini menggunakan algoritma K-Means dan algoritma K-Medoid dan mengukur validitas kelompok hasil pengelompokan.

II. METODE

Metode penelitian yang dilakukan adalah menggunakan metode eksperimen dimana dalam penelitian menggunakan perlakuan pada atribut parameter dan melakukan tes adapun metodenya sebagai berikut:

2.1. Pengumpulan data

Data didapat dari pengumpulan data UMKM yang ada di 26 Kecamatan Kebumen (DIY). Total keseluruhan data UMKM adalah 36.851 dengan berbagai jenis produk (perdagangan; peternakan, pertanian, perikanan; konveksi dan pakaian; jasa; makanan/minuman dan sejenisnya; kerajinan dan batik).

2.2. Pengolahan Data

Data yang terkumpul kemudian dibuat tabel per kecamatan serta jenis produknya. Tabel yang terbentuk, nantinya diolah kedalam algoritma K-Means dan K-Medoid.

2.3. Metode yang Diajukan

1) Algoritma K-Means

Berikut ini adalah tahapan-tahapan untuk menghitung metode K-Means [11][12]:

1. Tentukan jumlah K (*Cluster*), nilai K harus lebih kecil dari jumlah data.
2. Pilih titik *Cluster* secara acak sebanyak K-buah, dimana titik ini akan menjadi pusat *Centroid* dari masing-masing kelompok (*Cluster*).
3. Hitung jarak dan alokasikan masing-masing data ke *Centroid* atau rata-rata terdekat, dengan rumus:

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

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

4. Tentukan *Centroid* baru, lokasi *centroid* (pusat kelompok) dihitung dengan rata-rata semua data yang bergabung dalam setiap kelompok. Kembali ke nomor 3 apa bila masih ada data yang berpindah *Cluster* atau ada perubahan nilai *Centroid*, jika tidak ada maka hentikan proses *clustering*nya.

2) Algoritma K-Medoid

1. Melakukan inisialisasi pusat cluster sebanyak jumlah cluster (k).
2. Mengalokasikan setiap objek ke cluster terdekat menggunakan Euclidean Distance
3. Memilih objek secara acak pada masing-masing cluster sebagai calon medoid yang baru.
4. Menghitung jarak setiap objek pada masing-masing cluster dengan calon medoid yang baru.
5. Total simpangan (S) didapatkan dengan menghitung nilai total jarak yang baru dikurangi total jarak yang lama. Jika nilai S kurang dari 0, maka dilakukan penukaran objek dengan data cluster untuk membentuk sekumpulan jumlah cluster objek yang baru sebagai medoid.
6. Mengulangi tahap c sampai dengan e apabila masih terjadi perubahan medoid, jika tidak ada perubahan maka diperoleh *cluster* serta anggota dari *cluster* masing-masing.

a Eksperimen dan Pengujian Model

Pada tahap eksperimen dan pengujian model dilakukan menggunakan tool Rapidminer. Pengelompokan dilakukan menggunakan Algoritma K-Means dan Algoritma K-Medoid. Hasil pengelompokan kemudian dilakukan pengukuran kinerja algoritma.

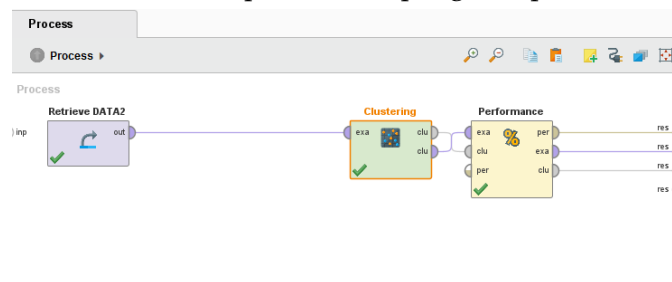
b Evaluasi dan Validasi Hasil

Tahap ini membahas *Davies-Bouldin Index* (DBI) metode yang digunakan untuk mengukur validitas *cluster* pada metode pengelompokan. Pengukuran dengan menggunakan DBI untuk memaksimalkan jarak antar *cluster* dan untuk meminimalkan jarak antar titik dalam *cluster*, jika jarak antar *cluster*

maksimal, berarti kesamaan karakteristik antar masing-masing *cluster* lebih sedikit sehingga perbedaan antar *cluster* terlihat lebih jelas. DBI yang diperoleh (non-negatif ≥ 0), maka semakin baik *cluster* yang diperoleh dari pengelompokan.

III. HASIL DAN PEMBAHASAN

Pada data UMKM Yogyakarta, terdapat 26 kecamatan dengan total jumlah UMKM 36.851. Jumlah UMKM yang tersebar di semua kecamatan dikelompokkan berdasarkan jenis produk yang nantinya data yang telah diolah dilakukan proses pengelompokan menggunakan algoritma K-Means dan algoritma K-Medoid. Proses pengelompokan menggunakan tools RapidMiner. Dimana langkah pertaman memasukkan data, kemudian menggunakan algoritma *clustreing* K-Means dengan menentukan cluster 3. Hasil pengelompokan dihitung performance menggunakan *Davies-Bouldin Index* (DBI). Tampilan proses pengelompokan terlihat pada Gambar 1. Gambar 2 merupakan hasil pengelompokan kecamatan.



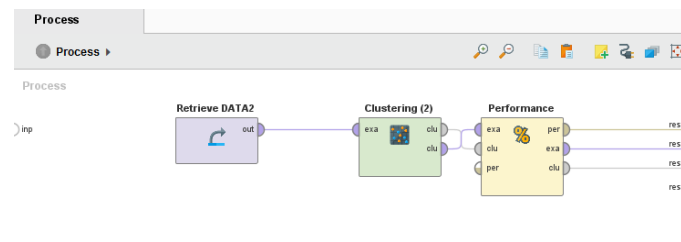
Gambar 1. Proses Cluster Algoritma K-Means

kabupaten	Cluster	Makanan/Minuman dan Sembako	Perdagangan	Pesiraman, Pertanian, Pertambangan, Perikanan	Konveksi dan Pakai	Jasa	Kerajinan dan Batik
KURONG	cluster_1	1504	65	544	67	477	345
KARANGSAMBUNG	cluster_0	145	0	50	21	50	15
SADANG	cluster_0	133	9	84	6	70	14
BONOROWO	cluster_0	386	7	178	10	122	26
SRIWENG	cluster_0	854	24	213	32	368	188
SEMPOR	cluster_0	654	61	117	55	136	46
ROWOKELE	cluster_0	613	37	207	42	159	100
PURING	cluster_0	893	25	394	42	233	48
PREMBUN	cluster_0	455	24	152	33	134	39
BULUSPESANTREN	cluster_1	1548	36	570	97	701	228
PONCOWARNO	cluster_0	221	10	427	9	172	26
PETANAHAH	cluster_0	330	4	119	9	86	100
PADURESO	cluster_0	252	7	88	6	77	25
MIRIT	cluster_0	375	10	120	19	95	19
KUWARASAN	cluster_0	533	5	121	19	177	22
KUTOWINANGUN	cluster_0	406	11	106	15	216	42
KEBUMEN	cluster_2	3564	264	626	520	1073	351
KARANGANYAR	cluster_0	359	4	48	15	124	72
GOMBONG	cluster_2	2857	169	350	125	800	126
BUAYAN	cluster_0	8	0	1	0	5	0
AYAH	cluster_0	440	11	101	48	77	113
AMBAL	cluster_0	497	31	136	36	201	139
ADIMULYO	cluster_0	690	62	447	29	262	132
PEJAGOAN	cluster_0	360	22	60	32	86	67
ALIAN	cluster_1	1421	46	509	175	628	181
KARANGGAYAM	cluster_0	114	16	52	16	30	18

Gambar 2. Data Kabupaten Berdasarkan Hasil Cluster K-Means

Berdasarkan hasil pengelompokan menggunakan algoritma K-Means menghasilkan 3 *cluster*. *Cluster* 1 merupakan kecamatan dengan jumlah UMKM sedikit, *cluster* 2 jumlah UMKM sedang, *cluster* 3 dengan jumlah UMKM banyak. *Cluster* 2 terdiri dari kecamatan Kurong, Bulus Pesantern, dan Alian. *Cluster* 3 terdiri dari Kebumen dan Gombong. Kecamatan yang lainnya masuk pada *cluster* 1.

Proses pengelompokan menggunakan algoritma K-Medoid sama dengan langkah pada pengelompokan K-Means, hanya merubah algoritma *clustering* nya. Gambar 3 merupakan gambar proses *clustering* menggunakan algoritma K-Medoid dan Gambar 4 merupakan hasil pengelompokan kecamatan berdasarkan *cluster*.



Gambar 3. Proses Cluster Algoritma K-Medoid

Kabupaten	Cluster	Makanan/Minuman dan Sembako	Perdagangan	Pertanian, Pertambangan, Perikanan	Konveksi dan Pakai	Jasa	Kerajinan dan Batik
KURONG	cluster_1	1504	65	544	67	477	345
KARANGSAMBUNG	cluster_0	145	0	50	21	50	15
SADANG	cluster_0	133	9	84	6	70	14
BONOROWO	cluster_0	386	7	178	10	122	26
SRUWENG	cluster_1	854	24	213	32	368	188
SEMPOR	cluster_0	654	61	117	55	136	46
ROWOKELE	cluster_0	613	37	207	42	159	100
PURING	cluster_1	833	25	394	42	233	48
PREMBUN	cluster_0	455	24	152	33	134	39
BULUSPESANTREN	cluster_1	1548	36	570	97	701	228
PONCOWARNO	cluster_0	221	10	427	9	172	26
PETANAHAN	cluster_0	330	4	119	9	86	100
PADURESO	cluster_0	252	7	88	6	77	25
MIRIT	cluster_0	375	10	120	19	95	19
KUWARASAN	cluster_0	533	5	121	19	177	22
KUTOWINANGUN	cluster_0	406	11	106	15	216	42
KEBUMEN	cluster_2	3564	264	626	520	1073	351
KARANGANYAR	cluster_0	359	4	48	15	124	72
GOMBONG	cluster_2	2857	169	350	125	800	126
BUAYAN	cluster_0	8	0	1	0	5	0
AYAH	cluster_0	440	11	101	48	77	113
AMBAL	cluster_0	497	31	136	36	201	139
ADIMULYO	cluster_0	690	62	447	29	262	132
PEJAGOAN	cluster_0	360	22	60	32	86	67
ALIAN	cluster_1	1421	46	509	175	628	181
KARANGGAYAM	cluster_0	114	16	52	16	30	18

Gambar 4. Data Kabupaten Berdasarkan Hasil Cluster K-Medoid6

Pengelompokan menggunakan algoritma K-Medoid menghasilkan Cluster 2 terdiri dari kecamatan Kurong, Puring, Sruwing, dan Bulus Pesantern, dan Alian. Cluster 3 terdiri dari Kebumen dan Gombang. Kecamatan yang lainnya masuk pada cluster 1.

Sedangkan berdasarkan kinerja pengelompokan nilai *Davies-Bouldin Index* (DBI) algoritma K-Means seneilai 0,324 dan algoritma K-Medoid sebesar 0,536. Tabel 1 menunjukkan nilai DBI tiap algoritma, terlihat bahwa terdapat selisih 0,212. Berdasarkan nilai DBI algoritma K-Means bekerja lebih bagus dibandingkan dengan algoritma K-Medoid dalam pengelompokan.

Tabel 1. Nilai DBI

Algoritma	Nilai DBI
K-Means	0,324
K-Medoid	0,536

IV. KESIMPULAN

Hasil penelitian terbetuk 3 cluster, baik menggunakan algoritma K-Means dan algoritma K-Medoid. Algoritma K-Means cluster 1 sebanyak 21 kecamatan, cluster 2 sebanyak 3 kecamatan, dan cluster 3 sebanyak 2 kecamatan. Algoritma K-Medoid pada cluster 1 sebanyak 19 kecamatan, cluster 2 sebanyak 5 kecamatan dan cluster 3 sebanyak 2 kecamatan. Berdadarkan nilai DBI algoritma K-Means bernilai 0,324 dan algoritma K-Medoid sebesar 0,536. Maka dapat disimpulkan bahwa algoritma K-Means pada kasus pengelompokan UMKM Kebumen lebih bagus kinerjanya, sebab nilai DBI mendekati nilai 0. Dari hasil pengelompokan berdasarkan hasil pengelompokan dapat membantu dalam memberikan gambaran bagi pihak terkait dalam mendorong atau memberikan pendampinhan terhadap kecamatan yang masuk dalam cluster rendah jumlah UMKM nya.

DAFTAR PUSTAKA

- [1] A. S. Sunge and H. Hermawanto, "Pengelompokan Industri Mikro Di Indonesia Dengan Metode K-Means Clustering," *SIGMA - J. Teknol. Pelita Bangsa*, vol. 10, no. 1, pp. 167–172, 2019.
- [2] T. Zulyanti and Noeryanti, "Perbandingan Pengelompokan Usaha Mikro Kecil Dan Menengah Di Kabupaten Klaten Tahun 2019 Dengan Metode K-Means Dan Clustering Large Application," *J. Stat. Ind. dan Komputasi*, vol. 7, no. 1, pp. 46–59, 2022.
- [3] W. Sudrajat, I. Cholid, and J. Petrus, "Penerapan Algoritma K-Means Untuk Pengelompokan UMKM Menggunakan Rapidminer," *JUPITER*, vol. 14, no. 1, pp. 27–36, 2022.
- [4] S. Nirmal, "Comparative study between k-means and k-medoids clustering algorithms," *Int. Res. J. Eng. Technol.*, vol. 839, pp. 839–844, 2019.
- [5] P. Puntoriza and C. Fibriani, "Analisis Persebaran UMKM Kota Malang Menggunakan Cluster K-means," *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 86–94, 2020, doi: 10.33633/joins.v5i1.3469.
- [6] E. Rouza, Basorudin, and Efrida, "Identifikasi dan Klasifikasi UMKM di Kabupaten Rokan Hulu Menggunakan Metode K-Means," *J. Ilm. Univ. Pengaraian*, vol. 7, no. 01, pp. 32–40, 2021.
- [7] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electron.*, vol. 9, no. 8, pp. 1–12, 2020, doi: 10.3390/electronics9081295.
- [8] 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.
- [9] N. Sureja, B. Chawda, and A. Vasant, "An improved K-medoids clustering approach based on the crow search algorithm," *J. Comput. Math. Data Sci.*, vol. 3, no. March, p. 100034, 2022, doi: 10.1016/j.jcmds.2022.100034.
- [10] Q. Zhang and I. Couloigner, "A new and efficient K-medoid algorithm for spatial clustering," *Lect. Notes Comput. Sci.*, vol. 3482, no. III, pp. 181–189, 2005, doi: 10.1007/11424857_20.
- [11] S. Saifulloh and N. I. F. Nisa, "Penerapan Metode K-Means dan Fuzzy Sugeno dalam Pemetaan Tingkat Produksi Masker Kain di Masa Pandemi," *Expert J. Manaj. Sist. Inf. dan Teknol.*, vol. 11, no. 1, p. 01, 2021, doi: 10.36448/expert.v11i1.1972.
- [12] D. T. Larose, *Discovering Knowledge In Data: an introduction to data mining. In Automotive Industries AI*. Canada: A John Wiley & Sons, Inc, 2005.