

Analisis Data Mining Pada Klasterisasi UMKM Dengan Menggunakan Algoritma K-Means

Budi Arifitama
Universitas Trilogi
Jl.TMP Kalibata No.1
budiarif@trilogi.ac.id

Ade Syahputra
Universitas Trilogi
Jl.TMP Kalibata No 1
adesyahputra@trilogi.ac.id

Penulis Korespondensi : Budi Arifitama

Abstrak— UMKM merupakan usaha mikro kecil menengah yang dimiliki oleh seseorang atau perorangan. UMKM sangat berpengaruh dalam sektor ekonomi Indonesia yaitu salah satu yang paling berguna adalah memberikan lowongan kerja dengan modal yang minim. Peran masyarakat dalam sektor pembangunan, khususnya dalam pembangunan ekonomi, adalah Usaha Mikro, Kecil dan Menengah (UMKM). Posisi UMKM dalam perekonomian nasional memiliki peran penting dan strategis. Kondisi ini sangat dimungkinkan karena keberadaan UMKM cukup dominan dalam perekonomian Indonesia. Usaha mikro kecil menengah memiliki beberapa kriteria seperti berapa jumlah pendapatan maksimal dari usaha mikro, usaha kecil, dan usaha menengah. Dalam pemodelan clustering data mining dengan metode K-means yang kami gunakan, kami akan mengklasifikasikan yang mana usaha mikro, kecil, dan menengah. Pengelompokan UMKM memiliki sangat banyak manfaat, contohnya adalah ketika jika kita sudah mendapatkan data UMKMnya, kita dapat meningkatkan kualitas UMKM rendah agar dapat berkembang lebih lagi. Dari hasil perhitungan menggunakan rapidminer untuk menghasilkan data.

Kata Kunci— UMKM, K-Means, Data Mining, Clustering, Rapidminer.

Abstract— Small and Medium Enterprise (SME) owned by a person or individual. MSMEs are very influential in the Indonesian economic sector, one of the most useful is to provide job vacancies with minimal capital. The role of the community in the development sector, especially in economic development, is Micro, Small and Medium Enterprises (MSMEs). The position of MSMEs in the national economy has an important and strategic role. This condition is very possible because the existence of MSMEs is quite dominant in the Indonesian economy. Micro, small and medium enterprises have several criteria such as what is the maximum amount of income from micro, small and medium enterprises. In modeling clustering data mining with the K-means method that we use, we will classify which are micro, small, and medium enterprises. The grouping of MSMEs has many benefits, for example, when we have obtained the MSME data, we can improve the quality of low MSMEs so that they can develop even more. From the results of calculations using rapidminer to generate data.

Keywords— UMKM, K-Means, Data Mining, Clustering, Rapidminer



I. PENDAHULUAN

Pembangunan merupakan salah satu perkembangan yang dilakukan oleh masyarakat untuk masa yang akan datang. UMKM atau usaha mikro kecil menengah adalah usaha ekonomi produktif yang dimiliki oleh seseorang sesuai dengan kriteria-kriteria yang sudah ditetapkan di undang-undang, tepatnya UU No. 20 tahun 2008 (Menengah et al., 2017). Peran masyarakat dalam pembangunan nasional utamanya dalam pembangunan ekonomi adalah Usaha Mikro, Kecil, dan Menengah. Posisi Usaha Mikro, Kecil, dan Menengah (UMKM) dalam perekonomian nasional memiliki peran yang penting dan berkontribusi banyak (Budiman & Herkulana, 2021),(et al., 2019).

Data berdasarkan dari Badan Pusat Statistik merilis keadaan tersebut pasca krisis ekonomi jumlah UMKM tidak berkurang, malah jumlah UMKM meningkat pertumbuhannya terus, bahkan mampu menyerap 85 juta hingga 107 juta tenaga kerja sampai tahun 2012. Pada tahun itu jumlah pengusaha di Indonesia sebanyak 56.539.560 unit. Dari jumlah yang disebutkan, UMKM sebanyak 56.534.592 unit atau sebesar 99,99%. Sisanya sekitar 0,01% atau sebesar 4.968 unit adalah Usaha berskala besar.

Berdasarkan undang-undang No. 20 tahun 2008. Sebuah UMKM ditentukan berdasarkan beberapa kriteria yaitu, Usaha Mikro adalah usaha yang memiliki hasil bersih dari aset paling tinggi Rp.50.000.000 (Lima puluh juta rupiah) dan omset paling tinggi Rp.300.000.000(Tiga ratus juta rupiah). Usaha Kecil memiliki aset paling tinggi hingga Rp.500.000.000(Lima ratus juta rupiah) dan hasil penjualan tahunan Rp. 300.000.000 (tiga ratus juta rupiah) hingga sampai dengan Rp.2.500.000.000(Dua Milyar Lima Ratus juta rupiah) sedangkan usaha menengah memiliki hasil aset paling tinggi Rp.10.000.000.000 (Sepuluh Milyar rupiah) dan penjualan tahunan paling banyak Rp.50.000.000.000,00 (lima puluh milyar rupiah).

Pembagian cluster antara UMKM dapat dikembangkan dengan rapidminer yang bertujuan agar pembagian untuk mengetahui masing-masing keberadaan dari setiap pelaku UMKM dan bertujuan untuk pengembangan dan peningkatan UMKM. pengembangan yang dilakukan agar setiap UMKM mengalami peningkatan yang seimbang serta dapat mencegah terjadinya kesalahan dalam pemerataan dan kebijakan sasaran UMKM.

Penelitian sebelumnya telah dilakukan oleh (Mustapita & Slamet, 2020), yang mengumpulkan jenis atau pengelompokan di kota malang, kemudian ada magdalena yang melakukan penelitian di jawa barat dengan menggunakan metode algoritma K-means(Magdalena & Fahrudin, 2020) dan Penelitian terakhir dilakukan oleh wahyu sudrajat dengan menggunakan K-means juga.

II. METODE

2.1 Kerangka Penelitian

Kerangka Kerja yang ditujukan untuk membagi alur dari berbagai tahap yang akan dilakukan dalam melakukan penelitian.

Berikut penjelasan dari kerangka Penelitian:

- a) Identifikasi Masalah
Tidak tersedianya data secara umum dan informasi yang detail mengenai pengelompokan data dari UMKM
- b) Pendekatan Sosial dan Identifikasi Masalah
menggunakan data yang ada dari UMKM yang tersedia dan Clustering menggunakan Algoritma K-Means
- c) Pengolahan data
Mengimplementasikan hasil dari data yang diperoleh ke dalam RapidMiner
- d) Pengujian data
Menguji data dari setiap object yang telah tersedia atau yang telah ditemukan
- e) hasil
hasil dari pengelompokan data UMKM sesuai dengan kelas-kelas yang telah ditentukan menggunakan k-means.

Data relevan untuk mengklasifikasi UMKM dalam menentukan cluster yang akan dicapai.

Berikut adalah data UMKM berdasarkan omset dan aset:

2.2 K-means clustering

K-means (Clustering Algorithms) adalah algoritma clustering(Darwis et al., 2021) paling umum dan paling sederhana. Algoritma clustering ini membagi n pengamatan menjadi menjadi kelompok kedalam kumpulan data(Janßen & Wan, 2020),(Abadi et al., 2018).

NO	Daftar UMKM	Omset	Asset
1.	Warung Bunda Tuti	Rp. 720,000,000	Rp. 100,000,000
2.	Warung Dua Bersaudara	Rp. 360,000,000	Rp. 500,000,000
3.	Warung Tirto	Rp. 120,000,000	Rp. 30,000,000
4.	Toko Zidan	Rp. 1,440,000,000	Rp. 100,000,000
5.	Toko Tulis Nazwa	Rp. 180,000,000	Rp. 75,000,000

		00	0
6.	Bengkel Harry	Rp. 120,000,000	Rp. 100,000,000
7.	Toko Grosir Hj Mamat	Rp. 2,540,000,000	Rp. 45,000,000
8.	Warung Bambang	Rp. 388,000,000	Rp. 50,000,000
9.	Toko Baju Tita	Rp. 54,000,000	Rp. 50,000,000
10.	TK. Tasiya	Rp. 1,800,000,000	Rp. 100,000,000
11.	H. Sidi Zainudin	Rp. 240,000,000	Rp. 400,000,000
12.	Bambang	Rp. 1,350,000,000	Rp. 50,000,000
13.	Aida	Rp. 1,080,000,000	Rp. 100,000,000
14.	M. Yusuf / Bakso Kepala Sapi	Rp. 60,300,000	Rp. 40,000,000
15.	Nin / Yoga Moobilindo	Rp. 6,500,000,000	Rp. 1,000,000,000

Tujuan dari clustering algoritma ini adalah untuk menemukan data prototipe pada setiap cluster, kemudian dikumpulkan dan membentuk sebuah cluster.

Cara kerja dari algoritma K-Means Clustering (Suharjo & Utama, 2021),(Kumar & Verma, 2018),(Ghazal et al., 2021) sebagai berikut:

- Menentukan Jumlah Centroid atau Menentukan nilai pusat (centroid)
- Tetapkan point data, Pengukuran jarak Euclidean adalah ukuran kedekatan yang paling umum, meskipun ukuran lain seperti ukuran Manhattan dan koefisien Jaccard dapat digunakan. Jarak Euclidean antara dua titik data X (x1, x2,...xn) dan C (c1, c2, ..., cn) dengan n atribut

deucliden

$$(X,Y)=\sqrt{\sum_i^n =1(Xi - Yi).....(1)}$$

ket: X dan Y representasi dari dua record

- Menghitung Centroid baru.

$$c=\sum nm$$

ket: Centroid

- Ulangi tugas dan hitung centroid baru sampai titik data dipindahkan ke Centroid baru
- Penghentian

III. HASIL DAN PEMBAHASAN

3.1 Pengolahan Data dan Informasi
 Data UMKM yang didapatkan digunakan untuk perhitungan algoritma K-Means

No	Nama UMKM	Transformasi Omset	Transformasi aset
1.	Warung Bunda Tuti	2	2
2.	Warung Dua Bersaudara	2	2
3.	Warung Tirto	1	1
4.	Toko Zidan	2	2
5.	Toko Tulis Nazwa	1	2
6.	Bengkel Harry	1	2
7.	Toko Grosir Hj Mamat	3	1
8.	Warung Bambang	2	1
9.	Toko Baju Tita	1	1
10.	TK. Tasiya	2	2
11.	H. Sidi Zainudin	1	2
12.	Bambang	2	1
13.	Aida	2	2



14.	M. Yusuf / Bakso Kepala Sapi	1	2
-----	------------------------------------	---	---

15.	Nin / Yoga Moobilindo	3	3
-----	--------------------------	---	---

3.2 Perhitungan dengan Algoritma K-Means data UMKM

a) Menentukan Jumlah Cluster

Jumlah Cluster adalah jumlah kelompok yang dihasilkan. disesuaikan dengan kebutuhan analisis dalam hal ini terdiri dari 3 cluster yaitu Micro, Kecil dan Menengah.

b) Menghitung Titik Pusat Centroid secara Acak

Centroid ditentukan secara acak, centroid mengacu pada cluster yang akan didapatkan. cluster pertama adalah : $m_1 = b(2,2)$, $m_2 = f(1,2)$ dan $m_3 = o(3,3)$

c) Menghitung distance space ke titik centroid

Hasil Perhitungan dan Pengelompokan Data Pada Iterasi Pertama

NO	Nama UMKM	TRANSFORMASI OMSET	TRANSFORMASI ASET	JARAKA KE CENTROID 1	JARAKA KE CENTROID 2	JARAKA KE CENTROID 3	CLUSTER
1	Warung Bunda Tuti	2	2	0	1	2	1
2	Warung Dua Bersaudara	2	2	0	1	2	1
3	Warung Tirto	1	1	2	1	6	2
4	Toko Zidan	2	2	0	1	2	1
5	Toko Tulis Nazwa	1	2	1	0	3	2
6	Bengkel Harry	1	2	1	0	3	2
7	Toko Grosir Hj Mamat	3	1	2	3	4	1
8	Warung Bambang	2	1	1	2	5	1
9	Toko Baju Tita	1	1	2	1	6	2
10	TK. Tasiya	2	2	0	1	2	1
11	H. Sidi Zainudin	1	2	1	0	3	2
12	Bambang	2	1	1	2	5	1
13	Aida	2	2	0	1	2	1
14	M. Yusuf / Bakso Kepala Sapi	1	2	1	0	3	2



15	Nin / Yoga Moobilindo	3	3	2	3	0	3
----	-----------------------	---	---	---	---	---	---

Hasil Perhitungan dan Pengelompokan Data Pada Iterasi Kedua

NO	NAMA UMKM	TRANSFORMASI OMSET	TRANSFORMASI ASET	JARAKA KE CENTROID 1	JARAKA KE CENTROID 2	JARAKA KE CENTROID 3	CLUSTER
1	Warung Bunda Tuti	2	2	0.265625	1.111111	2	1
2	Warung Dua Bersaudara	2	2	0.265625	1.111111	2	1
3	Warung Tirto	1	1	1.515625	0.444444	6	2
4	Toko Zidan	2	2	0.265625	1.111111	2	1
5	Toko Tulis Nazwa	1	2	1.265625	0.111111	3	2
6	Bengkel Harry	1	2	1.265625	0.111111	3	2
7	Toko Grosir Hj Mamat	3	1	1.265625	2.444444	4	1
8	Warung Bambang	2	1	0.515625	1.444444	5	1
9	Toko Baju Tita	1	1	1.515625	0.444444	6	2
10	TK. Tasiya	2	2	0.265625	1.111111	2	1
11	H. Sidi Zainudin	1	2	1.265625	0.111111	3	2
12	Bambang	2	1	0.515625	1.444444	5	1
13	Aida	2	2	0.265625	1.111111	2	1
14	M. Yusuf / Bakso Kepala Sapi	1	2	1.265625	0.111111	3	2
15	Nin / Yoga Moobilindo	3	3	2.765625	3.777778	0	3

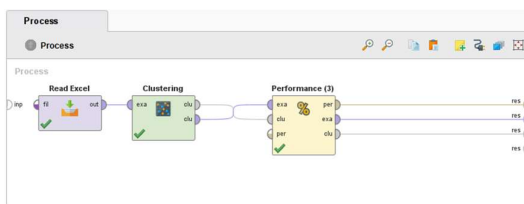


Berikut adalah hasil dari pengelompokan data UMKM dengan menggunakan K-Means Clustering adalah sebagai berikut :

- Cluster 1 : Warung Bunda Tuti, Warung Dua Bersaudara, Toko Zidan, Toko Grosir Hj Mamat, Warung Bambang, TK. Tasiya, Bambang, dan Aida.
- Cluster 2 : Warung Tirto, Toko Tulis Nazwa, Bengkel Harry, Toko Baju Tirta, H. Sidi Zainudin, dan M. Yusuf/Bakso Kepala Sapi.
- Cluster 3 : Nin/Yoga Mobilindo.

3.3 Pengujian dengan rapidminer

Perhitungan yang dilakukan harus dilakukan dengan Rapidminer untuk memastikan hasil yang sesuai. dibawah ini merupakan hasil pengelompokan berdasarkan algoritma K_Means di rapidminer :



Proses Clustering pada Rapid Miner

Row No.	Nama UMKM	cluster	Transforma...	Transforma...
1	Warung Bund...	cluster_2	2	2
2	Warung Dua ...	cluster_2	2	2
3	Warung Tirto	cluster_1	1	1
4	Toko Zidan	cluster_2	2	2
5	Toko Tulis N...	cluster_1	1	2
6	Bengkel Harry	cluster_1	1	2
7	Toko Grosir ...	cluster_2	3	1
8	Warung Bam...	cluster_2	2	1
9	Toko Baju Tita	cluster_1	1	1
10	TK Tasiya	cluster_2	2	2
11	H. Sidi Zainu...	cluster_1	1	2
12	Bambang	cluster_2	2	1
13	Aida	cluster_2	2	2
14	M. Yusuf / Ba...	cluster_1	1	2
15	Nin / Yoga Mo...	cluster_0	3	3

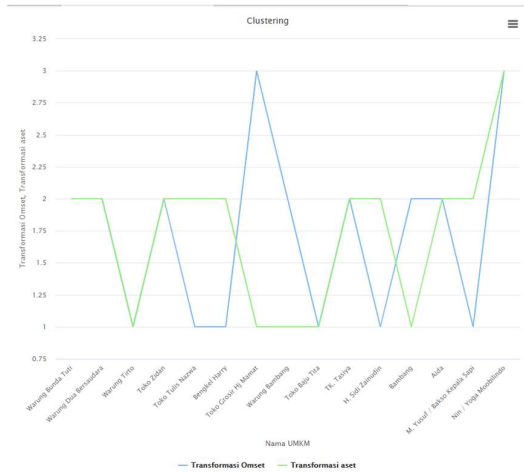
Data UMKM di Rapid Miner

The screenshot shows a window titled 'Cluster Model (Clustering)'. Inside, it displays the following information:
 Cluster 0: 1 items
 Cluster 1: 6 items
 Cluster 2: 8 items
 Total number of items: 15

Cluster Model di Rapid Miner

The screenshot shows the project tree in Rapid Miner. The root folder contains three sub-folders: cluster_0, cluster_1, and cluster_2. Each folder contains a list of data files corresponding to the items in that cluster:
 - cluster_0: Nin / Yoga Moobilindo
 - cluster_1: Warung Tirto, Toko Tulis Nazwa, Bengkel Harry, Toko Baju Tita, H. Sidi Zainudin, M. Yusuf / Bakso Kepala Sapi
 - cluster_2: Warung Bunda Tuti, Warung Dua Bersaudara, Toko Zidan, Toko Grosir Hj Mamat, Warung Bambang, TK. Tasiya, Bambang, Aida

Pengelompokan Data di Rapid Miner



Visualisasi Cluster UMKM di Rapid Miner,

IV. KESIMPULAN

Berdasarkan hasil yang telah dicapai pengujian data UMKM menggunakan rapidminer dan Dengan menggunakan teknik clustering K-means, kita dapat mengelompokan data prototipe pada setiap cluster. Setelah itu kita bisa mendapatkan hasil seperti yang tertera pada tabel diatas. Kami menggunakan aplikasi Rapidminer untuk menggunakan metode K-means. Penelitian ini memiliki kekurangan dalam mengetahui data detail dari pelaku UMKM, sehingga harus dikembangkan lagi agar sesuai dengan hasil yang diharapkan dan dapat menunjukkan hasil yang lebih detail.

V. DAFTAR PUSTAKA

Abadi, S., Mat The, K. S., Nasir, B. M., Huda, M., Ivanova, N. L., Sari, T. I., Maselena, A., Satria, F., & Muslihudin, M. (2018). Application model of k-means clustering: Insights into promotion strategy of vocational high school. *International Journal of Engineering and Technology(UAE)*.
<https://doi.org/10.14419/ijet.v7i2.11491>

Budiman, J., & Herkulana, H. (2021). PERAN UMKM DALAM PENYERAPAN TENAGA KERJA DI KOTA PONTIANAK. *Jurnal Ekonomi Integra*.
<https://doi.org/10.51195/iga.v11i2.164>

Darwis, M., Hasibuan, L. H., Firmansyah, M., Ahady, N., & Tiaharyadini, R. (2021). Implementation

of K-Means clustering algorithm in mapping the groups of graduated or dropped-out students in the Management Department of the National University. *JISA(Jurnal Informatika Dan Sains)*.
<https://doi.org/10.31326/jisa.v4i1.848>

Ghazal, T. M., Hussain, M. Z., Said, R. A., Nadeem, A., Hasan, M. K., Ahmad, M., Khan, M. A., & Naseem, M. T. (2021). Performances of k-means clustering algorithm with different distance metrics. *Intelligent Automation and Soft Computing*.
<https://doi.org/10.32604/iasc.2021.019067>

Janßen, A., & Wan, P. (2020). k-means clustering of extremes. *Electronic Journal of Statistics*.
<https://doi.org/10.1214/20-ejs1689>

Kumar, M., & Verma, A. (2018). Clustering Techniques - A Review. *International Journal of Computer Sciences and Engineering*, 6(6), 1091–1099.
<https://doi.org/10.26438/ijcse/v6i6.10911099>

Magdalena, L., & Fahrudin, R. (2020). Penerapan Data Mining Untuk Koperasi Se-Jawa Barat Menggunakan Metode Clustering pada Kementerian Koperasi dan UKM. *Jurnal Digit*.
<https://doi.org/10.51920/jd.v9i2.120>

Menengah, D., Situs, M., Indriasari, A., Suryanti, N., & Afriana, A. (2017). UNDANG-UNDANG NOMOR 20 TAHUN 2008 TENTANG USAHA MIKRO, KECIL, DAN MENENGAH. *ACTA DIURNAL Jurnal Ilmu Hukum Kenotariatan*.

Mustapita, A. F., & Slamet, A. R. (2020). PENGEMBANGAN EKONOMI LOKAL KABUPATEN MALANG MELALUI KAJIAN POTENSI KLASTER INDUSTRI KECIL. *Buletin Studi Ekonomi*.
<https://doi.org/10.24843/bse.2020.v25.i02.p07>

Sarfiah, S., Atmaja, H., & Verawati, D. (2019). UMKM Sebagai Pilar Membangun Ekonomi Bangsa. *Jurnal REP (Riset Ekonomi Pembangunan)*.
<https://doi.org/10.31002/rep.v4i2.1952>

Suharjo, B., & Utama, M. S. Y. (2021). K-Means Cluster Analysis of Sex, Age, and Comorbidities in the Mortalities of Covid-19 Patients of Indonesian Navy Personnel. *JISA(Jurnal Informatika Dan Sains)*.
<https://doi.org/10.31326/jisa.v4i1.869>