

Penerapan Algoritma K-Means Clustering untuk Pengelompokan UMKM Menggunakan Rapidminer

Wahyu Sudrajat¹⁾, Idahm Cholid²⁾, Johannes Petrus³⁾

¹Program Studi Manajemen Informatika, Universitas Multi Data Palembang,

²Program Studi Manajemen, Universitas Multi Data Palembang,

³Program Studi Informatika, Universitas Multi Data Palembang,

Jl. Rajawali No. 14 Palembang, Sumatera Selatan 30113

e-mail: ¹wahyu.sudrajat@mdp.ac.id, ²idham@mdp.ac.id, ³johannes@mdp.ac.id

Abstrak

Pendampingan UMKM secara tepat diharapkan dapat meningkatkan kualitas UMKM, namun demikian untuk memberikan pendampingan secara tepat kepada setiap UMKM di perlukan identifikasi dan pengelompokan usaha UMKM tersebut. Pengelompokan usaha di Indonesia dibagi dalam tiga kelompok, yaitu Mikro, Kecil dan Menengah, dimana setiap kelompok didasarkan atas omset dan aset. Tujuan dari penelitian ini adalah pengelompokan UMKM yang ada di Kabupaten di Sumatera Selatan dengan menerapkan algoritma K-Means Clustering. Tahapan penelitian yang dilakukan diantaranya adalah business understanding phase, data understanding phase, data processing phase, Modeling Phase, evaluation phase dan desimination phase. Pada pengujian 15 data usaha, penelitian ini berhasil menerapkan algoritma K-Means Clustering untuk mengidentifikasi dan mengklasifikasi UMKM dengan hasil pengujian sebesar 53% data ke cluster 1 sebanyak 8 data, 40% data ke cluster 2 sebanyak 6 data dan 7% data ke cluster 3 sebanyak 1 data. Hasil perhitungan ini juga telah diuji coba menggunakan software rapidminer dan menghasilkan data yang sama.

Kata kunci— K-Means, UMKM, data Mining, Clustering

Abstract

SMEs assistance appropriately is expected to improve the quality of SMEs, however, to provide proper assistance to each SMEs requires identification and grouping of SMEs businesses. The grouping of businesses in Indonesia is divided into three groups, namely Micro, Small and Medium, where each group is based on turnover and assets. The purpose of this research is to group SMEs in districts in South Sumatra by applying the K-Means Clustering algorithm. The stages of the research carried out include the business understanding phase, the data understanding phase, the data processing phase, the Modeling Phase, the evaluation phase and the desimination phase. In testing 15 business data, this study succeeded in applying the K-Means Clustering algorithm to identify and classify SMEs with test results of 53% of data to cluster 1 of 8 data, 40% of data to cluster 2 of 6 data and 7% of data to cluster 3 as much as 1 data. The results of this calculation have also been tested using rapidminer software and produce the same data.

Keywords— K-Means, SMEs, data Mining, Clustering

1. PENDAHULUAN

Salah satu faktor pendorong pertumbuhan ekonomi Indonesia adalah sektor Usaha Mikro Kecil dan Menengah (UMKM). Saat ini terdapat sekitar 64,2 juta unit usaha yang tergolong dalam UMKM

dan mampu berkontribusi terhadap Produk Domestik Bruto (PDB) sebesar 61,7% yang setara Rp. 8.573,89 trilyun. Selain itu, UMKM juga mampu menyerap tenaga kerja sekitar 97% dari total tenaga kerja[1]. Namun, sejalan dengan perkembangan teknologi dan persaingan usaha, pengembangan UMKM memiliki tantangan lebih besar lagi. Dalam menganalisa daya saing UMKM, maka perlu diperhatikan beberapa hal, diantaranya adalah: tingkat produktifitas dan inovasi pelaku usaha, kemudahan berusaha, akses permodalan, akses pemasaran, dukungan infrastruktur dan siklus bisnis[2]. Hal ini tentunya memerlukan peran semua pihak dan tidak tergantung hanya pada pemerintah semata saja, namun semua elemen masyarakat perlu mengambil peran dalam peningkatan daya saing UMKM ini.

Banyak kebijakan, bantuan, pelatihan kegiatan lainnya telah diarahkan untuk mendukung pengembangan dan peningkatan daya saing dan kapasitas UMKM. Namun terkadang kegiatan ini masih belum optimal antara lain dikarenakan redansi kegiatan, tidak tepat sasaran program, masih belum berkesinambungan program sampai dengan kegiatan yang terkadang kurang efektif. Salah satu faktor yang menyebabkan hal ini adalah belum terdatanya dengan baik serta belum adanya pemetaan yang komprehensif terhadap UMKM yang ada. Selain itu, database yang mumpuni terhadap kegiatan dan kebijakan kepada sasaran UMKM juga belum ada. Untuk itulah perlu dilakukan pembuatan kelompok-kelompok UMKM berdasarkan karakteristik yang dimiliki oleh masing-masing unit usaha. Diharapkan dengan pengelompokan (clusterisasi) ini, maka pengambil kebijakan dalam hal ini pemerintah atau lembaga-lembaga yang kompeten dapat memetakan karakteristik ataupun kebutuhan UMKM. Informasi pengelompokan UMKM ini menjadi salah satu informasi yang sangat penting dan solusi untuk mendapatkan informasi ini adalah dengan menggunakan konsep data mining.

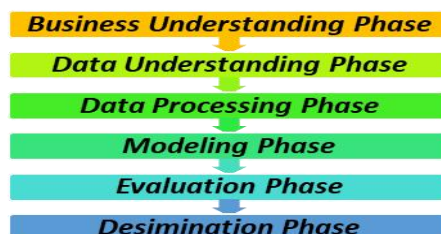
Data mining adalah analisis kumpulan data observasional untuk menemukan hubungan yang tidak terduga dan untuk meringkas data dengan cara baru yang dapat dimengerti dan berguna bagi pemilik data[3]. Data mining dikelompokkan berdasarkan tugas yang dilakukan, salah satunya adalah clustering. Clustering merupakan proses mempartisi sekumpulan objek data menjadi subset. Subset adalah cluster, sehingga objek dalam cluster sama satu sama lain, namun berbeda dengan objek di cluster lain[4]. Salah satu algoritma dalam pengelompokan data adalah Algoritma K-Means. Clustering berbeda dari klasifikasi karena tidak ada variable target untuk clustering[5]. K-Means merupakan salah satu algoritma clustering tertua dan paling banyak digunakan. Penggunaan algoritma K-Means dalam pengelompokan UMKM telah dilakukan oleh beberapa peneliti, diantaranya adalah puntoriza. Dimana dalam penelitiannya pengelompokan UMKM didasarkan oleh jenis UMKM yang ada di 5 (lima) kecamatan dikota Malang. Perhitungan algoritma K-Means yang dilakukan menghasilkan 3 cluster[6]. Sedangkan pada penelitian yang dilakukan oleh Magdalena, penggunaan algoritma K-Means digunakan untuk pengelompokan koperasi di Jawa barat. Pengelompokan yang dilakukan didasarkan atas aspek modal sendiri, modal luar dan volume usaha, dimana hasil perhitungan yang diperoleh adalah 3 (tiga) cluster yaitu cluster tinggi, sedang dan rendah[7].

Pada penelitian yang dilakukan pengelompokan UMKM dilihat dari aspek asset dan omset yang dimiliki oleh UMKM. Hal ini dipilih mengacu pada undang-undang RI Nomor 20 Tahun 2008 tentang Usaha Mikro, Kecil, dan Menengah[8]. Hasil dari perhitungan algoritma K-Means Clustering dapat dijadikan inputan bagi para pemangku kepentingan terhadap pendampingan, pengembangan dan pengelolaan UMKM yang ada, seperti Dinas Koperasi dan UMKM atau instansi yang lain yang berkenaan dengan hal ini. Dengan harapan UMKM yang ada dapat berkembang lebih baik.

2. METODE PENELITIAN

2.1 Kerangka Kerja Penelitian

Kerangka kerja penelitian merupakan gambaran proses yang sistematis, sehingga tujuan dari penelitian dapat dicapai dan berjalan dengan baik. Kerangka kerja dalam penelitian ini ditunjukkan pada gambar 1.



Gambar 1. Framework

Berikut ini adalah penjelasan dari masing-masing langkah dari kerangka kerja penelitian, yaitu:

a. Business understanding phase

Langkah pertama dan terpenting dalam data mining adalah pemahaman bisnis, yaitu mengajukan pertanyaan bisnis yang tepat. Pada tahapan ini perlu didefinisikan secara jelas untuk kebutuhan manajerial akan pengetahuan baru. Pada penelitian ini pemahaman bisnis yang ingin dicapai adalah mengetahui klasifikasi atau kelompok UMKM, sehingga pemangku kepentingan dapat melakukan proses pendampingan yang tepat. Hal ini penting dilakukan karena proses pendampingan dapat diberikan kepada UMKM secara tepat ketika sudah diketahui bahwa UMKM tersebut ada di kelompok mana.

b. Data understanding phase

Setelah diketahui pemahaman bisnis atau kebutuhan bisnis oleh para pemangku kepentingan, langkah selanjutnya adalah menentukan data yang relevan dan sesuai dari data yang ada. Penelitian ini digunakan untuk mengklasifikasikan UMKM, dengan demikian data yang akan digunakan tentu saja data-data UMKM. Tabel 1. Merupakan data UMKM yang akan digunakan dalam pengklasifikasian, dimana terdapat atribut nama UMKM, modal dan asset. Data UMKM yang digunakan merupakan data sekunder, yang berjumlah 15 *record*.

Tabel 1 Data UMKM

No.	Nama UMKM	Omset	Asset
a	Warung Tati	Rp. 720,000,000	Rp. 100,000,000
b	Toko Zhi-Zhi	Rp. 360,000,000	Rp. 500,000,000
c	Bengkel Las Kemang Tanjung	Rp. 120,000,000	Rp. 30,000,000
d	Toko Tasya	Rp. 1,440,000,000	Rp. 100,000,000
e	Fotokopi Naura	Rp. 180,000,000	Rp. 75,000,000
f	Bengkel Las Sony	Rp. 120,000,000	Rp. 100,000,000
g	Toko Grosir 3 Saudara	Rp. 1,440,000,000	Rp. 45,000,000
h	warung wahyu	Rp. 288,000,000	Rp. 160,000,000
i	Toko Baju Yossy YM	Rp. 54,000,000	Rp. 50,000,000
j	Rudi / Tk. Tasiya	Rp. 1,800,000,000	Rp. 100,000,000
k	H. Sidi Zainudin	Rp. 240,000,000	Rp. 400,000,000
l	Bambang	Rp. 1,350,000,000	Rp. 50,000,000

m	Aida	Rp. 1,080,000,000	Rp. 100,000,000
n	M. Yusuf / Bakso Kepala Sapi	Rp. 60,300,000	Rp. 40,000,000
o	Nin / Yoga Mobilindo	Rp. 6,500,000,000	Rp. 1,000,000,000

c. Data processing phase

Tujuan dari data preprocessing adalah untuk mengambil data yang diidentifikasi pada langkah sebelumnya dan mempersiapkannya menjadi data yang dapat diolah menggunakan algoritma data mining yang telah ditentukan. Data UMKM yang telah ditentukan selanjutnya ditransformasi kedalam bentuk numerik sehingga dapat diolah dengan algoritma data mining. Tabel 2 adalah kriteria transformasi modal UMKM dan Tabel 3 untuk kriteria transformasi asset UMKM. Sedangkan tabel 4 merupakan hasil dari transformasi data UMKM.

Tabel 2 Kriteria Transformasi data Jumlah Modal UMKM

Kriteria Transformasi Jumlah Modal UMKM	
Jumlah Omset	Transformasi
Maksimal Rp. 300 Juta	1
> Rp 300 Juta – Rp 2,5 Milyar	2
> Rp 2,5 Milyar – Rp 50 Milyar	3

Tabel 3 Kriteria Transformasi data Jumlah Asset UMKM

Kriteria Transformasi Jumlah Asset UMKM	
Jumlah Asset	Transformasi
Maksimal Rp. 50 Juta	1
> Rp 50 Juta – Rp 500 juta	2
> Rp 500 Juta – Rp 10 Milyar	3

Tabel 4 Transformasi Data UMKM

No.	Nama UMKM	Omset	Asset	Transformasi Omset	Transformasi Asset
a	Warung Tati	Rp. 720,000,000	Rp. 100,000,000	2	2
b	Toko Zhi-Zhi	Rp. 360,000,000	Rp. 500,000,000	2	2
c	Bengkel Las Kemang Tanjung	Rp. 120,000,000	Rp. 30,000,000	1	1
d	Toko Tasya	Rp. 1,440,000,000	Rp. 100,000,000	2	2
e	Fotokopi Naura	Rp. 180,000,000	Rp. 75,000,000	1	2
f	Bengkel Las Sony	Rp. 120,000,000	Rp. 100,000,000	1	2
g	Toko Grosir 3 Saudara	Rp. 2,540,000,000	Rp. 45,000,000	3	1
h	warung wahyu	Rp. 388,000,000	Rp. 50,000,000	2	1
i	Toko Baju Yossy YM	Rp. 54,000,000	Rp. 50,000,000	1	1
j	Rudi / Tk. Tasiya	Rp. 1,800,000,000	Rp. 100,000,000	2	2
k	H. Sidi Zainudin	Rp. 240,000,000	Rp. 400,000,000	1	2
l	Bambang	Rp. 1,350,000,000	Rp. 50,000,000	2	1
m	Aida	Rp. 1,080,000,000	Rp. 100,000,000	2	2
n	M. Yusuf / Bakso Kepala Sapi	Rp. 60,300,000	Rp. 40,000,000	1	2
o	Nin / Yoga Mobilindo	Rp. 6,500,000,000	Rp. 1,000,000,000	3	3

d. Modelling phase

Berdasarkan data yang telah diidentifikasi sebelumnya, kemudian diterapkan model data mining yang sesuai dengan kebutuhan. Pada penelitian ini, algoritma data mining yang akan digunakan adalah algoritma k-means clustering. Pemodelan data mining akan dilakukan sesuai dengan langkah kerja algoritma K-means clustering.

e. Evaluation phase

Setelah model data didapatkan, langkah selanjutnya adalah menilai sejauh mana model yang diperoleh dapat memenuhi tujuan bisnis yang telah diinisiasi diawal. Dari model yang diperoleh tentunya dapat digunakan oleh pemangku kepentingan dalam menentukan kebijakan, dalam hal ini adalah bentuk pendampingan yang tepat pada UMKM yang ada pada masing-masing cluster yang ada.

f. Desimination phase

Pengetahuan yang telah diperoleh dari modeling data selanjutnya perlu divisualkan dengan bentuk yang mudah dipahami dan dimanfaatkan oleh pengguna akhir atau pemangku kepentingan.

2.2 K-Means Clustering

Data mining merupakan proses menemukan pola yang menarik dan pengetahuan dari sekumpulan data[9]. Algoritma pengelompokan k-means didasarkan pada karya Stuart Lloyd dan E.W. Forgy dan kadang-kadang disebut sebagai algoritma Lloyd-Forgy atau algoritma Lloyd. Secara visual, algoritma k-means membagi ruang data menjadi k partisi atau batas, dimana centroid pada setiap partisi merupakan centroid dari cluster. k-means clustering merupakan metode clustering berbasis centroid dimana kumpulan data dibagi menjadi k cluster. k-means clustering adalah salah satu algoritma clustering yang paling sederhana dan paling umum digunakan. Dalam teknik ini, pengguna menentukan jumlah cluster (k) yang perlu dikelompokkan dalam kumpulan data. Tujuan dari k-means clustering adalah untuk menemukan titik data prototipe untuk setiap cluster; semua titik data kemudian ditugaskan ke prototipe terdekat, yang kemudian membentuk sebuah cluster.

Langkah-langkah kerja algoritma K-Means Clustering adalah sebagai berikut[5]:

a. Langkah 1: Menentukan Jumlah Centroids

Langkah pertama dalam algoritma k-means adalah menginisiasi k centroid secara acak. Jumlah cluster k harus ditentukan oleh pengguna.

b. Langkah 2: Tetapkan Poin Data

Setelah centroid telah dimulai, semua titik data sekarang ditugaskan ke centroid terdekat untuk membentuk sebuah cluster. Dalam konteks ini "terdekat" dihitung dengan ukuran kedekatan. 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 $X(x_1, x_2, \dots, x_n)$ dan $C(c_1, c_2, \dots, c_n)$ dengan n atribut.

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

Keterangan:

X dan y: representasi nilai atribut dari dua record.

c. Langkah 3: Hitung Centroid Baru

Langkah ini adalah menghitung centroid baru, berdasarkan keanggotaan dari cluster yang baru. Centroid baru ini merupakan titik data yang paling representatif dari semua titik data dalam cluster.

$$c = \frac{\sum m}{n} \dots \dots \dots (2)$$

Keterangan:

c = centroid

m = nilai data yang termasuk dalam centroid tertentu

n = jumlah data dari anggota centroid tertentu

d. Langkah 4: Ulangi Tugas dan Hitung Centroid Baru

Setelah centroid baru telah diidentifikasi, menetapkan titik data ke centroid terdekat diulang sampai semua titik data dipindahkan ke centroid baru.

e. Langkah 5: Penghentian

Langkah 3, menghitung centroid baru dan langkah 4, menetapkan titik data ke centroid baru, bersifat iteratif sampai tidak ada perubahan lebih lanjut dalam penugasan titik data yang terjadi atau, dengan kata lain, tidak ada perubahan signifikan pada centroid yang dicatat.

3. HASIL DAN PEMBAHASAN

3.1 Data Hasil Transformasi UMKM

Data yang yang didapatkan selanjutnya ditransformasi sesuai dengan ketentuan yang telah dibuat, sehingga dapat digunakan dalam proses perhitungan dalam algoritma k-means. Tabel 5 merupakan hasil transformasi data UMKM yang diperoleh.

Tabel 5 Data UMKM

No.	Nama UMKM	Transformasi Omset	Transformasi Asset
a	Warung Tati	2	2
b	Toko Zhi-Zhi	2	2
c	Bengkel Las Kemang Tanjung	1	1
d	Toko Tasya	2	2
e	Fotokopi Naura	1	2
f	Bengkel Las Sony	1	2
g	Toko Grosir 3 Saudara	3	1
h	warung wahyu	2	1
i	Toko Baju Yossy YM	1	1
j	Rudi / Tk. Tasiya	2	2
k	H. Sidi Zainudin	1	2
l	Bambang	2	1
m	Aida	2	2
n	M. Yusuf / Bakso Kepala Sapi	1	2
o	Nin / Yoga Mobilindo	3	3

3.2 Penerapan Algoritma K-Means Clustering terhadap Data UMKM

Dari data UMKM yang telah didapatkan, selanjutnya dilakukan proses modeling dengan algoritma K-Means clustering, Adapun langkah-langkahnya adalah sebagai berikut:

a. Menentukan Jumlah cluster

Jumlah cluster akan mengacu pada pengkategorian UMKM sesuai dengan undang-undang No. 20 Tahun 2008 tentang UMKM. Dalam undang-undang tersebut UMKM diklasifikasikan berdasarkan asset dan omset tiap skala usaha, seperti ditunjukkan pada tabel 2.

Tabel 2 Klasifikasi UMKM

Skala Usaha	Kekayaan Bersih/Asset	Hasil Penjualan/Omset
Usaha Mikro	Maksimal Rp. 50 Juta	Maksimal Rp. 300 Juta
Usaha kecil	> Rp 50 Juta – Rp 500 juta	> Rp 300 Juta – Rp 2,5 Milyar
Usaha Menengah	> Rp 500 Juta – Rp 10 Milyar	> Rp 2,5 Milyar – Rp 50 Milyar

Berdasarkan tabel tersebut, maka jumlah k yang akan digunakan sebanyak 3 (tiga).

b. Inisiasi k centroid secara acak

Centroid pertama ditentukan secara acak, jumlah centroid mengacu pada jumlah cluster yang akan didapatkan. Dalam penelitian ini pusat 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

Untuk menghitung distance atau jarak pada setiap cluster digunakan formula k-means yang telah didefinisikan sebelumnya. Tabel 3 merupakan perhitungan distance atau jarak ke masing-masing centroid dan hasil pengelompokan data pada iterasi pertama.

Tabel 3 Hasil Perhitungan dan Pengelompokan Data Pada Iterasi Pertama

No.	Nama UMKM	Transformasi Omset	Transformasi Asset	Jaraka ke Centroid 1	Jaraka ke Centroid 2	Jaraka ke Centroid 3	Cluster
a	Warung Tati	2	2	0	1	2	1
b	Toko Zhi-Zhi	2	2	0	1	2	1
c	Bengkel Las Kemang Tanjung	1	1	2	1	6	2
d	Toko Tasya	2	2	0	1	2	1
e	Fotokopi Naura	1	2	1	0	3	2
f	Bengkel Las Sony	1	2	1	0	3	2
g	Toko Grosir 3 Saudara	3	1	2	3	4	1
h	warung wahyu	2	1	1	2	5	1
i	Toko Baju Yossy YM	1	1	2	1	6	2
j	Rudi / Tk. Tasiya	2	2	0	1	2	1
k	H. Sidi Zainudin	1	2	1	0	3	2
l	Bambang	2	1	1	2	5	1
m	Aida	2	2	0	1	2	1
n	M. Yusuf / Bakso Kepala Sapi	1	2	1	0	3	2
o	Nin / Yoga Mobilindo	3	3	2	3	0	3

Berdasarkan hasil perhitungan dan pengelompokan awal, menunjukkan bahwa pengelompokan data pada iterasi pertama berbeda dengan pengelompokan data awal, maka dari itu akan dilanjutkan pada iterasi berikutnya. Langkah yang dilakukan selanjutnya adalah menentukan kembali nilai titik pusat (centroid) berdasarkan data dari cluster yang didapatkan. Dengan menggunakan formula yang telah didefinisikan sebelumnya berikut ini adalah centroid baru yang didapatkan, yaitu:

Centroid 1 : $(2, 2) \rightarrow (2, 125, 1625)$

Centroid 2 : $(1, 2) \rightarrow (1, 1, 666667)$

Centroid 3 : $(3, 3) \rightarrow (3, 3)$

Setelah nilai centroid baru didapatkan, lakukan perhitungan distance atau jarak ke masing-masing centroid kembali dengan menggunakan formula yang sama digunakan pada iterasi pertama. Hasil perhitungan dan pengelompokan data pada iterasi kedua dapat dilihat pada tabel 4.

Tabel 4 Hasil Perhitungan dan Pengelompokan Data Pada Iterasi Kedua

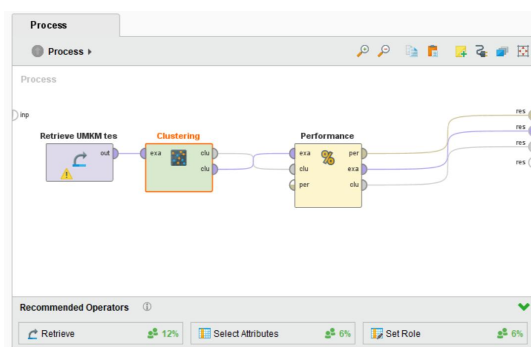
No.	Nama UMKM	Transformasi Omset	Transformasi Asset	Jaraka ke Centroid 1	Jaraka ke Centroid 2	Jaraka ke Centroid 3	Cluster
a	Warung Tati	2	2	0.265625	1.111111	2	1
b	Toko Zhi-Zhi	2	2	0.265625	1.111111	2	1
c	Bengkel Las Kemang Tanjung	1	1	1.515625	0.444444	6	2
d	Toko Tasya	2	2	0.265625	1.111111	2	1
e	Fotokopi Naura	1	2	1.265625	0.111111	3	2
f	Bengkel Las Sony	1	2	1.265625	0.111111	3	2
g	Toko Grosir 3 Saudara	3	1	1.265625	2.444444	4	1
h	warung wahyu	2	1	0.515625	1.444444	5	1
i	Toko Baju Yossy YM	1	1	1.515625	0.444444	6	2
j	Rudi / Tk. Tasiya	2	2	0.265625	1.111111	2	1
k	H. Sidi Zainudin	1	2	1.265625	0.111111	3	2
l	Bambang	2	1	0.515625	1.444444	5	1
m	Aida	2	2	0.265625	1.111111	2	1
n	M. Yusuf / Bakso Kepala Sapi	1	2	1.265625	0.111111	3	2
o	Nin / Yoga Mobilindo	3	3	2.765625	3.777778	0	3

Hasil pengelompokan data yang diperoleh pada iterasi kedua menunjukkan bahwa tidak ada perpindahan cluster, maka proses perhitungan distance atau jarak dihentikan. Dengan demikian hasil akhir dari pengelompokan data UMKM dengan menggunakan algoritma K-Means Clustering adalah sebagai berikut:

- Cluster 1 : Terdapat 8 UMKM, yaitu: Warung Tati, Toko Zhi-Zhi, Toko Tasya, Toko Grosir 3 Saudara, warung Wahyu, Rudi / Tk. Tasiya, Bambang dan Aida.
- Cluster 2 : Terdapat 6 UMKM, yaitu: Bengkel Las Kemang Tanjung, Fotokopi Naura, Toko Baju Yossy YM, H. Sidi Zainudin dan M. Yusuf / Bakso Kepala Sapi
- Cluster 3 : Terdapat 1 UMKM, yaitu: Nin / Yoga Mobilindo

3.3 Pengujian dengan RapidMiner

Untuk lebih memastikan hasil perhitungan yang dilakukan, pada penelitian ini juga dilakukan proses perhitungan dengan menggunakan software RapidMiner dengan data yang sama sebagai pembandingan. Gambar 2 merupakan proses clustering dengan menggunakan algoritma k_means di RapidMiner. Centroid yang digunakan juga sebanyak 3 (tiga). Sedangkan gambar 3 merupakan data yang UMKM yang sama juga digunakan pada proses perhitungan manual.

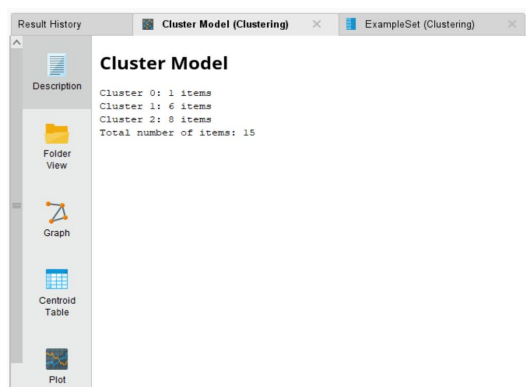


Gambar 2 Proses Clustering pada RapidMiner

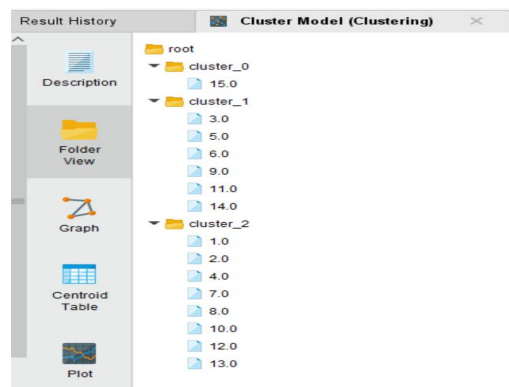
Row No.	id	Nama UMMK	cluster	Omset	Asset
1	1	Warung Tati	cluster_2	2	2
2	2	Toko Zhi-Zhi	cluster_2	2	2
3	3	Bengkel Las Kemang Tanjung	cluster_1	1	1
4	4	Toko Tasya	cluster_2	2	2
5	5	Fotokopi Naura	cluster_1	1	2
6	6	Bengkel Las Sony	cluster_1	1	2
7	7	Toko Grosir 3 Saudara	cluster_2	3	1
8	8	warung wahyu	cluster_2	2	1
9	9	Toko Baju Yossy YM	cluster_1	1	1
10	10	Rudi / Tk Tasya	cluster_2	2	2
11	11	H. Sidi Zainudin	cluster_1	1	2
12	12	Bambang	cluster_2	2	1
13	13	Aida	cluster_2	2	2

Gambar 3 Data UMMK pada RapidMiner

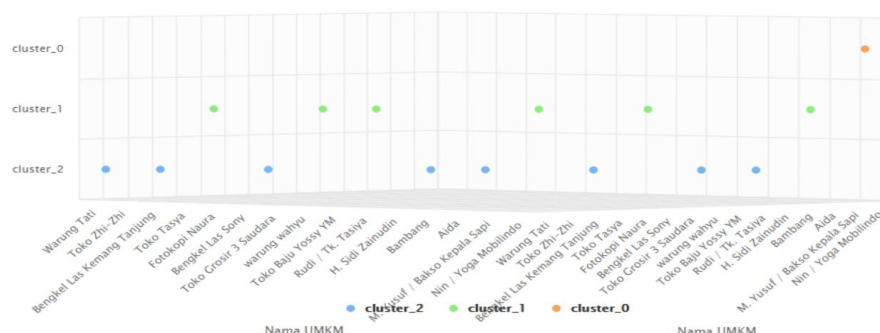
Setelah proses dijalankan, hasil menunjukkan bahwa jumlah cluster sebanyak 3 (tiga) diperoleh dengan total data item sebanyak 15 record. Hasil cluster model pada RapidMiner ditunjukkan pada gambar 4. Sedangkan pada gambar 5 menunjukkan pengelompokan data yang dihasilkan dimana pada cluster_0.



Gambar 4 Hasil Cluster Model Pada RapidMiner



Gambar 5 Hasil Pengelompokan Data Pada RapidMiner



Gambar 6 Visualisasi Cluster UMMK Pada RapidMiner

3.4 Analisis Hasil

Hasil dari pengelompokan dengan algoritma K-Means Clustering berdasarkan asset dan omset berhenti pada iterasi kedua dan menghasilkan 3 (tiga) cluster. Dimana UMMK Nin / Yoga Mobilindo berada pada cluster 3, yang artinya UMMK ini memiliki omset lebih dari 2,5 Milyar dan omset lebih dari 500 Juta. UMMK pada cluster 2 (dua) diantaranya adalah Terdapat 6 UMMK, yaitu: Bengkel Las Kemang Tanjung, Fotokopi Naura, Toko Baju Yossy YM, H. Sidi Zainudin dan M. Yusuf / Bakso Kepala Sapi. Dengan omset lebih besar dari 300 juta dan asset lebih besar dari 50 juta. Sedangkan untuk yang lainnya masuk pada cluster ke 3 (tiga).

Dengan pengelompokan yang telah dilakukan dapat mempermudah instansi atau pihak terkait dalam memberikan pendampingan bagi UMKM kearah yang lebih baik.

4. KESIMPULAN

Berdasarkan perhitungan dan uji coba menggunakan aplikasi rapidminer, peneliti menarik beberapa kesimpulan, yaitu:

- a. Dari tiga cluster yang telah ditentukan, untuk cluster 1 (satu) terdapat 1 (satu) UMKM, cluster 2 (dua) terdapat 6 (enam) UMKM dan cluster 3 (tiga) terdapat 8 (delapan) UMKM.
- b. Tidak ada perbedaan hasil akhir antara perhitungan dengan excel dengan aplikasi rapidminer, yaitu diperoleh tiga cluster UMKM.
- c. Hasil perhitungan diperoleh pada iterasi ke 2 (dua).

5. SARAN

Pada penelitian ini masih memiliki banyak kekurangan sehingga diperlukan rencana pengembangan di penelitian selanjutnya. Beberapa saran yang dapat dilakukan pada penelitian selanjutnya adalah memasukan aspek tenaga kerja dan umur usaha dalam kriteria pengelompokkan. Selain itu juga jumlah data yang digunakan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Kepala LPPM Universitas Multi Data Palembang yang telah memberi dukungan sehingga penelitian ini dapat terlaksana.

DAFTAR PUSTAKA

- [1] K. K. B. PEREKONOMIAN and R. INDONESIA, "UMKM Menjadi Pilar Penting dalam Perekonomian Indonesia," 2021. <https://www.ekon.go.id/publikasi/detail/2939/dukungan-pemerintah-bagi-umkm-agar-pulih-di-masa-pandemi>.
- [2] Departemen Pengembangan UMKM Bank Indonesia, "Pemetaan dan strategi peningkatan daya saing umkm dalam menghadapi masyarakat ekonomi asean (MEA 2015)," *J. RISALAH*, vol. 27, no. 1, p. 68, 2016, [Online]. Available: <http://ejournal.uin-suska.ac.id/index.php/risalah/article/view/2510%0A>.
- [3] D. T. Larose and C. D. Larose, *Discovering Knowledge in Data*. 2014.
- [4] Jiawei Han Micheline Kamber, *DataMining Concepts and Techniques*. 2006.
- [5] C. D. L. Daniel T. Larose, *Data Mining and Predictive Analytics*. Canada: John Wiley & Sons, 2015.
- [6] 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.
- [7] L. Magdalena and R. Fahrudin, "Penerapan Data Mining Untuk Koperasi Se-Jawa Barat Menggunakan Metode Clustering pada Kementerian Koperasi dan UKM," *J. Digit*, vol. 9, no. 2, p. 190, 2020, doi: 10.51920/jd.v9i2.120.
- [8] UU No. 20 Tahun 2008, "UU No. 20 Tahun 2008," *UU No. 20 Tahun 2008*, no. 1, pp. 1–31, 2008.
- [9] M. Liang, *Data Mining: Concepts, Models, Methods, and Algorithms*, vol. 36, no. 5. 2004.