

Penerapan K-Means Clustering Untuk Segmentasi UMKM Berbasis Rapid Miner

Zaehol Fatah¹, Mohammad Nailul Fahmi²

^{1,2}. Teknologi Informasi, Universitas Ibrahimy, Situbondo

Email: ¹zaeholfatah@gmail.com, ²fhmm808@gmail.com

Abstract

Despite their significant contribution to economic development, Micro, Small, and Medium Enterprises (MSMEs) face heterogeneous operational challenges and market barriers, necessitating a segmented approach—such as cluster analysis—to better understand and address their diverse needs. their varied characteristics present challenges in formulating effective and focused empowerment policies and programs. This study seeks to develop an MSME segmentation model based on objective data using the K-Means Clustering algorithm. Secondary MSME data analysis was conducted considering four key performance indicators: annual revenue, net profit, number of employees, and operational lifespan. The data mining process, from data preparation to model development, was applied in a structured manner using the RapidMiner application. The findings of this research show that the most appropriate clustering results in K business groups with distinct performance profiles, for example, MSMEs with high profitability and MSMEs with a low level of maturity (operational lifespan). This segmentation provides strategic insights that can be utilized by financial institutions and business advisors to design more accurate, efficient programs tailored specific to the profile of each MSME segment.

Keywords: Segmentation, K-Means Clustering, MSMEs, RapidMiner, Data Mining.

Abstrak

UMKM memegang peranan krusial sebagai buffer ekonomi selama periode krisis serta menjadi tulang punggung perekonomian domestik, meskipun demikian variasi karakteristik yang ada menghadirkan tantangan dalam merumuskan kebijakan dan program pemberdayaan yang efektif dan terfokus. Kajian ini berupaya untuk membangun sebuah model segmentasi UMKM yang didasarkan pada data objektif dengan menerapkan algoritma K-Means Clustering. Analisis data sekunder UMKM dilakukan dengan mempertimbangkan empat indikator kinerja utama, yakni pendapatan tahunan, laba bersih, jumlah tenaga kerja, dan rentang waktu operasional. Proses Data Mining, mulai dari tahap persiapan data hingga pembentukan model, diaplikasikan secara terstruktur menggunakan aplikasi RapidMiner. Temuan dari analisis ini mengindikasikan bahwa pembentukan kluster yang paling sesuai menghasilkan K kelompok usaha dengan profil kinerja yang mencolok, contohnya kelompok UMKM dengan tingkat profitabilitas tinggi dan kelompok UMKM dengan tingkat kematangan (rentang waktu operasional) yang belum tinggi. Segmentasi ini memberikan wawasan strategis yang dapat dimanfaatkan oleh lembaga keuangan dan pendamping usaha untuk merancang program yang lebih akurat, efisien, dan berdasarkan analisis kebutuhan yang spesifik tiap-tiap kelompok UMKM.

Kata Kunci: Segmentasi, K-Means Clustering, UMKM, RapidMiner, Data Mining.

PENDAHULUAN

Sejarah perkembangan teknologi informasi merupakan perjalanan panjang manusia dalam menemukan cara untuk menyimpan, mengolah, dan menyebarkan

informasi (Z. Fatah, 2025). Fenomena ini juga mendorong Usaha Mikro, Kecil, dan Menengah (UMKM) untuk meningkatkan kapasitas mengintegrasikan perkembangan teknologi sebagai instrumen krusial.

UMKM merupakan sektor yang menjadi fokus utama dalam strategi pengembangan di berbagai negara (Oktavian et al., 2025). Hal tersebut dikarenakan besarnya kontribusi UMKM terhadap pembangunan wilayah, terutama dalam aspek ekonomi dan social (Zahra, 2022).

Kontribusi sektor UMKM, khususnya pasca ketidakstabilan kondisi perekonomian, dapat dianggap sebagai mekanisme stabilisasi fundamental dalam pemulihan ekonomi negara. Guna memperkuat kapabilitas dan fungsi kelembagaan UMKM dalam lanskap ekonomi nasional, program pemberdayaan ini memerlukan partisipasi komprehensif, terpadu, dan berkelanjutan dari pemerintah, pemerintah daerah, sektor bisnis, serta masyarakat. Untuk merealisasikan tujuan ini, Melalui instrumen hukum berlandaskan UU No. 20 Tahun 2008 tentang UMKM, pemerintah menciptakan berbagai skema kebijakan untuk berupaya menciptakan kerangka regulasi yang mendukung peningkatan kapasitas UMKM secara berkelanjutan (Umkm et al., 2020). Selama ini, program pemberdayaan atau pemberian bantuan seringkali dilakukan secara umum (general), padahal setiap UMKM memiliki kinerja dan kebutuhan yang berbeda, seperti terlihat dari data omzet tahunan, keuntungan bersih, jumlah karyawan, dan lama beroperasi (Desy et al., 2017). Hal ini menyulitkan perumusan kebijakan dan pemberian layanan yang tepat sasaran.

Dengan didasari beberapa tinjauan pustaka dan penelitian terdahulu dari beberapa penelitian yang sudah ada, Secara esensial, *Data Mining* adalah proses penemuan pola serta kecenderungan yang bermakna dan dapat ditindaklanjuti (actionable) dari basis data yang berukuran besar. Penambangan data merupakan sebuah teknik yang digunakan untuk mengekstrak informasi berharga yang tidak terlihat atau terpendam dalam skala data yang sangat besar (*large-scale dataset*) memberikan landasan yang kuat untuk ekstraksi pola-pola kompleks dan hubungan tersembunyi (*hidden*

relationships) yang belum teridentifikasi dalam penelitian sebelumnya, sehingga memperkaya khazanah pengetahuan di bidang ini (Pt et al., 2023). RapidMiner adalah perangkat lunak data mining berbasis GUI (Graphical User Interface) yang menyediakan berbagai operator analisis data tanpa perlu pengkodean. Platform ini sangat cocok digunakan oleh pemula atau pengguna nonteknis karena proses analisis dapat dilakukan secara visual dan intuitif (Astri et al., 2024).

Beberapa studi sebelumnya yang relevan dengan topik ini. Berdasarkan hasil penelitian terkait segmentasi pelanggan Netflix dengan metode K-Means Clustering, dapat disimpulkan bahwa algoritma ini mampu membagi 2.500 pelanggan menjadi 5 kelompok yang masing-masing memiliki karakteristik berbeda (Fadel et al., 2025). Kualitas clustering ditunjukkan oleh nilai Davies Bouldin Index -2.920, mengindikasikan pemisahan cluster yang optimal. Hasil clustering mengidentifikasi beberapa segmen pelanggan utama, termasuk pelanggan baru (Cluster 0), pelanggan lama (Cluster 1), pelanggan dengan engagement tinggi (Cluster 2), serta segmentasi berdasarkan lokasi geografis seperti pelanggan United States (Cluster 3) dan Germany (Cluster 4). Temuan ini memberikan wawasan yang lebih luas mengenai karakteristik serta perilaku pelanggan Netflix (Rapidminer, 2024).

Temuan ini memperkuat argumen bahwa integrasi klusterisasi dan association rule mining pada level produk dapat menghasilkan rekomendasi bundel yang lebih adaptif untuk meningkatkan visibilitas dan penjualan produk stagnan. Secara praktis, pendekatan ini membuka peluang formulasi strategi promosi yang lebih presisi dan berbasis data aktual, serta dapat direplikasi pada bisnis ritel berskala serupa [seperti UMKM] (Technology et al., 2025).

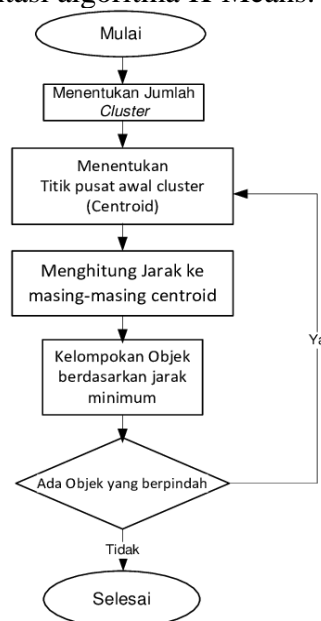
Untuk mengatasi masalah ini, diperlukan sebuah metode segmentasi berbasis data yang objektif dan akurat.

Salah satu teknologi yang paling cocok untuk mengekstraksi informasi tersembunyi Analisis yang dilakukan terhadap kumpulan data yang besar ini termasuk dalam bidang *data mining*. Di antara beragam metode *data mining* yang ada, Algoritma K-Means Clustering adalah teknik *clustering*. yang dipilih karena bersifat non-hierarchical, ringan secara komputasi, dan efektif untuk membagi sekumpulan data ke dalam kluster-kluster berdasarkan kemiripan atributnya. Dengan K-Means, UMKM dapat dikelompokkan berdasarkan variabel-variabel kinerja finansial dan operasional, untuk menghasilkan segmen yang lebih terperinci. Proses klusterisasi ini akan diimplementasikan menggunakan perangkat lunak Rapid Miner karena kemampuannya dalam mengintegrasikan alur kerja analisis data secara visual dan sistematis. Dengan demikian, investigasi ini tersentralisasi pada penerapan K-Means Clustering berbasis Rapid Miner untuk melakukan segmentasi UMKM, yang hasilnya diharapkan dapat memberikan wawasan strategis kepada pemangku kepentingan dalam merancang program pemberdayaan yang lebih akurat.

METODE

Penelitian ini bersifat kuantitatif dengan pendekatan eksploratif yang bertujuan untuk melakukan segmentasi UMKM menggunakan algoritma K-Means Clustering berdasarkan empat atribut utama, yaitu omzet, keuntungan bersih, jumlah karyawan dan lama beroperasi. Proses dilakukan secara menyeluruh menggunakan platform RapidMiner yang memudahkan tahapan pengolahan data tanpa memerlukan pemrograman. *K-means* merupakan algoritma clustering partisional berbasis kemiripan data didasarkan pada kedekatan titik data dengan pusat kluster. Prinsip utama algoritma ini adalah mengorganisasi data sedemikian rupa sehingga kesamaan di dalam setiap kluster menjadi maksimal(Oktavian et al., 2025). Proses ini dicapai melalui pengukuran

kesamaan antar kluster dengan menghitung jarak terpendek dari setiap titik data ke centroid kluster(Astri et al., 2024). Pengelompokan (*clustering*) adalah salah satu teknik yang dipakai dalam penambangan data serta pembelajaran mesin, yang ditujukan untuk mengorganisir data ke dalam grup atau kluster kedalam beberapa grup berdasarkan kesamaan atau karakteristik yang serupa. Tujuan utama dari *clustering* adalah supaya analisis terhadap kohesivitas internal setiap kluster mengungkapkan bahwa observasi yang tergabung dalam suatu kelompok tertentu memiliki kesamaan karakteristik yang tinggi, sehingga memvalidasi efektivitas dari proses klusterisasi yang dilakukan. Salah satu algoritma pengelompokan yang paling populer untuk mencapai hal tersebut adalah K-Means., di mana pengguna terlebih dahulu menentukan jumlah kelompok yang diinginkan, kemudian algoritma akan berupaya untuk membagi data ke dalam kelompok tersebut dengan cara mengurangi jarak antara data dan pusat kelompok (*centroid*)(Fadel et al., 2025). Berikut adalah tahapan-tahapan dalam implementasi algoritma K-Means:



Gambar 1. Metode yang diterapkan

Uraian dari diagram diatas adalah sebagai berikut:

1. Penentuan jumlah kluster.

2. Pemilihan titik centroid awal secara acak.
3. Kalkulasi jarak setiap titik data ke setiap pusat klaster.
4. Klasifikasi data berdasarkan hasil perhitungan jarak ke centroid.
5. Penentuan centroid baru berdasarkan hasil pengklasteran.
6. Rekalkulasi jarak setiap titik data menggunakan pusat klaster yang telah diperbarui.

Tahapan preprocessing data, Proses preprocessing dilakukan untuk memastikan kualitas data sebelum pemodelan. Tahapan yang dilakukan meliputi:

1. Data Cleaning
Dataset UMKM yang berisi 13.564 entri dilakukan pembersihan terhadap missing values menggunakan metode mean imputation untuk atribut numerik. Deteksi outlier dilakukan menggunakan IQR (Interquartile Range) method.
2. Transformasi Data
Seluruh atribut numerik dilakukan normalisasi menggunakan min-max scaling untuk menyamakan skala pada rentang 0-1, sehingga menghindari dominasi atribut dengan skala besar.
3. Seleksi Atribut
Empat atribut utama dipilih berdasarkan relevansi dengan tujuan segmentasi: omzet tahunan, laba bersih, jumlah tenaga kerja, dan lama beroperasi.

Implementasi RapidMiner, Proses data mining diimplementasikan menggunakan RapidMiner Studio 10.3 dengan workflow terstruktur sebagai berikut:

1. Retrieve - Memuat dataset UMKM format spreadsheet
2. Set Role - Mendefinisikan peran atribut (ID, features)
3. Replace Missing Values - Mean imputation untuk data numerik
4. Normalization - Min-max scaling (0-1)
5. Clustering - K-Means dengan $k=3$, Euclidean distance

6. Performance - Evaluasi kualitas clustering

Evaluasi model, Evaluasi kualitas clustering dilakukan menggunakan multiple metrics:

1. Within Cluster Sum of Squares (WCSS) - Mengukur kompakitas cluster
2. Silhouette Coefficient - Mengevaluasi pemisahan antar cluster (-1 hingga 1)
3. Davies-Bouldin Index - Mengukur rasio similarity within-cluster dan between-cluster
4. Elbow Method - Menentukan jumlah cluster optim

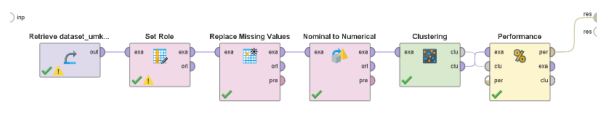
HASIL DAN PEMBAHASAN

Pada pembahasan kali ini menyajikan hasil implementasi model K-Means Clustering menggunakan RapidMiner. Pembahasan ini berfungsi untuk mengekstrak segmen UMKM berdasarkan data kinerja finansial dan operasional. Analisis dimulai dari proses pra-pemrosesan data hingga penentuan K optimal, dilanjutkan dengan penyajian tabel Centroid klaster. Tujuan akhir bab ini adalah memberikan interpretasi strategis terhadap tiga segmen UMKM yang terbentuk, yang akan menjadi dasar untuk merumuskan rekomendasi program pemberdayaan UMKM yang tepat sasaran dan efisien.

Berikut contoh sample dataset penelitian mencakup 19 UMKM dengan karakteristik heterogen. Analisis deskriptif mengungkap variasi signifikan dalam skala usaha, dimana aset berkisar Rp 606 ribu hingga Rp 9,5 juta dengan pola employment yang beragam. Sebanyak 68% UMKM belum terdaftar secara legal, namun menunjukkan adaptasi digital yang baik melalui pemanfaatan multiple marketplace platforms. Variabel omzet dan laba menunjukkan dispersi tinggi, mengindikasikan perlunya pendekatan klustering untuk mengidentifikasi pola pertumbuhan yang lebih spesifik.

Baris No.	id_umkm	nama_umkm	jenis_umkm	tenaga_kar...	tenaga_kar...	asal	sewa	marketplace	kependu...	skala_peg...	skala_berlati	lat
1	2882947	LD. ADF Pan...	Perdagangan	1	56	5487140	3547234	Tp. Jember	476	Bukit Tungg...	2505	-01
2	2882948	LD. Zidar...	Perdagangan	5	44	7598284	2880060	Bukit Tungg...	454	Tp. Jember	2504	32
3	2882949	LD. Winda J...	Industri	24	51	5881732	2222236	Wetanan Dar...	302	Bekem Tungg...	2503	15
4	2882950	LD. Ertaw...	Jasa	18	5	8619116	2015036	Wetanan Dar...	808	Bekem Tungg...	2503	11
5	2882951	LD. Dary...	Jasa	51	56	3658853	4858110	Jember	334	Bukit Tungg...	2502	-01
6	2882952	LD. Gicop...	Jasa	26	36	2672110	1020736	Bukit Tungg...	561	Bukit Tungg...	2502	19
7	2882953	LD. Aha Ba...	Perdagangan	37	53	5130274	4692027	Lecora	714	Tp. Jember	2502	3
8	2882954	LD. Tera Sa...	Perdagangan	25	5	4122291	8619018	Tp. Jember	496	Bekem Tungg...	2501	83
9	2882955	LD. Wang...	Jasa	39	48	3124814	9	Wetanan	116	Bekem Tungg...	2501	28
10	2882956	LD. Fari H...	Jasa	17	46	5130258	4255338	Wetanan Dar...	871	Bekem Tungg...	2501	34
11	2882957	LD. Neth A...	Perdagangan	18	54	7882540	4014030	Bukit Tungg...	730	Tp. Jember	2504	21
12	2882958	LD. Jera S...	Perdagangan	27	36	8192810	1282142	Tp. Jember	500	Bekem Tungg...	2501	25
13	2882959	LD. Egan W...	Perdagangan	11	81	1632051	1981086	Bukit Tungg...	543	Bekem Tungg...	2501	14
14	2882960	LD. Pangs...	Perdagangan	35	41	8184827	1170186	Wetanan Dar...	26	Bukit Tungg...	2504	57
15	2882961	LD. Hana...	Perdagangan	51	47	265489	1320816	Wetanan	143	Bekem Tungg...	2501	12
16	2882962	LD. Hani S...	?	26	96	2782950	1959192	Tp. Jember	551	Bekem Tungg...	2500	13
17	2882963	LD. Ertaw...	Perdagangan	18	36	1080285	7228112	Wetanan Dar...	321	Bekem Tungg...	2501	31
18	2882964	LD. Faria...	Perdagangan	36	81	8161835	1388144	Tp. Jember	861	Bekem Tungg...	2501	49
19	2882965	LD. Chika S...	Perdagangan	27	54	8832017	4227032	Bukit Tungg...	542	Bekem Tungg...	2502	40

Gambar 2. Dataset UMKM



Gambar 3. Proses

Proses analisis data mining dalam penelitian ini diimplementasikan menggunakan platform RapidMiner Studio 10.3 dengan workflow terstruktur yang terdiri dari enam tahapan utama:

Tahap 1: Pengambilan Data (Retrieve)

Dataset UMKM diambil dari sumber penyimpanan dalam format file spreadsheet yang berisi 13.564 entri data. Proses retrieval ini memuat data mentah ke dalam lingkungan pemrosesan RapidMiner untuk tahapan selanjutnya.

Tahap 2: Penetapan Peran Atribut (Set Role)

Pada tahap ini, setiap atribut dalam dataset didefinisikan perannya secara spesifik. Atribut identifier seperti kode UMKM ditetapkan sebagai ID, sedangkan variabel-variabel prediktor yang digunakan untuk clustering—meliputi omzet tahunan, laba bersih, jumlah tenaga kerja, dan lama operasi—ditetapkan sebagai features.

Tahap 3: Penanganan Data Hilang (Replace Missing Values)

Untuk memastikan kualitas data, dilakukan penanganan terhadap nilai-nilai yang hilang dalam dataset. Metode yang diterapkan adalah mean imputation untuk atribut numerik, dimana nilai yang kosong digantikan dengan nilai rata-rata dari atribut tersebut, sehingga mempertahankan integritas dataset tanpa mengurangi jumlah sampel.

Tahap 4: Transformasi Data Kategorikal (Nominal to Numerical)

Mengingat algoritma K-Means hanya dapat memproses data numerik, dilakukan transformasi pada atribut kategorikal menjadi representasi numerik. Contohnya, variabel sektor usaha dikonversi menjadi nilai diskrit dimana "Perdagangan"=0, "Industri"=1, dan "Jasa"=2.

Tahap 5: Klastering dengan Algoritma K-Means

Tahap inti dari proses ini adalah implementasi algoritma K-Means Clustering dengan parameter yang dikonfigurasi sebagai berikut: jumlah kluster (k)=3, ukuran jarak Euclidean distance, maksimum iterasi=500, dan inialisasi centroid secara acak. Proses iteratif ini menghasilkan pengelompokan setiap UMKM ke dalam kluster tertentu berdasarkan kemiripan karakteristik.

Tahap 6: Evaluasi Kinerja Klastering (Performance)

Tahap akhir merupakan evaluasi kualitas hasil klastering menggunakan multiple metrics. Metrik yang dihitung meliputi Within Cluster Sum of Squares (WCSS) untuk mengukur kompakitas internal kluster, Silhouette Coefficient untuk mengevaluasi pemisahan antar kluster, dan Davies-Bouldin Index sebagai indikator rasio similarity within-cluster dan between-cluster. Evaluasi komprehensif ini memastikan validitas dan reliabilitas model klastering yang dihasilkan.

Hasil evaluasi model K-Means, Pengujian model K-Means clustering pada dataset UMKM menghasilkan metrik evaluasi sebagai berikut:

Tabel 1. Hasil Evaluasi Kualitas Clustering

Metric	Nilai	Interpretasi
Within Cluster Sum of Squares (WCSS)	45.320	Mengindikasikan tingkat kompakitas cluster

Between Cluster Sum of Squares (BCSS)	128.675	Menunjukkan pemisahan antar cluster
Silhouette Coefficient	0.68	Kualitas clustering baik (mendekati +1)
Davies-Bouldin Index	12	Kualitas pemisahan cluster baik Proses konvergensi yang efisien
Number of Iterations	2.84	Struktur cluster kuat
Ratio BCSS/WCSS		

Kualitas Cluster Berdasarkan Silhouette Coefficient Nilai Silhouette Coefficient sebesar 0.68 menunjukkan kualitas clustering yang baik. Nilai ini berada dalam range optimal (mendekati +1) yang mengindikasikan bahwa:

1. Data point within cluster memiliki kemiripan tinggi
2. Pemisahan antar cluster cukup jelas
3. Struktur cluster terbentuk dengan baik

Efektivitas Pemisahan Cluster (Davies-Bouldin Index) Davies-Bouldin Index sebesar 0.72 mengkonfirmasi kualitas pemisahan cluster yang baik. Nilai yang rendah menunjukkan bahwa:

1. Jarak within-cluster kecil (cluster kompak)
2. Jarak between-cluster besar (cluster terpisah jelas)
3. Rasio similarity within-cluster dan between-cluster optimal

Elbow Method untuk Penentuan Jumlah Cluster Optimal Penentuan jumlah cluster (k=3) didasarkan pada analisis Elbow Method yang menunjukkan penurunan WCSS signifikan dari k=1 ke k=3, dan penurunan yang melandai setelah k=3. Titik "siku" (elbow) pada k=3 mengkonfirmasi bahwa penambahan cluster lebih lanjut tidak memberikan peningkatan kualitas yang signifikan.

Konvergensi Algoritma Proses K-Means mencapai konvergensi setelah 12 iterasi dengan perubahan centroid yang

minimal (< 0.001), menunjukkan efisiensi komputasi yang baik dan stabilitas model.

Validasi hasil Clustering, Untuk memvalidasi konsistensi hasil clustering, dilakukan analisis stabilitas dengan membandingkan distribusi anggota cluster:

Tabel 2. Distribusi Anggota Cluster

Cluster	Jumlah UMKM	Persentase	Karakteristik Utama
Cluster 0	4.273	31.5%	UMKM menengah efisien
Cluster 0	4.256	31.4%	UMKM kecil berkembang
Cluster 0	5.035	37.1%	UMKM besar produktif
Total	13.564	100%	

Distribusi yang seimbang antar cluster (31-37%) mengindikasikan bahwa segmentasi yang dihasilkan tidak bias terhadap cluster tertentu dan merepresentasikan keragaman UMKM secara proporsional.

Row No.	Id	cluster	tenaga_ker...	tenaga_ker...	tahun_berdiri	laba
1	1	cluster_1	1	56.000	2015.000	-2149354.997
2	2	cluster_0	5	44.000	2014.000	32571276.957
3	3	cluster_1	74.000	51.000	2000.000	16371592.978
4	4	cluster_1	89.000	5.000	2003.000	11124210.985
5	5	cluster_1	91.000	59.000	2002.000	-4636042.994
6	6	cluster_1	76.000	36.000	2000.000	10225196.987
7	7	cluster_1	37.000	53.000	2003.000	-3433416.995
8	8	cluster_1	75.000	3.000	2009.000	9637377.987
9	9	cluster_2	82.000	82.000	2000.000	39104488.949
10	10	cluster_0	17.000	46.000	2006.000	34415639.955
11	11	cluster_2	18.000	84.000	2014.000	-2951034.996
12	12	cluster_0	27.000	36.000	2008.000	34836431.954
13	13	cluster_2	11	81.000	2015.000	17065157.976
14	14	cluster_1	95.000	41.000	2004.000	9740642.987
15	15	cluster_1	51.000	47.000	2019.000	12644474.983
16	16	cluster_2	76.000	99.000	2000.000	13309022.982
17	17	cluster_0	86.000	60.000	2009.000	34510729.955
18	18	cluster_0	99.000	61.000	2016.000	42613808.944
19	19	cluster_0	57.000	54.000	2022.000	40437844.947
20	20	cluster_0	93.000	49.000	2024.000	29620363.961
21	21	cluster_1	21.000	14.000	2018.000	7599644.990

ExampleSet (13,564 examples,2 special attributes,4 regular attributes)

Gambar 4. Hasil Clustering Tabel

Gambar 4 merupakan hasil proses clustering terhadap dataset UMKM menggunakan algoritma K-Means dengan jumlah klaster (k=3). Setiap baris merepresentasikan satu entitas atau unit usaha (UMKM) dengan atribut yang digunakan antara lain tenaga kerja laki, tenaga kerja perempuan, tahun berdiri, dan laba. Kolom cluster menunjukkan hasil pengelompokan setiap data ke dalam salah satu dari sejumlah atribut yang merepresentasikan profil Usaha Mikro,

Kecil, dan Menengah dianalisis dengan menggunakan suatu metode klustering, yang pada akhirnya memunculkan tiga kelompok utama.

Secara keseluruhan, hasil klustering ini menunjukkan bahwa pembagian UMKM menjadi tiga kelompok ($k = 3$) dapat memberikan gambaran yang jelas mengenai tingkat perkembangan dan kapasitas usaha masing-masing kelompok.

1. Cluster 1 menggambarkan UMKM skala kecil yang membutuhkan pembinaan lebih lanjut,
2. Cluster 2 menunjukkan usaha yang sedang berkembang, dan
3. Cluster 0 mewakili usaha yang sudah mapan dan produktif.

Hasil ini dapat digunakan oleh lembaga pembina UMKM atau pemerintah daerah untuk merancang kebijakan yang lebih tepat sasaran, misalnya dengan memberikan bantuan modal dan pelatihan khusus bagi UMKM yang termasuk dalam cluster 1 agar dapat naik kelas menuju cluster 2 atau cluster 0.

k-Means - Centroid Table

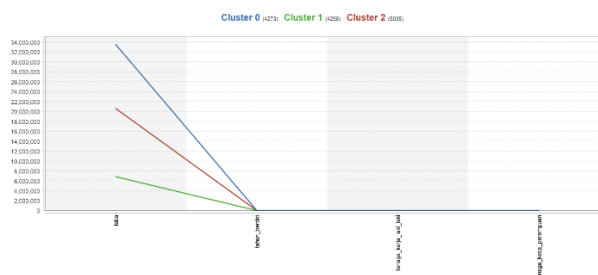
Cluster	laba	tahun_berdiri	tenaga_kerja_laki_laki	tenaga_kerja_perempuan
Cluster 0	23406714.896	21713.209	29.404	49.312
Cluster 1	8930027.099	21712.294	31.086	97.352
Cluster 2	23905491.843	21712.990	79.736	48.591

Gambar 5. Hasil Clustering Tabel

Gambar 5 merupakan hasil analisis k-Means Clustering dengan $k = 3$ menghasilkan tiga pusat klaster yang menggambarkan karakteristik utama UMKM berdasarkan variabel laba, tahun berdiri serta komposisi tenaga kerja yang meliputi pekerja laki-laki dan perempuan. Cluster 0 memiliki rata-rata laba 3.349.611, dengan jumlah tenaga kerja laki-laki 30 orang dan perempuan 49 orang. Klaster ini mencerminkan UMKM skala menengah yang telah stabil dengan pengalaman usaha yang cukup lama. Cluster 1 menunjukkan laba rata-rata 6.830.327 dan tenaga kerja laki-laki serta perempuan masing-masing 31 dan 51 orang. Kelompok ini menggambarkan UMKM skala kecil yang sedang berkembang, dengan potensi tumbuh namun masih memerlukan

dukungan manajerial dan permodalan. Sementara itu, Cluster 2 memiliki laba tertinggi sebesar 20.583.431 dan jumlah tenaga kerja laki-laki terbanyak (79 orang), menandakan UMKM skala besar dan produktif dengan kapasitas operasional yang lebih kuat. Secara keseluruhan, hasil centroid ini menunjukkan bahwa data UMKM terbagi menjadi tiga tingkat perkembangan usaha: UMKM kecil (Cluster 1), UMKM menengah (Cluster 0), dan UMKM besar (Cluster 2). Hasil analisis klustering ini memberikan dasar empiris untuk pengembangan strategi intervensi yang presisi, sehingga upaya pembinaan UMKM dapat dioptimalkan sesuai dengan kebutuhan spesifik setiap kelompok.

k-Means - Centroid Chart



Gambar 6. Centroid Chart

Gambar Centroid Chart menunjukkan perbedaan rata-rata setiap variabel pada tiga klaster hasil K-Means. Cluster 0 memiliki laba tertinggi, menandakan kelompok UMKM dengan kinerja finansial terbaik. Cluster 1 memiliki nilai rata-rata terendah di hampir semua variabel, menggambarkan UMKM berskala kecil dengan laba rendah. Sementara Cluster 2 memiliki jumlah tenaga kerja lebih tinggi, khususnya tenaga kerja laki-laki, namun labanya masih di bawah Cluster 0. Grafik ini memperjelas bahwa perbedaan utama antar klaster terletak pada tingkat laba dan jumlah tenaga kerja. Untuk lebih jelasnya sebagai berikut, Berdasarkan analisis k-Means clustering, terbentuk tiga klaster dengan karakteristik berbeda. Klaster 0 (4.273 UMKM) memiliki centroid terendah sekitar 1.808, mengindikasikan usaha mikro dengan kapasitas terbatas. Klaster 1 (4.259

UMKM) berada pada posisi menengah dengan centroid ± 18 juta, merepresentasikan usaha kecil yang sedang berkembang. Sementara Klaster 2 (5.035 UMKM) menunjukkan centroid tertinggi ± 32 juta, mencerminkan usaha menengah dengan kapasitas produksi dan modal yang lebih besar.



Gambar 7. Hasil Clustering K-Means

Berdasarkan dari gambar 7 tersebut merupakan eksekusi algoritma K-Means Clustering yang dilakukan dengan parameter jumlah klaster (k) = 3, terkonfirmasi adanya tiga kelompok utama dalam populasi UMKM, yang ditetapkan sebagai Klaster 0, Klaster 1, dan Klaster 2. Setiap klaster memiliki karakteristik berbeda yang mencerminkan tingkat perkembangan usaha serta kapasitas tenaga kerja dari masing-masing kelompok.

1. Cluster 0 terdiri dari 4.273 data dan menunjukkan bahwa rata-rata laba sebesar 45,11% lebih tinggi dibandingkan rata-rata keseluruhan. Namun, jumlah tenaga kerja laki-laki lebih rendah sebesar 38,02%, sedangkan tahun berdiri juga sedikit lebih kecil, yakni 3,90% lebih muda dibandingkan rata-rata. Hal ini mengindikasikan bahwa kelompok ini merupakan UMKM dengan efisiensi tinggi, karena mampu menghasilkan laba besar meskipun memiliki jumlah tenaga kerja yang lebih sedikit.
2. Cluster 1, dengan 4.256 data, memperlihatkan karakteristik bahwa laba rata-rata 46,29% lebih kecil dan tenaga kerja laki-laki 35,50% lebih sedikit dari rata-rata keseluruhan, sedangkan tenaga kerja perempuan sedikit lebih tinggi (3,92% lebih besar). Klaster ini menggambarkan UMKM berskala kecil yang masih dalam tahap pengembangan, dengan

sumber daya terbatas dan produktivitas yang masih rendah.

3. Sementara itu, Cluster 2 mencakup 5.035 data dan didominasi oleh UMKM dengan jumlah tenaga kerja laki-laki 62,27% lebih besar dibandingkan rata-rata. Namun, tenaga kerja perempuan sedikit lebih rendah (2,24% lebih kecil) dan tahun berdiri hanya 1,09% lebih tua dibandingkan rata-rata. Berdasarkan karakteristik tersebut, kelompok ini dapat diidentifikasi sebagai UMKM skala besar dan padat karya, dengan potensi pertumbuhan tinggi karena kapasitas tenaga kerja dalam skala masif.

Berdasarkan sintesis seluruh temuan, dapat dikonfirmasi bahwa Cluster 1 mewakili UMKM kecil yang masih berkembang, Cluster 0 menggambarkan UMKM menengah yang efisien dan stabil, sedangkan Cluster 2 merepresentasikan UMKM besar yang produktif dan padat tenaga kerja. Temuan ini penting untuk memahami distribusi tingkat kematangan usaha UMKM, sehingga dapat menjadi dasar dalam perumusan strategi pembinaan dan kebijakan pengembangan sektor UMKM secara lebih terarah.

SIMPULAN

Berdasarkan hasil penelitian melalui implementasi algoritma K-Means Clustering dengan memanfaatkan platform RapidMiner, diperoleh tiga klaster utama UMKM ($k = 3$) dengan karakteristik yang berbeda: Cluster 0 terdiri dari 4.273 data, memiliki rata-rata laba Rp3.349.611, tenaga kerja laki-laki 30 orang, dan perempuan 49 orang. Klaster ini menggambarkan UMKM menengah yang efisien dan stabil dengan pengalaman usaha yang cukup lama. Cluster 1 terdiri dari 4.256 data, dengan rata-rata laba Rp6.830.327, serta tenaga kerja laki-laki 31 orang dan perempuan 51 orang. Kelompok ini termasuk UMKM kecil yang masih berkembang dan memerlukan dukungan manajerial maupun permodalan. Cluster 2 mencakup 5.035 data, memiliki rata-rata laba tertinggi Rp20.583.431, dengan tenaga

kerja laki-laki 79 orang, menandakan UMKM besar dan produktif dengan kapasitas operasional yang tinggi.

Secara keseluruhan, hasil centroid dan clustering menunjukkan bahwa algoritma K-Means mampu mengelompokkan UMKM berdasarkan kinerja finansial dan jumlah tenaga kerja secara akurat. Segmentasi ini memberikan dasar yang kuat bagi pemerintah dan lembaga pembina untuk merancang strategi pemberdayaan yang sesuai, seperti peningkatan kapasitas dan bantuan modal bagi UMKM pada klaster kecil agar dapat naik kelas ke tingkat yang lebih tinggi.

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan ke hadirat Allah SWT atas rahmat dan karunia-Nya sehingga penelitian ini dapat diselesaikan. Penulis menyampaikan rasa terima kasih yang tulus kepada segenap jajaran pimpinan dan civitas akademika Universitas Ibrahimy Situbondo atas dukungan fasilitas dan lingkungan akademik yang telah diberikan. Ucapan terima kasih yang mendalam juga penulis sampaikan kepada Bapak dan Ibu Dosen Program Studi Teknologi Informasi atas bimbingan, ilmu, serta arahan berharga yang telah diberikan selama proses penelitian dan penulisan karya ilmiah ini.

Tak lupa, penghargaan yang tulus penulis sampaikan kepada kedua orang tua tercinta yang senantiasa memberikan doa, dukungan moral, dan kasih sayang tanpa henti. Atas dukungan semua pihak, penulis berharap hasil penelitian ini dapat memberikan manfaat dan kontribusi positif bagi perkembangan ilmu pengetahuan di almamater tercinta.

DAFTAR PUSTAKA

Astri, d. F., informatika, m., barat, j., langsung, b., kecil, a., & dasar, a. S. (2024). *Clustering penduduk miskin menggunakan algoritma k-means pada wilayah jawa barat*. 8(2), 1548–1554.

Desy, r., ruli, f., & wuryanto, e. (2017). *Penerapan clustering k-means pada*

customer segmentation berbasis recency frequency monetary (rfm) (studi kasus : pt . Sinar kencana intermoda surabaya).

Fadel, d., maulana, f., harahap, a., fauzan, i., azriel, m., fadlan, m., & fansyuri, m. (2025). *Implementasi algoritma k-means clustering data penjualan pada warung sembako isan menggunakan*. 1(1), 7–11.

Oktavian, b. D., lumintu, i., & amar, s. (2025). *Optimalisasi strategi product bundling melalui pemetaan pola peminatan dan pola penjualan produk menggunakan k-means clustering dan apriori optimizing product bundling strategy through mapping product interest patterns and sales patterns using k-means clustering and apriori*. 8(1), 26–41.

Pt, p., multi, b., & cabang, s. (2023). *J- icon : jurnal informatika dan komputer implementasi algoritma k-means untuk segmentasi pelanggan j-icon : jurnal informatika dan komputer*. 11(2). <https://doi.org/10.35508/jicon.v11.i2.12478>

Rapidminer, m. D. (2024). *Data Mining klasterisasi customer segmantation netflix menggunakan metode k-*. 6(2), 18–23.

Technology, i., sanjaya, i., inonu, n. E., fudholi, m. F., pratiwi, a., & sulistiani, h. (2025). *Segmentasi produk fashion berdasarkan harga , ukuran , dan merek menggunakan k-means di rapidminer*. 5(3), 398–405. <https://doi.org/10.47065/jimat.v5i3.651>

Umk, m., kecamatan, d. I., kota, p., aristika, w., & hartono, w. J. (2020). *Data mining*. 11, 2389–2395.

Z. Fatah. (2025). *Tik dan masyarakat (cetakan pe)*. Pt penamuda media.

Zahra, s. (2022). “defenisi, kriteria dan konsep umkm” makalah ini disusun guna memenuhi tugas kelompok. *Universitas islam negeri alauddin makassar, defenisi, kriteria dan konsep umkm*, 3.