

ANALISIS DATA MINING PENGELOMPOKAN UMKM BERDASARKAN JENIS USAHA DI PROVINSI JAWA BARAT MENGUNAKAN K-MEANS

Anindya Daniswara¹, Yuliazmi^{2*}

^{1,2} Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ¹2012500670@student.budiluhur.ac.id, ^{2*}yuliazmi@budiluhur.ac.id
(* : corresponding author)

Abstrak- Jawa Barat mengalami peningkatan perekonomian khususnya di sektor UMKM. Merujuk pada Portal JabarProv Kab. Sumedang 2023, Jawa Barat mengalami pencapaian pertumbuhan ekonomi pada sektor UMKM. Selanjutnya Ridwan Kamil menyatakan bahwa pencapaian tersebut tidak terlepas dari kontribusi Menteri Koordinator Bidang Perekonomian, Airlangga Hartarto, yang berhasil mengimplementasikan arahan Presiden Joko Widodo melalui kinerjanya. "Sekitar 80 persen ekonomi di Jawa Barat didorong oleh UMKM," ujar Ridwan Kamil. Tujuan pada penelitian ini untuk mengklasterisasi Jenis Usaha UMKM di Jawa Barat sehingga dapat mengetahui jenis usaha apa saja yang sedang meningkat di Jawa Barat. Data yang digunakan adalah data UMKM di Jawa Barat pada tahun 2019 sampai 2023. Metode penelitian yang dipakai adalah *K-Means Clustering*. Hasil penelitian menunjukkan terdapat 2 *cluster* yang paling optimal dengan nilai DBI 0,830 yaitu *cluster* terbaik dari *cluster* yg lain. Penelitian ini mendapatkan hasil yaitu, *cluster* terbaik pada jenis usaha Craft dan Makanan dengan nilai jarak terdekat yaitu 0 dan 3. Hasil akhir penelitian ini didapatkan 2 *cluster* yaitu *cluster* 0 dan *cluster* 1 dengan nilai rata-rata keduanya 15,5. Informasi ini dapat membantu pelaku UMKM khususnya di bidang Craft dan Makanan untuk lebih meningkatkan strategi pemasaran terhadap usahanya tersebut. Selain itu, hasil penelitian ini juga dapat membantu Pemerintah Pusat di Jawa Barat khususnya Dinas Koperasi dan Usaha Kecil dalam pengambilan keputusan terhadap sektor UMKM.

Kata Kunci: Data Mining, Clustering, K-Means, UMKM

DATA MINING ANALYSIS FOR CLUSTERING UMKM BASED ON BUSINESS TYPES IN WEST JAVA PROVINCE USING K-MEANS

Abstract- West Java has experienced economic growth, particularly in the MSME sector. According to the Sumedang District Portal of JabarProv in 2023, West Java achieved significant economic growth in the MSME sector. This achievement, as noted by Ridwan Kamil, is largely due to the efforts of Coordinating Minister for Economic Affairs Airlangga Hartarto, who successfully interpreted and implemented the directives of President Joko Widodo. "Almost 80 percent of West Java's economy is driven by MSMEs." The objective of this research is to cluster MSME business types in West Java to identify which types of businesses are currently thriving in the region. The research data used includes MSME data from West Java between 2019 and 2023. The research method applied is *K-Means Clustering*. The results show that there are two clusters that are the most optimal, with a Davies-Bouldin Index (DBI) value of 0.830, indicating the best clusters among the others. The study found that the best clusters were in the Craft and Food business types, with the closest distance values of 0 and 3, respectively. The final results identified two clusters: Cluster 0 and Cluster 1, with an average value of 15.5. This information can help MSME owners, particularly those in the Craft and Food sectors, to enhance their marketing strategies. Additionally, these findings can assist the Central Government in West Java, especially the Department of Cooperatives and Small Enterprises, in making decisions regarding the MSME sector.

Keywords: Data Mining, Clustering, K-Means, MSME

1. PENDAHULUAN

Usaha mikro, kecil, dan menengah (UMKM) merupakan salah satu penopang utama perekonomian nasional yang perlu didukung, dilindungi, dan dikembangkan secara optimal melalui keterhubungan yang erat. Korporasi ekonomi publik ini harus diperhatikan tanpa mengabaikan peran perusahaan besar dan badan usaha milik negara. UMKM berpotensi menciptakan lapangan kerja kepada masyarakat. Usaha menengah adalah suatu jenis usaha yang termasuk dalam bidang ekonomi produksi, bukan bagian dari suatu perusahaan besar atau kecil

lainnya. Kriteria kekayaan bersih bagi usaha menengah ditentukan sesuai dengan peraturan perundang-undangan yang berlaku.[1]

Pada tahun 2012, UMKM di Indonesia mencetak tenaga kerja dalam jumlah yang besar, sekitar 85 hingga 107 juta orang. Di periode yang sama, jumlah wirausaha di Indonesia tercatat mencapai 56.539.560 unit. Saat ini, banyak UMKM yang muncul dan berkembang dengan pesat. Pertumbuhan UMKM perlu dipantau setiap tahunnya dengan pembaruan data untuk klasifikasi data yang baru.[2]

Penetapan kriteria UMKM diatur oleh UU No.20 Tahun 2008. Dalam Pasal 6 ayat (1), ayat (2), dan ayat (3), disebutkan bahwa nilai nominal kriteria tersebut dapat disesuaikan dengan perkembangan ekonomi melalui Peraturan Presiden. Kriteria UMKM terdiri dari tiga kategori, yaitu mikro, kecil, dan menengah. Penentuan kriteria UMKM didasarkan pada beberapa komponen dan atribut seperti: Nomor, Kecamatan, Desa, Nama Usaha, Nama Pemilik, Alamat, Telepon/Handphone, Jenis Usaha, Jumlah Karyawan, Aset, Pendapatan, Tahun Berdiri, dan standar sebagai merek. Penentuan kriteria UMKM ini membutuhkan waktu sekitar 12 bulan. Di sisi lain, data kriteria UMKM ini dapat digunakan untuk melakukan klasifikasi dengan algoritma klasifikasi yang tersedia, sehingga kriteria dapat ditentukan lebih cepat.[3]

Selain itu, tidak terdapat *database* yang mencatat aktivitas dan kebijakan yang ditujukan kepada UMKM. Oleh karena itu, diperlukan pengelompokan UMKM berdasarkan karakteristik unik dari masing-masing unit usaha. Dengan adanya pengelompokan ini, diharapkan para pembuat kebijakan, seperti pemerintah atau organisasi terkait, dapat memetakan karakteristik dan kebutuhan UMKM. Informasi mengenai klasterisasi UMKM sangat penting, dan solusi untuk memperoleh informasi tersebut adalah dengan menerapkan konsep data mining.[4]

Data mining adalah proses statistik matematika, kecerdasan buatan (*machine learning*) untuk menggali dan mengenali informasi penting dari berbagai *database* berukuran besar.[5] Untuk memenuhi kebutuhan manajemen, terdapat berbagai metode yang dapat digunakan, salah satunya adalah dengan memanfaatkan data perusahaan. *Data mining* dapat diklasifikasikan berdasarkan tugas yang dilakukannya, salah satunya adalah *clustering*. *Clustering* merupakan proses membagi objek data menjadi beberapa subset, yang disebut sebagai cluster. Objek dalam satu cluster memiliki kesamaan satu sama lain, namun berbeda dengan objek yang ada di cluster lain.[6] Salah satu algoritma yang digunakan dalam pengelompokan data adalah Algoritma *K-Means*. *K-Means* adalah metode penambangan data yang memungkinkan pembagian data menjadi beberapa cluster. Data yang memiliki karakteristik serupa akan dikelompokkan ke dalam satu cluster, sedangkan data yang memiliki karakteristik berbeda akan dikelompokkan ke dalam cluster lain.[7].

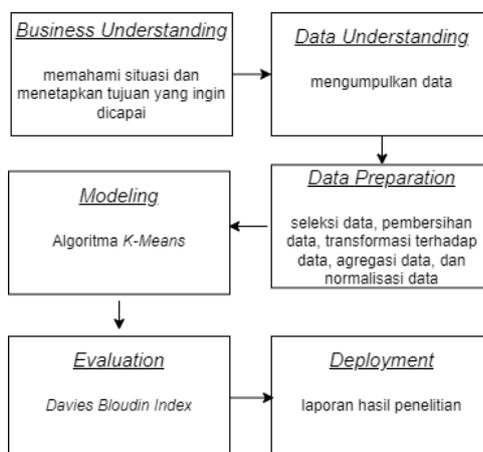
Penelitian terkait klasterisasi UMKM telah dilakukan menggunakan berbagai metode yaitu, “Penerapan Metode *K-Medoids* untuk Sintesis Data Usaha Mikro, Kecil dan Menengah (UMKM) Bidang Kuliner Kota Yogyakarta”[8], “Klasifikasi UMKM menggunakan algoritma *Naive Bayes* berdasarkan memiliki atau mempertahankan sertifikasi Halal”[9], “*Clustering* Jenis UMKM Berdasarkan Bidang Usaha di Kecamatan Tondano Selatan Kabupaten Minahasa Menggunakan Algoritma *K-Means Clustering*”[10]

Berdasarkan data yang diperoleh dari situs open data resmi Jawa Barat, pertumbuhan UMKM menunjukkan peningkatan yang signifikan di sektor UMKM menjadi usaha yang banyak diminati dan dapat meningkatkan keuangan setiap orang, serta memberikan kontribusi terhadap pembangunan perekonomian Provinsi Jawa Barat. Penelitian ini bertujuan untuk mengelompokkan UMKM di Jawa Barat berdasarkan jenis usaha di setiap kabupaten atau kota, untuk mengetahui jenis usaha yang paling diminati di setiap daerah. Tujuan Penelitian ini untuk mengklasifikasikan jenis-jenis usaha UMKM di Provinsi Jawa Barat dengan menerapkan teknik *data mining* menggunakan algoritma *clustering K-Means* dan melakukan evaluasi *clustering*. hasil menggunakan indeks *Davies-Bouldin*. Implikasi dari penelitian ini berpotensi besar untuk membantu para pemangku kepentingan UMKM memahami di mana sektor usaha mereka dikelompokkan, memberikan informasi baru kepada masyarakat tentang hasil dari pengelompokan ini dan mendukung pemerintah, baik daerah maupun pusat, untuk mengambil keputusan terkait dengan hal tersebut.

Algoritma *K-means* dipilih untuk penelitian *data mining* pada UMKM di Jawa Barat karena kemampuannya dalam mengelompokkan data ke dalam klaster-klaster yang *homogen*. Dengan *K-means*, pola-pola tersembunyi dalam data UMKM dapat ditemukan, membantu dalam pengambilan keputusan strategis yang lebih tepat sasaran. Algoritma ini juga relatif mudah diimplementasikan dan dapat diinterpretasikan dengan jelas, menjadikannya pilihan yang efisien dan efektif untuk analisis data UMKM.

2. METODE PENELITIAN

Penelitian ini menggunakan metode *CRISP-DM* adalah metode standar untuk pemrosesan data mining yang dikembangkan oleh analis dari berbagai industri.[11] untuk membantu mencari solusi dari suatu masalah. Untuk tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Business Understanding

Langkah pertama yang dilakukan adalah memahami situasi dan menetapkan tujuan yang ingin dicapai. Pada penelitian ini memiliki tujuan untuk mengelompokkan UMKM di Jawa Barat berdasarkan Jenis Usaha yang meningkat dari tahun ke tahun.

2.2 Data Understanding

Pada langkah ini dilakukan pengumpulan data UMKM di Jawa Barat yang berkualitas serta memahami data. Data yang digunakan adalah data UMKM 2019 sampai 2023 di Jawa Barat sebanyak 2.025 baris. Data tersebut diambil dari situs resmi Open Data Jabar.

2.3 Data Preparation

Di tahap ini dilakukan *Data Preprocessing* atau mempersiapkan data untuk dimodelkan, yang terdiri dari pemilihan data (*data selection*) dilakukan menggunakan *Ms Excel* untuk melakukan seleksi data, pembersihan data (*data cleaning*) menggunakan *python jupyter notebook data cleaning* dilakukan untuk mengecek ada tidaknya *missing value* dan *duplicated data* pada *dataset* yang digunakan, transformasi terhadap data (*datatransformation*) menggunakan *rapidminer* dengan metode transformasi yaitu *nominal to numeric* untuk mengubah tipe data *polynomial* menjadi *numeric* agar membantu dalam proses klusterisasi.

2.4 Modelling

Tahap ini penentuan metode yang dipakai dan implementasinya. Diawali dengan mengkomparasi model algoritma yaitu *k-means* dan *k-medoids* dengan jumlah *cluster* berdasarkan nilai DBI untuk mendapatkan *cluster* yang optimal yaitu bernilai mendekati nol atau kecil lalu mengimplementasikan model tersebut dengan perhitungan algoritma terpilih secara manual dan algoritma *k-means* pada *tools Rapidminer*.

2.5 Evaluation

Tahapan ini melakukan evaluasi atau pengujian kualitas dan apakah metode *K-Means Clustering* dapat memenuhi tujuan yang telah ditetapkan. Evaluasi dilakukan dengan membandingkan nilai DBI (*Davies Bouldin Index*) di beberapa jumlah *cluster*. Sehingga muncul perbandingan nilai DBI dari beberapa *cluster* dari hasil penyajian model dan didapatkan *cluster* terbaik dengan nilai DBI lebih kecil atau mendekati nol.

2.6 Deployment

Tahap terakhir melakukan pembuatan laporan hasil penelitian yang telah dilakukan. Dengan cara melakukan klusterisasi menggunakan algoritma terpilih dan melakukan perhitungan manual algoritma terpilih untuk mendapatkan hasil klusterisasi dan juga hasil akhir berupa nilai minimal, rata-rata, dan nilai maksimal dari hasil perhitungan manual.

3. HASIL DAN PEMBAHASAN

Hasil dari implementasi pada metode penelitian ini terdapat beberapa tahapan dengan penjabaran sebagai berikut :

3.1 Pengumpulan Data

Penelitian ini menggunakan data yang didapatkan melalui perantara atau tidak langsung. Data tersebut diambil dari situs resmi Open Data Jabar. Data yang digunakan adalah data UMKM yang terjadi pada tahun 2019 sampai 2023 di Jawa Barat sebanyak 2.025 baris dengan membahas Jenis Usaha UMKM, Jumlah UMKM, dan Tahun UMKM. Untuk bentuk tabulasi dari data asli dapat dilihat pada Tabel 1.

Tabel 1. Dataset Asli

Kode Provinsi	Nama Provinsi	Kode Kabupaten Kota	Nama Kabupaten Kota	Jenis Usaha	Jumlah Umkm	Satuan	Tahun
32	JAWA BARAT	3201	KABUPATEN BOGOR	AGRIBISNIS	9	UNIT	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	AKSESORIS	0	UNIT	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	BATIK	0	UNIT	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	BORDIR	0	UNIT	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	CRAFT	30	UNIT	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	DEKORASI	0	UNIT	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	MEBEL	0	UNIT	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	FASHION	13	UNIT	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	INDUSTRI	10	UNIT	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	JASA	11	UNIT	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	KONVEKSI	1	UNIT	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	KULINER	49	UNIT	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	MAKANAN	33	UNIT	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	MINUMAN	8	UNIT	2019

Tabel 1 diatas memuat dataset UMKM yang digunakan untuk penelitian.

3.2 Data Preprocessing

Tahapan pertama yang dilakukan adalah menyeleksi atribut yang berkesinambungan dengan topik atau masalah. Data asli memiliki 9 atribut lalu di seleksi menjadi 3 atribut untuk diteliti. Tampilan data menggunakan *Microsoft Excel* setelah dilakukan seleksi data dapat dilihat pada Tabel 2.

Tabel 2. Tampilan Data Terpilih

id	nama_kabupaten_kota	jenis_usaha	jumlah_umkm
1	KABUPATEN BOGOR	AGRIBISNIS	9
2	KABUPATEN BOGOR	AKSESORIS	0
3	KABUPATEN BOGOR	BATIK	0
4	KABUPATEN BOGOR	BORDIR	0
5	KABUPATEN BOGOR	CRAFT	30
6	KABUPATEN BOGOR	DEKORASI	0
7	KABUPATEN BOGOR	MEBEL	0
8	KABUPATEN BOGOR	FASHION	13
9	KABUPATEN BOGOR	INDUSTRI	10
10	KABUPATEN BOGOR	JASA	11
11	KABUPATEN BOGOR	KONVEKSI	1
12	KABUPATEN BOGOR	KULINER	49
13	KABUPATEN BOGOR	MAKANAN	33
14	KABUPATEN BOGOR	MINUMAN	8
15	KABUPATEN BOGOR	OBAT-OBATAN	0
16	KABUPATEN SUKABUMI	AGRIBISNIS	2
17	KABUPATEN SUKABUMI	AKSESORIS	2
18	KABUPATEN SUKABUMI	BATIK	0
19	KABUPATEN SUKABUMI	BORDIR	1
20	KABUPATEN SUKABUMI	CRAFT	15

Tabel 2 diatas memuat data terpilih yang sudah diseleksi untuk penelitian.

Transformasi data adalah proses mengubah format atribut data untuk memenuhi kebutuhan analisis *data mining*. Pada penelitian ini konversi data dilakukan menggunakan *RapidMiner* dengan operator nominal ke numerik dalam bentuk tabel atau tabular. Hasil Transformasi dapat dilihat pada Tabel 3.

Tabel 3. Tampilan Hasil Transformasi

id	Nama_Kabupaten_Kota	Jenis_usaha	Jumlah_umkm
1	0	0	9
2	0	1	0
3	0	2	0
4	0	3	0
5	0	4	30
6	0	5	0
7	0	6	0
8	0	7	13
9	0	8	10
10	0	9	11
11	0	10	1
12	0	11	49
13	0	12	33
14	0	13	8
15	0	14	0
16	1	0	2
17	1	1	2
18	1	2	0
19	1	3	1
20	1	4	15

Tabel 3 diatas adalah hasil dari proses Transformasi.

Normalisasi data dalam data mining adalah proses transformasi data ke dalam skala atau rentang yang konsisten, sehingga berbagai fitur data dapat dibandingkan dengan adil dan algoritma pembelajaran mesin dapat bekerja lebih efektif. Berikut ini hasil dari Normalisasi Data menggunakan *Rapidminer* dengan *Operator Normalize* dalam bentuk tabulasi atau *table*. Hasil Normalisasi dapat dilihat pada Tabel 4.

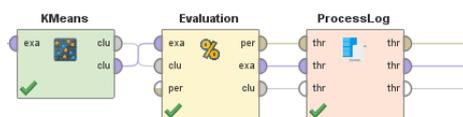
Tabel 4. Tampilan Hasil Normalisasi

id	Nama_Kabupaten_Kota	Jenis_usaha	Jumlah_umkm
1	0	0,000	0,045
2	0	0,071	0,000
3	0	0,143	0,000
4	0	0,214	0,000
5	0	0,286	0,149
6	0	0,357	0,000
7	0	0,429	0,000
8	0	0,500	0,065
9	0	0,571	0,050
10	0	0,643	0,055
11	0	0,714	0,005
12	0	0,786	0,244
13	0	0,857	0,164
14	0	0,929	0,040
15	0	1,000	0,000
16	0	0,000	0,010
17	0	0,071	0,010
18	0	0,143	0,000
19	0	0,214	0,005

Tabel 4 diatas adalah hasil dari proses Normalisasi.

3.2.1 Penyajian Model Terbaik

Nilai DBI atau *Davies Bouldin Index* didapatkan menggunakan *tools RapidMiner* dengan menggunakan operator *multiply* lalu ditarik garis ke 3 operator *clustering* dengan metode *K-Means* yang masing-masing operator memiliki nilai k yang berbeda yaitu 2, 3, dan 4. Selanjutnya masing-masing operator tersebut ditarik garis ke operator *Cluster Distance Performance* dengan *main criterion* atau kriteria utama *Davies Bouldin*. Untuk tampilan proses pada *tools RapidMiner* dapat dilihat pada Gambar 2.



Gambar 2. Penyajian Model Terbaik

Setelah proses tersebut, maka dapat diketahui jumlah *cluster* dari hasil penyajian model terbaik dapat dilihat pada Tabel 5.

Tabel 5. Nilai Penyajian Model Terbaik

Jumlah Cluster	Nilai DBI
2 cluster	0,830
3 cluster	1,066
4 cluster	0,883

Tabel 5 diatas adalah hasil dari Penyajian Model Terbaik menggunakan *K-Means* dan didapat Nilai 3 Cluster menurut DBI.

3.2.2 Perhitungan Manual Algoritma Terpilih

Pada tahap ini dilakukan perhitungan dengan menggunakan algoritma yang dipilih, khususnya algoritma *K-Means*. Perhitungan data dilakukan setelah transformasi sehingga data tersebut dapat diolah menggunakan algoritma clustering *K-Means*. Perhitungan dilakukan menggunakan bahan data yang terdiri dari 6 titik data. Data yang diproses mencakup atribut seperti Nama_Kabupaten Kota, Tipe_Bisnis, dan Nomor_Umkm. Data yang akan diolah dapat dilihat pada Tabel 6.

Tabel 6. Data Uji Terpilih

ID	nama_kabupaten_kota	jenis_usaha	jumlah_umkm
15	0	1,000	0,000
16	0,038	0,000	0,010
119	0,269	0,929	0,025
375	0,923	1,000	0,000
376	0,962	0,000	0,000
565	0,385	0,643	0,010

Tabel 6 diatas memuat Data Uji Terpilih yang digunakan untuk Perhitungan Manual Algoritma Terpilih.

Setelah data di transformasi menjadi numerik maka selanjutnya melakukan tahap berikut :

1. Menentukan Jumlah *cluster*, pada penelitian ini memilih 2 Cluster (k=2) yang akan dibuat.
2. Menentukan *centroid* tiap *cluster*, metode yang digunakan adalah mengambil data dari data yang tersedia secara acak. Berikut data *Centroid* tersebut.

Tabel 7. Titik Pusat Awal Tiap *Cluster*

Centroid Awal	nama_kabupaten_kota	jenis_usaha	jumlah_umkm
C1	0,538	0,929	0,040
C2	0,885	1	0,179

Tabel 7 diatas adalah Titik Pusat Awal Tiap Cluster yang digunakan untuk menghitung jarak data dengan centroid.

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

Berikut jarak data 1 dengan *centroid* 1:

$$\begin{aligned} D(1,1) &= \sqrt{(0.000-0.538)^2+(1.000-0.929)^2+(0.000-0.040)^2} \\ &= 0,543 \end{aligned}$$

Selanjutnya menghitung jarak data 1 dengan *centroid* 2:

$$\begin{aligned} D(1,2) &= \sqrt{(0.000-0.885)^2+(1.000-1.000)^2+(0.000-0.179)^2} \\ &= 0,904 \end{aligned}$$

Berikut jarak data 2 dengan *centroid* 1:

$$\begin{aligned} D(2,1) &= \sqrt{(0.038-0.538)^2+(0.000-0.929)^2+(0.010-0.040)^2} \\ &= 1,055 \end{aligned}$$

Selanjutnya menghitung jarak data 2 dengan *centroid* 2:

$$\begin{aligned} D(2,2) &= \sqrt{(0.038-0.885)^2+(0.000-1.000)^2+(0.010-0.179)^2} \\ &= 1,32 \end{aligned}$$

Berikut jarak data 3 dengan *centroid* 1:

$$\begin{aligned} D(3,1) &= \sqrt{(0.269-0.538)^2+(0.929-0.929)^2+(0.025-0.040)^2} \\ &= 0,268 \end{aligned}$$

Selanjutnya menghitung jarak data 3 dengan *centroid* 2:

$$\begin{aligned} D(3,2) &= \sqrt{(0.269-0.885)^2+(0.929-1.000)^2+(0.025-0.179)^2} \\ &= 0,638 \end{aligned}$$

Berikut jarak data 4 dengan *centroid* 1:

$$\begin{aligned} D(4,1) &= \sqrt{(0.923-0.538)^2+(1.000-0.929)^2+(0.000-0.040)^2} \\ &= 0,393 \end{aligned}$$

Selanjutnya menghitung jarak data 4 dengan *centroid* 2:

$$\begin{aligned} D(4,2) &= \sqrt{(0.923-0.885)^2+(1.000-1.000)^2+(0.000-0.179)^2} \\ &= 0,183 \end{aligned}$$

Berikut jarak data 5 dengan *centroid* 1:

$$\begin{aligned} D(5,1) &= \sqrt{(0.962-0.538)^2+(0.000-0.929)^2+(0.000-0.040)^2} \\ &= 1,022 \end{aligned}$$

Selanjutnya menghitung jarak data 5 dengan *centroid* 2:

$$\begin{aligned} D(5,2) &= \sqrt{(0.962-0.885)^2+(0.000-1.000)^2+(0.000-0.179)^2} \\ &= 1,019 \end{aligned}$$

Berikut jarak data 6 dengan *centroid* 1:

$$\begin{aligned} D(6,1) &= \sqrt{(0.385-0.538)^2+(0.643-0.929)^2+(0.010-0.040)^2} \\ &= 0,326 \end{aligned}$$

Selanjutnya menghitung jarak data 6 dengan *centroid* 2:

$$\begin{aligned}
 D(6,2) &= \sqrt{(0.385-0.885)^2 + (0.643-1.000)^2 + (0.010-0.179)^2} \\
 &= 0,637
 \end{aligned}$$

Hasil Perhitungan Jarak dapat dilihat pada Tabel 8.

Tabel 8. Hasil Perhitungan Setiap Data ke Setiap Pusat *Cluster*

ID	C0	C	Jarak Terdekat
		1	
15	0,543	0,904	0,543
16	1,055	1,32	1,055
119	0,268	0,638	0,268
375	0,393	0,183	0,183
376	1,022	1,019	1,019
565	0,326	0,637	0,326

Tabel 8 diatas adalah hasil perhitungan setiap data ke setiap pusat cluster.

Hasil *Cluster* menurut titik terdekat sesuai dengan hasil *Cluster* menggunakan *Rapidminer* dapat dilihat pada Tabel 9.

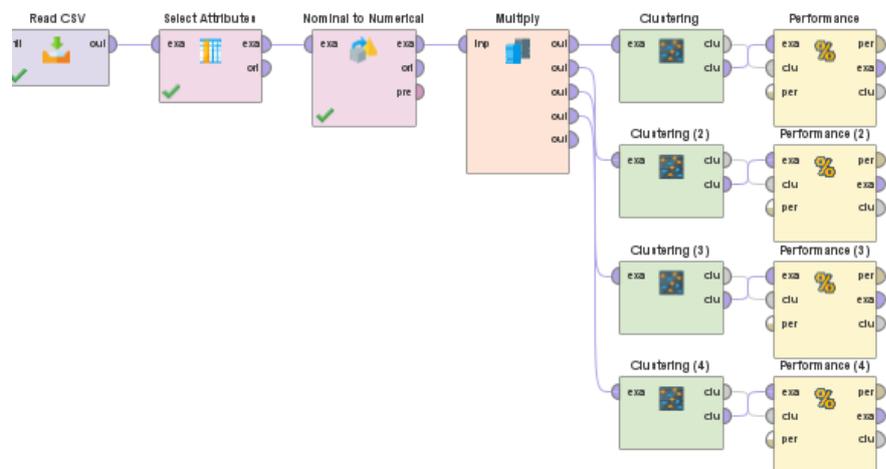
Tabel 9. Cluster Menurut Jarak Terdekat

ID	C0	C1	Jarak Terdekat	Cluster
15	0,543	0,904	0,543	C0
16	1,055	1,32	1,055	C0
119	0,268	0,638	0,268	C0
375	0,393	0,183	0,183	C1
376	1,022	1,019	1,019	C1
565	0,326	0,637	0,326	C0

Tabel 9 diatas adalah hasil perhitungan cluster menurut jarak terdekat.

3.3 Pengujian

Tahapan *evaluation* atau pengujian pada klasterisasi dilakukan dengan membandingkan nilai *DBI* (*Davies Bouldin Index*) algoritma *K-Means* pada jumlah 2 *cluster* hingga 5 *cluster*. Tahapan pengujian dapat dilihat pada Gambar 3.



Gambar 3. Pengujian pada Rapidminer

Berdasarkan pernyataan tersebut, nilai DBI paling kecil berada di jumlah 2 cluster sebesar 0,830 dapat dilihat pada Tabel 5 Nilai Penyajian Model Terbaik dan jumlah cluster tersebut dipakai untuk penelitian ini. Jadi, dapat disimpulkan jumlah cluster yang dipakai pada penelitian ini yang paling optimal dibanding dengan cluster lain dan dengan metode lain karena nilai dari 2 cluster adalah yang terbaik karena mendekati nilai 0 dan 2 cluster tersebut didominasi jenis usaha yang beraneka ragam khususnya di Kabupaten Bogor dengan jenis usaha craft dan makanan dengan jumlah UMKM yang lumayan banyak juga.

4. KESIMPULAN

Berdasarkan hasil penelitian dengan *RapidMiner* menggunakan algoritma *K-Means* untuk menganalisis perkembangan UMKM berdasarkan jenis usaha didapatkan nilai *Davies Bouldin Index* paling kecil atau mendekati nol dengan penentuan jumlah cluster terbaik adalah berjumlah 2 cluster yaitu 0,830 dan Hasil dari perhitungan manual didapat cluster menurut jarak terdekat didominasi oleh cluster C0. Kesimpulan yang dapat diambil oleh penulis terkait penelitian Perkembangan UMKM berdasarkan Jenis Usaha di Jawa Barat didapatkan hasil jenis usaha UMKM yang sedang meningkat yaitu Craft dan Makanan, Data ini didapat dari hasil klasterisasi menggunakan *k-means* dan melakukan perhitungan manual algoritma terpilih, dan didapat hasil yang dominan pada jenis usaha Craft dan Makanan. Dengan ini, Informasi ini dapat membantu pelaku UMKM khususnya owner UMKM di bidang Craft dan Makanan untuk lebih meningkatkan strategi pemasaran terhadap usahanya tersebut. Selain itu, hasil informasi ini juga dapat membantu Pemerintah Pusat di Jawa Barat khususnya Dinas Koperasi dan Usaha Kecil dalam pengambilan keputusan terhadap sektor UMKM.

5. SARAN

Penelitian ini dapat dikembangkan kembali dengan atribut, metode, dan studi kasus lainnya yang berbeda untuk memperoleh hasil yang lebih baik dan relevan. Dengan adanya hasil clustering yang dihasilkan dapat memberi masukan bagi Pemerintah Pusat khususnya Dinas Koperasi dan Usaha Kecil untuk lebih fokus memantau perkembangan ekonomi khususnya di sektor UMKM agar lebih meningkat lagi di Provinsi Jawa Barat.

DAFTAR PUSTAKA

- [1] L. Auliyah Alfiyah, "Prosedur Pelaporan Spt Tahunan Melalui E-Filing Di Kantor Pelayanan Penyuluhan Dan Konsultasi Perpajakan Majenang," 2021, pp. 22–52, 2021, [Online]. Available: L Aulia Alfiyah - 2021 - repositori.unsil.ac.id
- [2] N. Ameliana, N. Suarna, and W. Prihartono, "Analisis Data Mining Pengelompokkan Umkm Menggunakan Algoritma K-Means Clustering Di Provinsi Jawa Barat," *JATI (Jurnal Mhs. ...)*, vol. 8, no. 3, pp. 3261–3268, 2024, [Online]. Available:

- <https://ejournal.itn.ac.id/index.php/jati/article/view/9655><https://ejournal.itn.ac.id/index.php/jati/article/download/9655/5503>
- [3] T. Hardoyo and E. H. P. Eko, “Klasifikasi Usaha Mikro Kecil Menengah Menggunakan Jaringan Syaraf Tiruan Backpropagation,” *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 111–123, 2022, doi: 10.24002/konstelasi.v2i1.5625.
- [4] A. Azzam, A. Irma Purnamasari, and I. Ali, “Implementasi Algoritma K-Means Clustering Untuk Analisis Persebaran Umkm Di Jawa Barat,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 3062–3070, 2024, doi: 10.36040/jati.v8i3.8450.
- [5] Dodi Nofri Yoliad, “Data mining Dalam Analisis Tingkat Penjualan Barang Elektronik Menggunakan Algoritma K-means,” *Insearch (Information Syst. Res. J.*, vol. 3, no. 1, 2023.
- [6] W. Sudrajat, I. Cholid, and J. Petrus, “Penerapan Algoritma K-Means Clustering untuk Pengelompokan UMKM Menggunakan Rapidminer,” *J. JUPITER*, vol. 14, no. 1, pp. 27–36, 2022.
- [7] N. F. Adani *et al.*, “Implementasi Data Mining Untuk Pengelompokan Data Penjualan Berdasarkan Pola Pembelian Menggunakan Algoritma K-Means Clustering Pada Toko Syihan,” *J. Cyber Tech*, vol. x. No.x, no. x, pp. 1–11, 2019, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jct/article/view/4648><https://ojs.trigunadharma.ac.id/index.php/jct/article/download/4648/791>
- [8] U. Linarti, A. H. Soleliza Jones, L. Zahrotun, and A. Rahmawati, “Penerapan Metode K-Medoids Guna Pengelompokan Data Usaha Mikro, Kecil dan Menengah (UMKM) Bidang Kuliner Di Kota Yogyakarta,” *J. Ilmu Komput. dan Sist. Inf.*, vol. 7, no. 1, pp. 37–45, 2024, doi: 10.55338/jikomsi.v7i1.2194.
- [9] G. Wijaya, “Klasifikasi UMKM Menggunakan Algoritma Naive Bayes Berdasarkan Sudah Pernah Mempunyai Atau Mengurus Sertifikat Halal,” *J. Data Min. dan Sist. Inf.*, vol. 4, no. 1, p. 36, 2023, doi: 10.33365/jdmsi.v4i1.2634.
- [10] R. Ramba and V. P. Rantung, “... Berdasarkan Bidang Usaha Pada Kecamatan Tondano Selatan Kabupaten Minahasa Menggunakan Algoritma K-Means Clustering (Studi Kasus: Dinas UMKM dan ...,” *Innov. J. Soc. Sci. ...*, vol. 3, no. 19, pp. 6739–6747, 2023, [Online]. Available: <http://j-innovative.org/index.php/Innovative/article/view/7255><http://j-innovative.org/index.php/Innovative/article/download/7255/4932>
- [11] S. E. Damayanti and S. Kuswayati, “Analisis Dan Implementasi Framework CRISP-DM (Cross Industry Standard Process for Data Mining) Untuk Clustering Perguruan Tinggi Swasta,” *J. STT Bandung*, vol. 6, 2020.