ANALISIS DATA MINING PENGELOMPOKKAN UMKM MENGGUNAKAN ALGORITMA *K-MEANS CLUSTERING* DI PROVINSI JAWA BARAT

Nikan Ameliana [1], Nana Suarna [2], Willy Prihartono [3]
[1,2] Teknik Informatika, STMIK IKMI Cirebon
[3] Komputerisasi Akuntansi , STMIK IKMI Cirebon
Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135
nikanameliana40@gmail.com

ABSTRAK

Usaha Mikro, Kecil, dan Menengah (UMKM) memiliki peran penting dalam perekonomian Indonesia. Di Jawa Barat, terdapat banyak UMKM yang tersebar di berbagai wilayah. Mengetahui karakteristik dan persebaran UMKM menjadi penting untuk merumuskan kebijakan yang tepat dalam pengembangannya. Namun, persebaran UMKM di Jawa Barat tidak merata. Beberapa wilayah memiliki konsentrasi UMKM yang tinggi, sementara wilayah lain memiliki konsentrasi UMKM yang rendah. Penelitian ini bertujuan untuk menganalisis pengelompokan UMKM di Jawa Barat berdasarkan karakteristiknya menggunakan algoritma *K-Means Clustering*. Proses data mining dimulai dengan melakukan proses *Knowledge Discovery in Database* (KDD). Hasil penelitian menunjukkan nilai yang baik dalam *Davies Bouldin Index* adalah semakin kecil atau mendekati nol, terdapat jumlah *Cluster* yang memiliki nilai DBI terbaik adalah *Cluster* 2 yang terbaik, dengan nilai DBI mendekati nol yaitu 0.527. Hasil pengujian menghasilkan 10 anggota *Cluster*, yaitu *Cluster* 0 dengan hasil berjumlah 104 items, *Cluster* 1 dengan hasil berjumlah 24 item, *Cluster* 2 dengan hasil berjumlah 45 items, *Cluster* 3 dengan hasil berjumlah 302 items, *Cluster* 4 dengan hasil berjumlah 1 items, *Cluster* 5 dengan hasil berjumlah 50 Items, *Cluster* 6 dengan hasil berjumlah 282 items, *Cluster* 7 dengan hasil berjumlah 7 items, *Cluster* 8 dengan hasil berjumlah 304 items, *Cluster* 9 dengan hasil berjumlah 96 *items*.

Kata kunci : *UMKM, Jawa Barat, K-Means Clustering, Pengelompokan, Karakteristik, Davies Bouldin Index*

1. PENDAHULUAN

UMKM di Indonesia telah berhasil menyerap tenaga kerja dalam jumlah yang signifikan hingga tahun 2012, yaitu sekitar 85 juta hingga 107 juta orang. Pada periode yang sama, jumlah pengusaha di Indonesia mencapai 56.539.560 unit. Dari jumlah tersebut, sekitar 99,99% merupakan UMKM, sedangkan sisanya sekitar 0,01% atau sebanyak 4.968 unit merupakan usaha berskala besar [1]. Data ini menegaskan bahwa Usaha Mikro, Kecil dan Menengah (UMKM) memiliki potensi yang besar dalam mendukung pembangunan ekonomi di Indonesia. Meskipun UMKM di Jawa Barat menunjukkan potensi yang tinggi dalam menyerap tenaga kerja, namun UMKM masih dihadapkan pada sejumlah kendala yang belum terselesaikan. Karena keberadaan UMKM mampu bertahan dalam situasi apapun untuk tercapainya kesejahteraan masyarakat, maka secara teoritis dengan banyak nya umkm sehingga perlu dikelompokkan berdasarkan jenis usahanya untuk mengetahui UMKM mana yang paling banyak diminati dapat memberikan referensi terhadap pemerintah Jawa Barat dalam melakukan pemetaan UMKM.

Berdasarkan data yang diperoleh dari website resmi open data Jabar bahwa perkembangan UMKM mengalami peningkatan yang cukup signifikan menunjukan bahwa tren pelaku usaha bidang UMKM menjadi salah satu pilihan bidang usaha kekinian yang dapat meningkatkan taraf hidup perekonomian masyrakat pada khususnya dan merupakan perkembangan ekonomi pada skala Provinis Jawa barat pada umumnya pada penelitian ini melakukan pengelompokan jumlah UMKM di Jawa Barat berdasarkan jenis usaha di kabupaten atau kota sehingga dapat terlihat kelompok usaha apa saja yang paling diminati masyarakat pada kabupaten atau kota tersebut. Oleh karena itu Tujuan dari penelitian ini untuk mengetahui hasil mengklaster UMKM yang dihasilkan dengan menggunakan *K-Means Clustering* sehingga dapat mengetahui nilai *Davies Bouldin Index* DBI pada masing-masing klaster dan jarak titik pusat antar klaster nya. Penggunaan algoritma *K-Means Clustering* pada penelitian dimana algoritma tersebut banyak dipakai oleh para peneliti sebagai proses klasterisasi untuk menentukan pengelompokkan umkm.

Alasan penggunaan Algoritma *K-Means* adalah salah satu algoritma *Machine Learning* yang sederhana dan populer digunakan untuk memecahkan masalah pengelompokan data penggunaan algoritma *K-Means* diantaranya ialah karena algoritma ini memiliki ketelitian yang cukup tinggi terhadap ukuran objek, sehingga algoritma ini relatif lebih terukur dan efisien untuk pengolahan objek dalam jumlah besar [2]. Analisis data yang digunakan *Knowledge Discovery in Database* (KDD). Data mining sendiri adalah bagian dari tahapan proses KDD. Tahapan proses KDD meliputi seleksi data, *preprocessing data*, transformasi data, data *mining*, dan dan evaluasi.

2. TINJAUAN PUSTAKA

Dalam bagian ini, disajikan tinjauan pustaka yang memuat literatur ilmiah relevan dengan

penelitian yang sedang dilakukan. Literatur ilmiah ini akan menjadi fondasi teoretis bagi penelitian ini.

2.1. Data *Mining*

Data *mining*, atau pengembangan data, adalah sebuah metode untuk menemukan pola dan tren tersembunyi dari sejumlah besar data. Prosesnya dimulai dengan pemilihan data yang relevan, kemudian pencarian pola dan tren dengan menggunakan berbagai teknik statistik dan komputasi, dan akhirnya pemodelan data untuk memprediksi dan memahami fenomena yang terjadi. Data mining juga dikenal sebagai *Knowledge Discovery In Database* (KDD) atau pengenalan pola. KDD merujuk pada proses penemuan pengetahuan dari data. Tujuan utama data mining adalah memanfaatkan data dalam database untuk diolah dan menghasilkan informasi baru yang bermanfaat [3].

2.2. *Clustering*

Clustering adalah suatu proses di mana data dikelompokkan ke dalam beberapa kelompok (*Cluster*). Dalam proses ini, data yang tergabung dalam satu kelompok memiliki kesamaan karakteristik satu sama lain dan berbeda dengan data yang tergabung dalam kelompok lainnya [4]. Potensi dari menggunakan *Clustering* adalah dapat digunakan mengetahui struktur-struktur yang berada dalam data dan dapat digunakan dalam berbagai aplikasi seperti, pengenalan pola, pengolahan gambar serta klasifikasi [5].

2.3. Algoritma *K-Means*

K-Means adalah salah satu metode pengelompokan data yang populer. Metode ini membagi data menjadi beberapa kelompok (disebut *Cluster*) berdasarkan kesamaan karakteristik antar data. Prosesnya dimulai dengan menentukan jumlah *Cluster* yang diinginkan. Kemudian, algoritma *K-Means* akan menghitung jarak antara setiap data dan pusat *Cluster*. Data akan dikelompokkan ke dalam *Cluster* dengan jarak terdekat. Algoritma *K-Means* akan terus mengulang proses ini sampai tercapai kondisi stabil, di mana data dalam *Cluster* tidak lagi berpindah-pindah [6]. Kemampuan *K-Means* dalammengelompokkan data dalam jumlah besar dengan waktu pemrosesan yang relatif cepat dan efisien telah diakui [5]. Metode ini mengelompokkan data dengan karakteristik yang sama [7]. Berikut adalah tahapan pengelompokannya:

a. Menentukan Jumlah *Cluster* (k) harus dipilih terlebih dahulu, dan tidak boleh melebihi jumlah data.
b. Inisialisasi Pusat *Cluster*, Nilai awal pusat *Cluster* dapat ditentukan secara acak.
c. Menempatkan Data pada *Cluster* Setiap data akan ditempatkan pada *Cluster* yang paling dekat. Kedekatan dihitung berdasarkan jarak antar objek dengan menggunakan rumus *Euclidean Distance*. Rumus tersebut dapat dilihat pada persamaan (1).

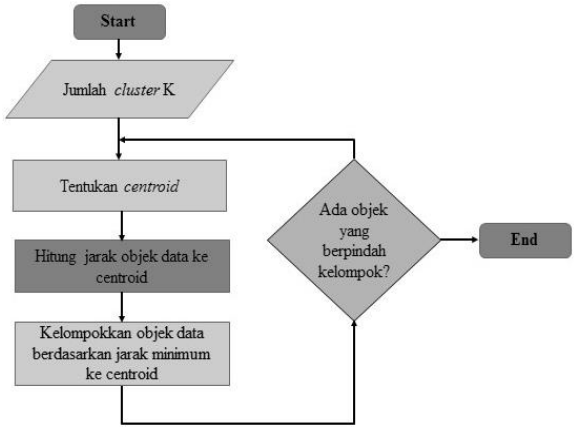
$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1)$$

Rumus jarak *Euclidean Distance*
Sumber Refrensi [7]

Keterangan :

d = Jarak
x_1 = Koordinat latitude 1
x_2 = Koordinat latitude 2
y_1 = Koordinat longitude 1
y_2 = Koordinat longitude 2

d. Menghitung dan Memperbarui Pusat *Cluster*. Data dikelompokkan berdasarkan jarak minimum. Pusat *Cluster* dihitung kembali dengan menggunakan data dalam *Cluster* baru.
e. Proses pengelompokan data dengan algoritma *K-Means Cluster*ing melibatkan perhitungan ulang jarak antar objek secara berulang. Jarak antar objek dihitung berdasarkan pusat *Cluster* yang baru, yang dihitung dari rata-rata data di dalam *Cluster*. Proses ini diulang terus-menerus hingga pusat *Cluster* tidak berubah lagi (konvergen). Gambar 1 menunjukkan langkah-langkah pengelompokan data dengan algoritma *K-Means Cluster*ing.



Gambar 1. Flowchart algoritma *k-means cluster*ing
sumber refrensi [6]

f. Menghitung *Davies Bouldin Index* (DBI), Setelah konvergen, nilai rata-rata setiap titik data dihitung dengan menggunakan *Davies Bouldin Index* (DBI) menggunakan persamaan (2).

2.4. Usaha Mikro, Kecil, dan Menengah (UMKM)

Usaha Mikro, Kecil, dan Menengah (UMKM) merupakan sektor usaha yang dimiliki oleh individu atau perorangan. Di dalam konteks perekonomian Indonesia, UMKM memegang peran yang sangat penting [8]. Salah satu manfaat yang paling nyata dari keberadaan UMKM adalah kemampuannya dalam menciptakan lapangan pekerjaan dengan modal yang terbatas [9].

2.5. *Davies Bouldin Indeks* (DBI)

Davies Bouldin Index (DBI) adalah metode validasi klaster yang diciptakan oleh D.L. Davies. DBI

bekerja dengan mengukur rasio distribusi data dalam klaster terhadap pemisahan antar klaster. Tujuannya adalah untuk mengoptimalkan jarak antar klaster. DBI digunakan untuk mengevaluasi validitas data di setiap klaster. Semakin kecil nilai DBI yang diperoleh (*non-negatif* ≥ 0), maka hasil pengelompokan data UMKM dengan model *Cluster*ing yang digunakan semakin baik. Nilai DBI dihitung untuk mendapatkan nilai K yang paling optimal, sehingga dapat disimpulkan Nilai DBI yang optimal dapat dicapai dengan mencari nilai DBI yang mendekati nol [10] [11], Rumus *Davies Bouldin Index* menggunakan persamaan (2).

$$DBI = \frac{1}{k}\Sigma_i^k = 1 \max_{i \neq j} \left(R_{ij}\right) \quad (2)$$

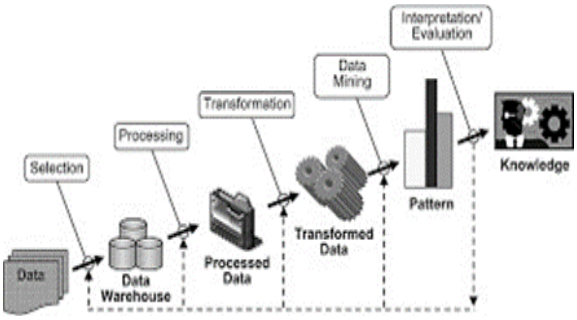
Rumus *Davies Bouldin Index*
Sumber Refrensi [12]

Keterangan :
K : jumlah *Cluster* yang digunakan
$R_{i,j}$ = rasio antara kluster i dan j
Max : dicari rasio antar kluster yang terbesar

Nilai DBI yang kecil, mendekati 0, menunjukkan kualitas pengelompokan data yang semakin baik. Nilai DBI dihitung menggunakan persamaan (2), untuk setiap jumlah *Cluster* yang diuji, dan jumlah *Cluster* dengan nilai DBI terkecil adalah yang optimal. Oleh karena itu, semakin kecil DBI, semakin baik hasil pengelompokan untuk jumlah *Cluster* yang telah dikelompokkan.

3. METODE PENELITIAN

Metode penelitian ini menggunakan eksperimen pendekatan kuantitatif. Pendekatan ini dipilih untuk melakukan analisis data secara sistematis dan objektif guna memastikan kesimpulan yang tepat dan dapat dipercaya. metode yang digunakan untuk menganalisis data adalah KDD (*Knowledge Discovery in Database*), proses tersebut dapat dilihat pada Gambar 2.



Gambar 2. Proses *knowledge discovery in database*

KDD melibatkan beberapa tahap, yaitu *Data Selection*, *Data Prepocessing*, *Data Transformation*, *Data Mining*, *Evaluation*. Metode penelitian eksperimental memungkinkan pengumpulan data melalui proses pengujian dan eksperimen. Data yang terkumpul kemudian dianalisis menggunakan algoritma *K-Means* untuk mengelompokkan UMKM berdasarkan karakteristiknya.

a. *Data Selection*
Data Selection merupakan tahap pemilihan data untuk memilih data yang akan digunakan dalam proses penggalian informasi atau data mining. Proses ini dilakukan setelah mendapatkan sekumpulan data operasional yang perlu diseleksi untuk memastikan bahwa data yang dipilih relevan dan sesuai dengan kebutuhan analisis. Atribut yang digunakan dalam proses data mining yaitu nama kabupaten atau kota, jenis UMKM, Jumlah UMKM dan tahun.
b. *Preprocessing*
Data Preprocessing adalah langkah awal dalam proses KDD, pada tahap ini, data yang tidak relevan, data yang hilang, dan data duplikat harus dibersihkan. Tahapan ini penting karena data yang relevan, tidak hilang dan tidak terduplikasi diperlukan untuk memulai proses data mining. Dalam penelitian ini perlu menyeleksi data agar tersetruktur.
c. *Transformation*
Transformation adalah tahapan ini bertujuan untuk mengubah data yang telah dipilih agar dapat digunakan untuk proses data mining. Proses transformasi dalam KDD sangat bergantung pada jenis atau pola informasi yang akan dicari di database.
d. *Data Mining*
Data Mining merupakan proses mencari pola atau informasi yang menarik di dalamnya, data dipilih menggunakan teknik atau metode tertentu berdasarkan kelengkapan proses KDD. Metode dalam penelitian ini adalah metode *Clustering* menggunakan Algoritma *K-Means*.
e. *Evaluation*
Pada tahap evaluasi, *RapidMiner* akan digunakan untuk mengetahui hasil performa dan akurasi algoritma *K-Means*. Hasil klaster yang ditemukan akan dievaluasi dan pola/informasi/pengetahuan dari hasil proses data mining akan ditampilkan dalam bentuk yang mudah dimengerti oleh pihak-pihak yang membutuhkan.

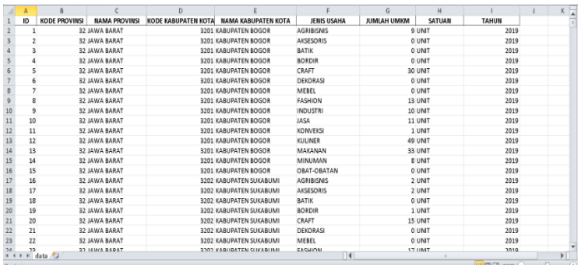
4. HASIL DAN PEMBAHASAN

Hasil penelitian yang dipaparkan dalam pembahasan ini akan menguraikan proses analisis dataset UMKM dengan menggunakan Algoritma *K-Means*. Pengelompokkan data dilakukan melalui proses pengujian menggunakan *RapidMiner* Studio. Penelitian ini mengikuti langkah - langkah proses data mining, yaitu: Data *Selection*, Data *Prepocessing*, Data *Transformation*, Data *Mining*, dan *Evaluation*.

4.1. Data

Persiapan data yang baik merupakan langkah pertama sebelum melakukan pemodelan *Cluster*ing di *RapidMiner*. Penelitian ini menggunakan dataset "Jumlah Usaha Mikro, Kecil, dan Menengah (UMKM) Binaan Berdasarkan Jenis Usaha di Jawa Barat". Dataset ini terdiri dari 1.215 data dengan 9 atribut, dan

mencakup periode tahun 2019-2021. Untuk menggunakan dataset tersebut di *RapidMiner*, diperlukan proses *import* data menggunakan operator yang tersedia. Salah satu operator yang dapat digunakan adalah *Read* CSV. Operator *Read CSV* berfungsi untuk membaca file CSV dan memasukkan data ke dalam *RapidMiner*. Gambar 3. menunjukkan dataset UMKM format file CSV untuk dilakukan proses mengimpor ke dalam *RapidMiner*.



Gambar 3. Dataset umkm

Pada operator *Read* CSV, digunakan parameter bawaan *RapidMiner* tanpa modifikasi. Operator ini berhasil mengimpor dan membaca file CSV berisi data mentah ke dalam program *RapidMiner*. Hasil pembacaan operator *Read* CSV disajikan pada Gambar 4.

Gambar 4. Operator *Read* csv

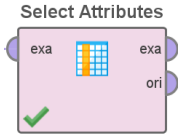
Operator membaca file CSV dengan menggunakan parameter bawaan. Berikut ini adalah informasi yang diperoleh dari hasil pembacaan file CSV, seperti yang terlihat pada Tabel 1.

Tabel 1. Statistik hasil operator *Read* csv

No	Uraian	Keterangan
1.	*Record*	1215
2.	*Special Attributes*	0
3.	*Regular Attributes*	9
4.	*Attributes:*	
	Nama Kabupaten Kota	*Nominal, missing* 0
	Jenis Usaha	*Nominal, missing* 0
	Jumlah UMKM	*Integer, missing* 0
	Tahun	*Integer, missing* 0
	Satuan	*Nominal, missing* 0
	Kode Kabupaten Kota	*Integer, missing* 0
	Nama Provinsi	*Nominal, missing* 0
	Id	*Integer, missing* 0
	Kode Provinsi	*Integer, missing* 0

4.2. Data Selection

Data selection merupakan tahap melibatkan pemilihan data yang relevan dari database. Hal ini penting karena tidak semua data diperlukan untuk menghasilkan model yang akurat dan bermanfaat. Dalam penelitian ini, Operator *Select Attributes* digunakan untuk memilih atribut tertentu, terdapat pada Gambar 5. Operator ini menggunakan filter jenis subset pada parameternya. Dataset yang digunakan untuk analisis data umumnya memiliki banyak atribut. Namun, tidak semua atribut relevan atau berkontribusi signifikan terhadap analisis. Oleh karena itu, diperlukan proses seleksi untuk memilih atribut yang tepat.



Gambar 5. *Select attributes*

Dataset pengelompokkan UMKM di Jawa Barat memiliki 9 atribut. Proses seleksi data akan memilih 4 atribut yang relevan untuk digunakan pada proses selanjutnya. Proses impor data dimulai dengan memilih file data Excel, menentukan penempatan data dalam dokumen, dan memasukkannya ke dalam proses data *RapidMiner*. Kemudian, atribut yang ingin digunakan dipilih, seperti nama kabupaten/kota, jenis usaha, jumlah UMKM, dan tahun. Pada operator *Select Attributes* terdapat parameter yang perlu disesuaikan. Parameter yang digunakan dalam operator ini dapat dilihat pada Tabel 2 sebagai berikut:

Tabel 2. Parameter select attributes

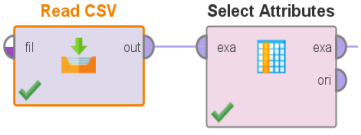
Parameter	Isi
Type	*Include Attributes*
Attribute filter type	*A Subset*
Select Attributes	Nama Kabupaten Kota, *Nominal, Missing* 0
	Jenis Usaha, *Nominal, Missing* 0
	Jumlah UMKM, Integer, *Missing* 0
	Tahun Integer, *Missing* 0

Dari hasil pembacaan operator *Select Attributes* yang telah dilakukan didapat informasi pada Tabel 3. sebagai berikut:

Tabel 3. Statistics Dataset

No	Uraian	Keterangan
1.	*Record*	1215
2.	*Special Attributes*	0
3.	*Regular Attributes*	4
4.	*Attributes:*	
	Nama Kabupaten Kota	*Nominal, missing* 0
	Jenis Usaha	*Nominal, missing* 0
	Jumlah UMKM	*Integer, missing* 0
	Tahun	*Integer, missing* 0

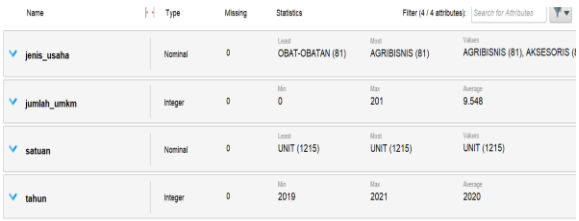
Proses pemodelan pada tahap *selection* di *RapidMiner*, dapat dilihat pada Gambar 6.



Gambar 6. Proses pemodelan tahap *selection*

4.3. Data *Preprocesing*

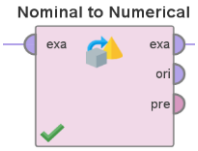
Pada tahap *Preprocessing* Data, peneliti memastikan bahwa data Usaha Mikro, Kecil, dan Menengah (UMKM) yang telah dipilih layak untuk diolah. Kegiatan yang dilakukan melibatkan pembersihan dan perbaikan data yang rusak, penghapusan data yang tidak digunakan, penyeragaman data yang dianggap sama namun memiliki nilai yang berbeda, serta konsistensi data yang memiliki nilai *Missing*. Hasil pemeriksaan menunjukkan tidak adanya data dengan kualitas buruk, hasil statistik data dapat dilihat pada Gambar 7.



Gambar 7. Tahap *preprocessing* data

4.4. Data Transformation

Transformasi data merupakan proses untuk mengubah format data atribut agar sesuai dengan kebutuhan analisis data mining. Pada penelitian ini, transformasi data dilakukan karena terdapat beberapa atribut dengan tipe data non-numerik yang perlu diubah menjadi bentuk numerik [13]. Hal ini dilakukan karena algoritma *K-Means Clustering* dan integrasi data memerlukan data numerik. Atribut *non-numerik* yang perlu diubah adalah: Jenis_usaha, Satuan, Kedua atribut ini kemudian ditransformasikan ke dalam bentuk numerik menggunakan operator *Nominal to Numerical*. Tampilan operator *Nominal to Numerical* dapat dilihat pada Gambar 8.



Gambar 8. Operator *nominal to numerical*

Operator *Nominal to Numerical* memiliki beberapa parameter yang perlu disesuaikan. Berikut adalah parameter - parameter tersebut beserta penjelasannya, seperti yang ditunjukkan pada Tabel 4.

Tabel 4. Parameter nominal to numerical

Parameter	Isi
Attribute filter type	*A Subset*
Coding type	*Unique Numericals*
Select Attributes	Jenis_usaha, *Numerical*, *missing* 0
	Satuan, *Numerical*, *missing* 0

Dari hasil pembacaan operator *Nominal to Numerical* yang telah dilakukan didapat informasi pada Tabel 5. sebagai berikut:

Tabel 5. Statistik dataset

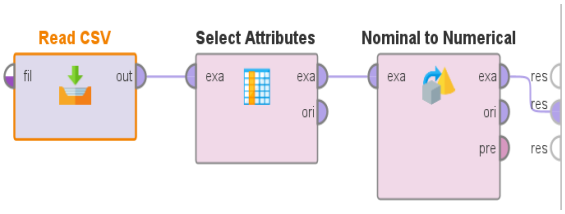
No	Uraian	Keterangan
1.	*Record*	1215
2.	*Special Attributes*	0
3.	*Regular Attributes*	4
4.	*Attributes:*	
	Jenis_usaha	*Numerical*, *missing* 0
	Satuan	*Numerical*, *missing* 0
	Nama_Kecamatan	*Numerical*, *missing* 0

Tabel 6. menunjukkan hasil dari penggunaan *Operator Nominal ke Numeric* pada data yang telah diolah. Tabel ini memuat informasi mengenai nilai data nominal yang telah diubah menjadi nilai numerik.

Tabel 6. Proses *operator nominal to numeric*

No	Nama Kabupaten kota	Jenis Usaha	Jumlah Umkm	Tahun
1	0	0	9	2019
2	0	1	0	2019
3	0	2	0	2019
4	0	3	0	2019
5	0	4	30	2019
6	0	5	0	2019
7	0	6	0	2019
8	0	7	13	2019
9	0	8	10	2019
10	0	9	11	2019
...	...	...	...	...
1206	26	5	2	2021
1207	26	6	3	2021
1208	26	7	9	2021
1209	26	8	5	2021
1210	26	9	9	2021
1211	26	10	0	2021
1212	26	11	4	2021
1213	26	12	64	2021
1214	26	13	14	2021
1215	26	14	0	2021

Gambar 9. di bawah ini menunjukkan proses transformasi data yang dilakukan dalam tahap Transformasi pada *RapidMiner*.



Gambar 9. Model proses transformasi data

4.5. Data Mining

Algoritma *K-Means Clustering* digunakan pada tahap data mining ini. Algoritma *K-Means Clustering* digunakan untuk melakukan pengelompokan pada dataset pengelompokkan jenis UMKM ke dalam jumlah klaster terbaik dengan nilai DBI (*Davies Bouldin Index*) teroptimal. Langkah selanjutnya adalah memasukan *operators Clustering* untuk pengelompokkan data menggunakan *algoritma K-*

Means Clustering pada *RapidMiner* seperti pada Gambar 10.

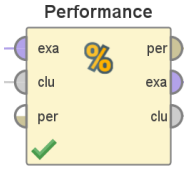
Gambar 10. *Operator k-means clustering*

Kemudian parameters yang akan digunakan dalam operator pada Tabel 7. berikut ini :

Tabel 7. Hasil proses operator *k-means clustering*

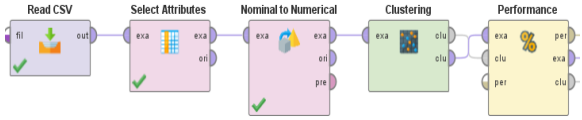
No	Parameter	Uraian
1	K (klaster yang ditentukan)	2
2	Max_run (maksimal perulangan)	10
3	*Measure Types*	*NumericalMeasures*
4	*Numerical Measure*	*EuclideanDistance*
5	*Max Optimization Steps*	10

Operator Performance kemudian ditambahkan untuk menilai kualitas pengelompokan (*Cluster Distance Performance*) menggunakan metode *Davies Bouldin Index* (DBI). Tujuannya adalah untuk mengetahui nilai DBI dari proses *Clustering* yang telah dilakukan. Gambar 11. menampilkan operator Performance yang digunakan.



Gambar 11. *Operator performance*

Gambar 12 mengilustrasikan proses pemodelan yang diterapkan dalam tahap Data *Mining* di *RapidMiner*.



Gambar 12. Model data *mining rapidminer*

Dari hasil pembacaan operator *K-Means Clustering* dengan parameter jenis *Numerical Measure* yaitu *Euclidean Distance* menggunakan persamaan (1), serta pembacaan operator *Cluster Distance Performance*, kemudian pada bagian *Main Criterion* pilih menggunakan *Davies Bouldin Index* (DBI) menggunakan persamaan (2), didapatkan informasi hasil DBI sebagai berikut:

Penulis melakukan percobaan K2 sampai dengan K10 dan maksimal iterasi 10 agar mendapatkan nilai K yang optimal untuk proses *K-Means Clustering* dan iterasi dilakukan 10 kali agar tidak ada perubahan yang signifikan terhadap posisi *centroid*. Pada penggunakan K2 dihasilkan bahwa Setelah menjalankan proses tersebut pengel`ompokkan dataset pada *RapidMiner* menghasilkan 2 *Cluster* model seperti pada Gambar 13.

Cluster Model

```
Cluster 0: 1116 items
Cluster 1: 99 items
Total number of items: 1215
```

Gambar 13. *Cluster* model

Selanjutnya melakukan percobaan dengan K=2 untuk mengkonfirmasi hasil *Davies Bouldin Index*, (DBI). Hasil percobaan ini dilakukan dengan membagi data menjadi dua *Cluster* yang berbeda. (DBI) yang dihasilkan sebesar 0.527 hasil percobaan pada K=2 seperti pada Gambar 14.

Davies Bouldin

```
Davies Bouldin: 0.527
```

Gambar 14. *Davies bouldin index* k=2

Selanjutnya melakukan percobaan dengan K=3 untuk mengkonfirmasi hasil *Davies Bouldin Index* (DBI). Hasil percobaan ini dilakukan dengan membagi data menjadi dua *Cluster* yang berbeda. (DBI) yang dihasilkan sebesar 0.584 hasil percobaan pada K=3 seperti pada Gambar 15.

Davies Bouldin

```
Davies Bouldin: 0.584
```

Gambar 15. *Davies bouldin index* k=3

Selanjutnya melakukan percobaan dengan K=4 untuk mengkonfirmasi hasil *Davies Bouldin Index (DBI)*. Hasil percobaan ini dilakukan dengan membagi data menjadi dua *Cluster* yang berbeda. Nilai (DBI) yang dihasilkan sebesar 0.727 hasil percobaan pada K=4 seperti pada Gambar 16.

Davies Bouldin

```
Davies Bouldin: 0.727
```

Gambar 16. *Davies bouldin index* k=4

Selanjutnya melakukan percobaan dengan K=5 untuk mengkonfirmasi hasil *Davies Bouldin Index (DBI)*. Hasil percobaan ini dilakukan dengan membagi data menjadi dua *Cluster* yang berbeda. Nilai *Davies Bouldin Index (DBI*) yang dihasilkan sebesar 0.690 hasil percobaan pada K=5 seperti pada Gambar 17.

Davies Bouldin

```
Davies Bouldin: 0.690
```

Gambar 17. *Davies bouldin index*s k=5

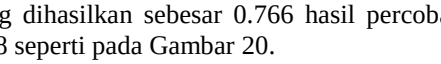
Selanjutnya melakukan percobaan dengan K=6 untuk mengkonfirmasi hasil *Davies Bouldin Index (DBI).* Hasil percobaan ini dilakukan dengan membagi data menjadi dua *Cluster* yang berbeda. Nilai (DBI) yang dihasilkan sebesar 0.698 hasil percobaan pada K=6 seperti pada Gambar 18.

Gambar 18. *Davies bouldin index* k=6

Selanjutnya melakukan percobaan dengan K=7 untuk mengkonfirmasi hasil *Davies Bouldin Index (DBI).* Hasil percobaan ini dilakukan dengan membagi data menjadi dua *Cluster* yang berbeda. Nilai *Davies Bouldin (DBI)* yang dihasilkan sebesar 0.721 hasil percobaan pada K=7 seperti pada Gambar 4.19.

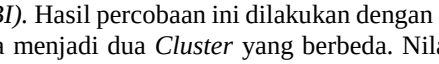
Gambar 19. *Davies bouldin index* k=7

Selanjutnya melakukan percobaan dengan K=8 untuk mengkonfirmasi hasil *Davies Bouldin Index (DBI).* Hasil percobaan ini dilakukan dengan membagi data menjadi dua *Cluster* yang berbeda. Nilai (DBI) yang dihasilkan sebesar 0.766 hasil percobaan pada K=8 seperti pada Gambar 20.



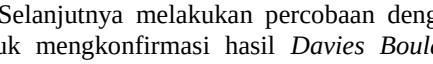
Gambar 20. *Davies bouldin index* k=8

Selanjutnya melakukan percobaan dengan K=9 untuk mengkonfirmasi hasil *Davies Bouldin Index (DBI).* Hasil percobaan ini dilakukan dengan membagi data menjadi dua *Cluster* yang berbeda. Nilai *Davies Bouldin (DBI)* yang dihasilkan sebesar 0.734 hasil percobaan pada K=9 seperti pada Gambar 21.



Gambar 21. *Davies bouldin index* k=9

Selanjutnya melakukan percobaan dengan K=10 untuk mengkonfirmasi hasil *Davies Bouldin Index* (DBI). Hasil percobaan ini dilakukan dengan membagi data menjadi dua *Cluster* yang berbeda. Nilai DBI yang dihasilkan sebesar 0.703 hasil percobaan pada K=10 seperti pada Gambar 22.



Gambar 22. *Davies bouldin index* k=10

4.6. Evaluasi

Selanjutnya adalah dilakukan evaluasi hasil dari proses *Clustering* tersebut dengan tujuan untuk mengukur kualitas hasil pengelompokkan atau klaster yang telah dilakukan dari Operator *Performance (Cluster Distance Performance).* Pada bagian parameter *Performance (Cluster Distance Performance),* kriteria utama yang digunakan adalah *Davies Bouldin Index* menggunakan persamaan (2). Selain itu, parameter *Maximize* dan *Normalize* dipilih untuk mengoptimalkan proses evaluasi. Dari hasil percobaan dapat disimpulkan bahwa nilai k optimal adalah k=2, nilai DBI = 0,527. Semakin kecil nilai DBI, semakin baik *Cluster* tersebut diperoleh. Setelah mendapatkan nilai optimal, mengimplementasikannya untuk parameter k=2 dari operator *K-Means*, hasil analisis terdapat Tabel 8.

Tabel 8. Hasil analisis nilai *davies bouldin index*

Percobaan	***Cluster* (K)**	***Davies Bouldin Indeks* (DBI)**
1	K2	0.527
2	K3	0.584
3	K4	0.727
4	K5	0.690
5	K6	0.698
6	K7	0.721
7	K8	0.766
8	K9	0.734
9	K10	0.703

Hasil dari evaluasi ini yang menampilkan nilai *Davie Bouldin Index* sebesar 0.527 Nilai ini menunjukkan bahwa hasil dari proses *Clustering* dengan K= 2 memiliki nilai *Davies Bouldin Index* yang paling optimal karena memiliki nilai dbi yang paling minimal atau mendekati 0 di antara percobaan lainnya. Penemuan nilai *Davies Bouldin Index* sebesar 0.527.

5. KESIMPULAN DAN SARAN

Dapat menentukan tujuan dan target penilaian suatu kelompok umkm dengan menerapkan data mining menggunakan algoritma *K-Means* sehingga mendapatkan Nilai *Davies bouldin Indeks*, Hasil analisis data *Cluster* UMKM dalam penelitian menunjukkan bahwa menurut prinsip DBI, nilai yang diinginkan dalam data mining adalah semakin kecil atau mendekati nol. Dari hasil analisis tersebut, terdapat jumlah *Cluster* yang memiliki nilai DBI terbaik, yaitu *Cluster* 2 dengan nilai DBI mendekati nol, sebesar 0.527 menunjukkan pengelompokan data yang cukup baik. Penggunaan parameter *Numerical Measures* dan *Euclidean Distance* dalam analisis ini membantu menghasilkan pengelompokan data yang optimal. Selain itu, hasil pengelompokan dan pengujian menggunakan *RapidMiner* menunjukkan bahwa terdapat 10 anggota *Cluster*, dengan jumlah item sebagai berikut: *Cluster* 0 (104 item), *Cluster* 1 (24 item), *Cluster* 2 (45 item), *Cluster* 3 (302 item), *Cluster* 4 (1 item), *Cluster* 5 (50 item), *Cluster* 6 (282

item), *Cluster* 7 (7 item), *Cluster* 8 (304 item), dan *Cluster* 9 (96 item).

Setelah melakukan penelitian ini, penulis memberikan beberapa saran yaitu Penelitian ini dapat membandingkan kinerja metode *K-Means* dengan algoritma *Clustering K-Medoids*, *X-Means* dan *Fuzzy C-Means* dalam mengelompokkan data UMKM yang merupakan algoritma *Clustering* yang berbeda dengan *K-Means*. Hal ini untuk mengetahui algoritma mana yang paling cocok untuk analisis data UMKM di Jawa Barat. dan Penelitian ini dapat dikembangkan dengan menggunakan dataset UMKM yang lebih besar dan beragam dari berbagai provinsi di Jawa Barat. Hal ini untuk memastikan bahwa hasil penelitian dapat membantu pemerintah dan pemangku kepentingan lainnya dalam merumuskan kebijakan dan program yang tepat sasaran untuk meningkatkan kinerja UMKM.

DAFTAR PUSTAKA

[1] B. Arifitama and A. Syahputra, "Analisis Data Mining Pada Klasterisasi UMKM Dengan Menggunakan Algoritma K-Means," *J. Ind. Kreat. dan Inform.*, vol. 2, no. 2, pp. 66–72, 2022.

[2] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," *IJCIT (Indonesian J. Comput. Inf. Technol.*, vol. 5, no. 1, pp. 75–82, 2020, doi: 10.31294/ijcit.v5i1.7951.

[3] F. Amin, D. S. Anggraeni, and Q. Aini, "Penerapan Metode K-Means dalam Penjualan Produk Souq.Com," *Appl. Inf. Syst. Manag.*, vol. 5, no. 1, pp. 7–14, 2022, doi: 10.15408/aism.v5i1.22534.

[4] N. Astuti, J. N. Utamajaya, and A. Pratama, "Penerapan Data Mining Pada Penjualan Produk Digital Konter Leppangeng Cell Menggunakan Metode K-Means Clustering," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 3, p. 754, 2022, doi: 10.30865/jurikom.v9i3.4351.

[5] Y. Dharma Putra, M. Sudarma, and I. B. A. Swamardika, "Clustering History Data Penjualan Menggunakan Algoritma K-Means," *Maj. Ilm. Teknol. Elektro*, vol. 20, no. 2, p. 195, 2021, doi: 10.24843/mite.2021.v20i02.p03.

[6] W. W. Kristianto and C. Rudianto, "Penerapan Data Mining Pada Penjualan Produk Menggunakan Metode K-Means Clustering (Studi Kasus Toko Sepatu Kakikaki)," *J. Pendidik. Teknol. Inf.*, no. 5, pp. 90–98, 2020.

[7] C. Purnama, W. Witanti, and P. Nurul Sabrina, "Klasterisasi Penjualan Pakaian untuk Meningkatkan Strategi Penjualan Barang Menggunakan K-Means," *J. Inf. Technol.*, vol. 4, no. 1, pp. 35–38, 2022, doi: 10.47292/joint.v4i1.79.

[8] A. Halim, "Pengaruh Pertumbuhan Usaha Mikro, Kecil Dan Menengah Terhadap Pertumbuhan Ekonomi Kabupaten Mamuju," *J. Ilm. Ekon. Pembang.*, vol. 1, no. 2, pp. 157–172, 2020, [Online]. Available: https://stiemmamuju.e-journal.id/GJIEP/article/view/39

[9] S. Sarfiah, H. Atmaja, and D. Verawati, "UMKM Sebagai Pilar Membangun Ekonomi Bangsa," *J. REP (Riset Ekon. Pembangunan)*, vol. 4, no. 2, pp. 1–189, 2019, doi: 10.31002/rep.v4i2.1952.

[10] N. Kholifa, N. Suarna, and W. Prihartono, "Analisis Data Transaksi Kue Mengggunakan Metode K-Means Clustering Pada Toko Rafa Cake," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3449–3457, 2024, doi: 10.36040/jati.v7i6.8221.

[11] S. Khairunnisa and M. I. Jambak, "Pengelompokan Cuaca Kota Palembang Menggunakan Algoritma K-Means Clustering Untuk Mengetahui Pola Karakteristik Cuaca," *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 2352, 2022, doi: 10.30865/mib.v6i4.4810.

[12] I. W. Septiani, A. C. Fauzan, and M. M. Huda, "Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru," *J. Sist. Komput. dan Inform.*, vol. 3, no. 4, p. 556, 2022, doi: 10.30865/json.v3i4.4055.

[13] G. Aprilianur and E. L. Hadisaputro, "Penerapan Data Mining Menggunakan Metode K-MeansClustering Untuk Analisa Penjualan Toko Myam HijabPenajam," *Jupiter*, vol. 14, no. 1, pp. 161–170, 2022.