

ANALISIS K-MEANS CLUSTERING UNTUK MENGIDENTIFIKASI KEBERADAAN BADAN USAHA MILIK DESA DI JAWA BARAT

Elin Muzilin, Martanto

Manajemen Informatika STMIK IKMI Cirebon

Jalan Perjuangan No 10 B Mejasem Kesambi Kota Cirebon, Jawa Barat 45131

elinmuzilin01@gmail.com

ABSTRAK

Badan Usaha Milik Desa Bisnis (BUMDes) merupakan entitas ekonomi yang dimiliki oleh masyarakat desa, beroperasi berdasarkan desa atau kelurahan di Jawa Barat. Penelitian ini bertujuan menerapkan metode K-Means Clustering dengan menggunakan data mining untuk menganalisis dan meningkatkan usaha BUMDes. Fokusnya adalah pada pengelolaan dan pengelompokan data masyarakat desa bisnis menggunakan metode K-Means Clustering melalui aplikasi RapidMiner. Dengan tujuan mendata dan menganalisis bisnis desa di Jawa Barat pada tahun 2021-2022, penelitian ini memproses 10.637 data bisnis desa/kelurahan. Metode K-Means Clustering digunakan untuk membagi data menjadi 9 cluster dan 2 cluster dengan nilai DBI sebesar 4.8.28. Melalui RapidMiner, dataset terdiri dari 123 rekaman yang mencakup Nama Kecamatan, Terdapat Penelitian, dan Tahun sebagai atribut. Penelitian ini mengaplikasikan K-Means Clustering untuk memahami pola dan karakteristik data Badan Usaha Milik Desa Bisnis di Jawa Barat. Metode ini membantu dalam pengelompokan data, memberikan wawasan tentang kemiripan atau perbedaan dalam karakteristik data, dan dapat digunakan sebagai dasar untuk pengelolaan lebih efektif dalam meningkatkan usaha BUMDes.

Kata Kunci: *k-means, Clustering Badan Usaha Milik Desa*

1. PENDAHULUAN

Badan usaha Milik Desa (BUMDES) merupakan lembaga ekonomi yang dimiliki oleh desa untuk meningkatkan perekonomian masyarakat desa milik usah bisnis. BUMDES memiliki peran penting dalam kegiatan yang dilakukan oleh BUMDES memberikan usaha kepada masyarakat desa untuk meningkatkan usaha mereka untuk menentukan keberadaan badan usaha milik desa bisnis berdasarkan desa/kelurahan di Jawa Barat. Penelitian ini bertujuan untuk menerapkan metode k-means clustering menentukan BUMDES keberadaan badan milik desa bisnis berdasarkan desa/kelurahan di Jawa Barat [1].

Menurut Data Open Data Jabar Dataset ini berisi tentang keberadaan Badan Usaha Milik Desa Bisnis Berdasarkan desa/kelurahan di Jawa Barat. Menentukan keberadaan desa bisnis di Jawa Barat dari tahun 2021 s.d 2022 Dataset ini merupakan terkait topik social ini yaitu yang dihasilkan oleh Dinas Badan pemberdayaan masyarakat dan desa yang dikeluarkan. Dan berikut ini tabel Data Badan Usaha.

Tabel 1. data Badan usaha provinsi Jawa Barat

Provinsi	Kabupaten	Kode	Nama Desa kelurahan	Jumlah Bumdes
Jawa Barat	Kabupaten Bogor	3201	Gunung Putri	32
Jawa Barat	Kabupaten Bogor	3201	Gunung Putri	32
Jawa Barat	Kabupaten Bogor	3201	Gunung Putri	32
Jawa Barat	Kabupaten Bogor	3201	Gunung Putri	32
Jawa Barat	Kabupaten Bogor	3201	Gunung Putri	32

Provinsi	Kabupaten	Kode	Nama Desa kelurahan	Jumlah Bumdes
Jawa Barat	Kabupaten Bogor	3201	Gunung Putri	32
Jawa Barat	Kabupaten Bogor	3201	Gunung piuting	32

Penelitian ini bertujuan menerapkan metode k-means clustering untuk melakukan pengelompokan berdasarkan data untuk menentukan keberadaan badan usaha milik desa bisnis berdasarkan desa/kelurahan di Jawa Barat. Yang diterapkan dari k-means clustering mengelompokkan data dari badan milik desa bisnis yang dapat diterapkan dengan mudah dan sederhana. Penerapan teknik pengelompokan k-means clustering membutuhkan langkah dasar yaitu memilih jumlah cluster dan centroid oleh karena itu penelitian penerapan metode k-means clustering untuk menentukan keberadaan badan milik usaha bisnis badan usaha untuk menganalisa dan mengelola data bisnis di Jawa Barat dan untuk menentukan keberadaan badan usaha milik desa bisnis berdasarkan desa/kelurahan di Jawa Barat. permasalahan pengelompokan masalah badan usaha bisnis di Jawa Barat pengelolaan data bisnis berdasarkan desa/kelurahan Jawa Barat untuk mencari data cluster [2]

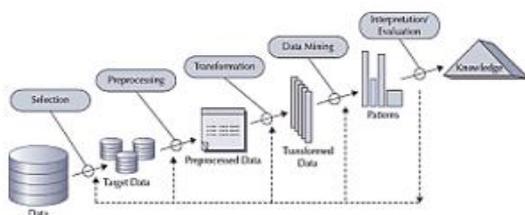
2. TINJAUAN PUSTAKA

2.1. Data Mining

Beberapa studi terkait telah dilakukan, termasuk yang mengulas topik data mining. Dalam konteks ini, tidak ada batasan spesifik terkait jumlah data yang dapat dimanfaatkan dalam pelaksanaannya. Pelaksanaan tersebut dapat melibatkan pemanfaatan

data transaksi penjualan selama satu bulan atau periode yang lebih lama, dengan tujuan mengidentifikasi keterkaitan antar barang menggunakan teknik data mining[3].

Data mining adalah suatu proses penelusuran pola atau informasi menarik yang dipilih dengan menggunakan berbagai teknik, metode, atau algoritma. Terdapat variasi yang cukup signifikan dalam penerapan teknik, metode, dan algoritma dalam praktik data mining. Pemilihan metode atau algoritma yang sesuai sangat tergantung pada tujuan yang ingin dicapai dan keseluruhan proses dalam Knowledge Discovery in Database (KDD).4].



Gambar 1. Knowledge Discovery in Database (KDD)

Berdasarkan beberapa konsep yang telah dijelaskan sebelumnya, dapat disimpulkan bahwa data mining merupakan suatu bidang ilmu yang memusatkan perhatian pada rangkaian langkah untuk mengekstrak pengetahuan atau mengidentifikasi pola dengan tujuan menghasilkan informasi yang diinginkan dari basis data atau kumpulan data yang besar. Pengetahuan yang diperoleh dari data tersebut kemudian diolah untuk membentuk informasi, yang pada akhirnya meningkatkan nilai dari data tersebut. [5]

Dalam pengertian lain, data mining dapat dianggap sebagai cabang ilmu komputer yang digunakan untuk mengekstraksi dan mengidentifikasi data dari basis data yang berskala besar. Di tengah kelimpahan data, kebutuhan akan informasi dapat dipenuhi melalui penerapan teknik-teknik seperti statistik, matematika, kecerdasan buatan, dan pembelajaran mesin. Dengan menggunakan pendekatan ini, informasi yang bermanfaat dapat diperoleh dari dataset yang besar dan kompleks. [6].

Berikut adalah perbedaan antara teknik clustering, regresi, dan klasifikasi:

a. Clustering

Clustering merupakan kategori unsupervised learning yang menganut pendekatan pengelompokan data ke dalam sejumlah kelompok yang saling mirip, baik dalam kelompok sendiri maupun berbeda dengan objek di kelompok lainnya. Tidak ada label yang telah ditentukan sebelumnya dalam proses clustering ini. Beberapa algoritma clustering yang umum digunakan mencakup K-Means, DBSCAN, dan Hierarchical clustering. Clustering menjadi metode efektif untuk mengelompokkan data dengan cara mempartisipasikan data ke dalam kelompok-kelompok

berdasarkan karakteristik tertentu yang diinginkan[7].

b. Regresi

Regresi merupakan metode analisis data yang digunakan untuk mengenali dan memahami hubungan antara dua variabel atau lebih. Tujuan utamanya adalah menyederhanakan pemahaman mengenai keterkaitan variabel-variabel tersebut. Melalui regresi, upaya dilakukan untuk meramalkan nilai suatu variabel berdasarkan nilai variabel lain yang memiliki hubungan dengannya. Oleh karena itu, regresi menjadi suatu teknik analisis yang membantu dalam menetapkan dan memahami relasi atau hubungan antara dua atau lebih variabel dalam suatu kumpulan data. [7].

c. Klasifikasi

Klasifikasi merupakan salah satu jenis supervised learning yang melibatkan proses pengelompokan data ke dalam kelompok yang telah memiliki label atau kategori sebelumnya. Dalam konteks klasifikasi, setiap data sudah diberikan label atau kategori tertentu sebagai panduan. Terdapat berbagai algoritma yang populer dalam klasifikasi, antara lain Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naïve Bayes Classifier, dan berbagai algoritma lainnya. Algoritma-algoritma ini digunakan untuk membantu mesin atau model pembelajaran untuk memahami pola dan karakteristik yang mendasari data, sehingga dapat melakukan prediksi atau klasifikasi terhadap data-data baru yang belum diberi label.

2.2. K-Means Clustering

K-Means Clustering merupakan metode pembelajaran mesin tanpa pengawasan yang bertujuan untuk mengelompokkan data ke dalam k cluster eksklusif. Cara kerja algoritma K-Means melibatkan proses pengolahan observasi pada dataset yang diberikan, membentuk kelompok berdasarkan karakteristik yang serupa. Langkah ini dilaksanakan dengan mempartisi data ke dalam kelompok berbeda, berdasarkan pola atau karakteristik yang ada dalam data tersebut. Dengan memanfaatkan algoritma K-Means, data dapat dikelompokkan ke dalam cluster yang mencerminkan struktur atau hubungan internal antar data, tanpa memerlukan label kategori sebelumnya. Algoritma ini bertujuan untuk membentuk cluster-cluster yang memiliki tingkat kesamaan yang tinggi di dalam setiap kelompok, dengan tingkat kesamaan yang lebih rendah antar kelompok. Dengan demikian, algoritma ini berusaha menciptakan pemahaman yang lebih mendalam terhadap struktur data yang kompleks, tanpa mengandalkan label kategori sebelumnya. [8].

2.3. Rapid Miner

Rapid Miner merupakan perangkat lunak yang dirancang oleh Dr. Markus Hofman dan Ralf Klittenberg dari Institut Teknologi Blanchardstown.

Perangkat lunak ini menyediakan fasilitas pemrograman dengan antarmuka pengguna grafis (Graphical User Interface/GUI). Selain digunakan untuk keperluan bisnis dan komersial, RapidMiner juga diaplikasikan dalam lingkup penelitian, pendidikan, pelatihan, pembuatan prototipe secara efisien, dan pengembangan aplikasi.[9].

2.4. Davies Bouldien Index

Indeks Davies-Bouldin (DBI) merupakan salah satu pendekatan evaluasi validitas internal yang digunakan untuk mengevaluasi kluster. DBI ditemukan oleh David L. Davies dan Donald W. Bouldin. Evaluasi validitas internal yang dijalankan oleh DBI mencakup penilaian sejauh mana kualitas kluster dapat diukur dengan menghitung sejumlah kuantitas dan turunan fitur dari dataset [10].

Proses penilaian kluster dengan memanfaatkan Davies-Bouldin Index melibatkan langkah-langkah berikut ini:

a. Sum of square within cluster (SSW)

SSW digunakan untuk menghitung jumlah jarak kuadrat antara setiap data dalam suatu kluster dan pusat kluster tersebut. Rumus SSW adalah sebagai berikut:

$$SSW_i = \sum (X_i - C_i)^2$$

Keterangan:

m adalah Jumlah data dalam cluster ke-i

x adalah data pada cluster

ci = Centroid cluster ke-i

d(X_i,C_i) adalah Jarak setiap data ke centroid i yang dihitung menggunakan jarak euclidean

b. Sum of square between cluster (SSB)

SSB digunakan untuk menghitung jumlah kuadrat jarak antara pusat kluster setiap kluster dan pusat kluster lainnya.

Rumus SSB:

$$SSB = \sum (d(c_i, c_j))^2$$

Keterangan:

d (c_i,c_j) adalah Jarak centroid dengan centroid lainnya.

c_i adalah cluster satu

c_j adalah lainnya

c. Pengukuran rasio (R_{ij})

Pengukuran rasio (R_{ij}) dipergunakan untuk membandingkan dua kluster berdasarkan kualitas kohesi dan separasinya. Perhitungan nilai rasio ini dilakukan dengan menggunakan rumus sebagai berikut:

$$R_{ij} = SSW_i / SSB_{i,j}$$

Keterangan:

R_{ij} adalah nilai rasio

SSW_i adalah Jumlah kuadrat dalam cluster terhadap centroid i

SSB_{i,j} adalah Jumlah kuadrat antara cluster data i dan j pada cluster yang berbeda

3. METODE PENELITIAN

3.1. Sumber Data

Sumber data dalam penelitian merujuk pada tempat atau lokasi dimana penelitian ini dapat memperoleh informasi atau data yang diperlukan. Data ini diambil melalui website www.opendata.jabar.com, yang bersumber dari Dinas Badam Usaha Milik Desa Bisnis Masyarakat Dan Desa di Jawa Barat, yang di perbarui dalam periode 1 tahun sekali.

Informasi ini merupakan hasil pengolahan data sekunder yang berasal dari instansi terkait di area Jawa Barat, khususnya fokus pada dataset yang mencakup informasi tingkat kelurahan/desa di Kabupaten Cirebon. Penelitian ini memanfaatkan data sekunder, yang berarti hanya melakukan analisis terhadap data yang telah ada sebelumnya. Sumber data berasal dari situs web dan publikasi Open Data Kabupaten Cirebon. Metode yang diterapkan dalam penelitian ini adalah Metode K-means clustering, suatu teknik pengelompokan data K-means yang bertujuan untuk menganalisis dan mengelompokkan data.

Dataset ini mencakup data Badan Usaha Milik Usaha di desa/ kelurahandi Jawa Barat dari tahun 2021 s.d 2022, dataset ini dihasilkan oleh Dinas Pemberdayaan Masyarakat Dan Desa dengan banyaknya 10.636 data, data yang di olah dalam penelitian ini hanya sebanyak 123 data, yaitu data di tahun 2019 - 2021. Dataset ini di buat pada tanggal 01 agustus 2022 dan terakhir di perbarui pada tanggal 09 agustus 2022.

3.2. Teknik Pengumpulan Data

Dalam penelitian ini penulis melakukan teknik pengumpulan data yang diterapkan dalam proyek penelitian ini adalah studi dokumentasi dengan memanfaatkan sumber data sekunder, yaitu informasi mengenai Daftar Data Badan Usaha Milik desa Bisnis di Jawa Barat. Data ini diperoleh dari Open Data Jabar. Menurut keterangan yang tertera di situs webnya, data ini disajikan oleh Badan Usaha Milik Desa Jawa Barat. Proses pengambilan data sekunder dilakukan secara manual dengan mengunduh informasi tersebut langsung dari situs Open Data Jabar. Berikut adalah langkah-langkah yang diambil dalam pengunduhan data manual dari Open Data Jabar

3.3. Tahapan Perancangan

Tahapan perancangan melibatkan serangkaian langkah perancangan dari tahap awal hingga tahap akhir sebagai panduan dalam penyelesaian penelitian dengan menerapkan proses Knowledge Discovery In Databases (KDD). Berikut Tahapan knowledge discovery in data base (KDD) yaitu dapat diuraikan sebagai berikut;

a. Pemilihan Data (Data Selection)

Data Selection merupakan suatu pengambilan data yang berhubungan dengan menganalisa dari basis data .Pada tahapan ini dilakukan teknik perolehan dan menganalisa data sebuah pengurangan representasi dari data permasalahan

tersebut. Dan meminimalkan informasi data dari bisnis badan usaha milik desa bisnis dimasyarakat desa di Jawa Barat pada tahun 2023 data yang digunakan yaitu 6000

b. Data Preprocessing

Data penelitian ini data Preprocessing ini yaitu digunakan untuk mengubah data mentah kedalam bentuk yang lebih mudah dipahami. Sebelum proses data mining dilakukan perlu proses cleaning pembersihan data, data badan usaha milik desa di Jawa Barat menggunakan metode k-means clustering. Pembersihan data diantaranya pembersihan yaitu menghapus dupikasi data dan memperbaiki kesalahan pada data .

c. Data Transformation

Data Transformation data penelitian ini digunakan untuk mengubah data kedalam format yang sesuai untuk diproses kedalam bentuk data mining tergantung dalam jenis pola data informasi yang akan kamu cari dalam database.

4. HASIL DAN PEMBAHASAN

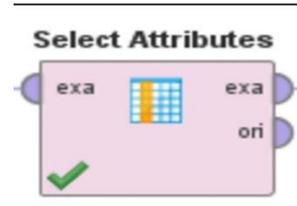
4.1. Hasil Tujuan Penelitian Ini Untuk Melakukan Penerapan K-Means Clustering Untuk Menentukan Keberadaan Badan Usaha Milik Desa Bisnis Berdasarkan Desa/Kelurahan Di Jawa Barat

4.1.1. Data Selection

Pada Penelitian ini, Nama Kecamatan, Terdapat penelitian, , Terdapat dan tahun digunakan sebagai atribut. Dataset yang digunakan dalam penelitian ini sebanyak 10.345 records. Dalam RapidMiner, langkah awal dalam proses data selection adalah melakukan data Badan Usaha Milik Desa Bisnis di Jawa Barat. berupa file dalam format excel (xlsx). Untuk memasukkan data ke dalam RapidMiner, menggunakan Import data, kemudian tentukan lokasi data yang sudah ada. Pilih cells yang ingin diimpor dan pilih form at yang sesuai pada kolom sesuai dengan jenis data.

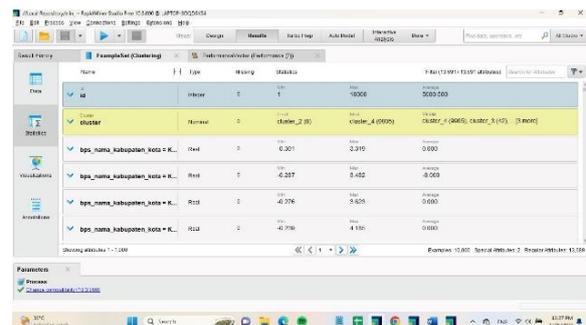
4.1.2. Preprocessing /Cleaning

Dalam penelitian ini, fase Pembersihan data, atau yang disebut juga dengan proses Cleaning, dilakukan untuk membersihkan data. Tujuan dari langkah ini adalah untuk mengatasi nilai yang mungkin hilang pada data yang akan digunakan, sehingga pada saat proses pengolahan, tidak akan muncul kesalahan atau masalah. Proses pembersihan data dilakukan menggunakan Preprocessing data dengan bantuan perangkat lunak RapidMiner versi 10.03. Setelah melalui proses preprocessing data, hasilnya menunjukkan bahwa tidak terdapat nilai yang hilang (missing value) pada data, baik pada atribut maupun label. Pada langkah selanjutnya, digunakan operator Select Attributes. Operator ini memungkinkan pengguna untuk membatasi jumlah atribut yang akan digunakan berdasarkan kriteria tertentu seperti nama, posisi, nilai rata-rata, atau variasi.



Gambar 2. operator select Atribut

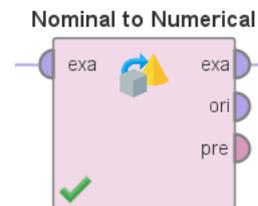
Kemudian mengubah attributes filter type menggunakan subset, dan memilih atribut pada subset dari data set yang digunakan dalam penelitian



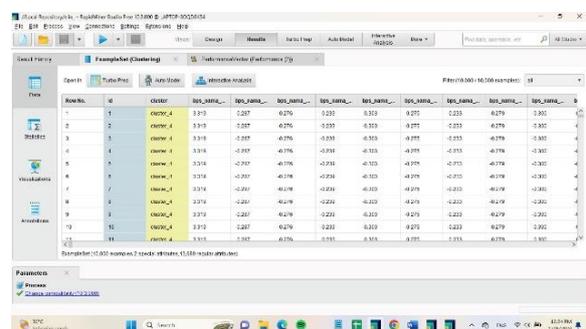
Gambar 3. Hasil Data Select Atribut

4.1.3. Data Transformations

Dalam langkah transformasi ini, peneliti mencantumkan transformasi karena bertujuan untuk menginisialisasi data dengan nilai yang sesuai dengan tipe data yang diperlukan dalam proses data mining. Dalam proses selanjutnya, dengan menggunakan operator nominal to numerical.



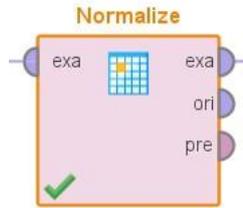
Gambar 4. Operator Nominal to Numerical



Gambar 5. Hasil Nominal to Numerical

Gambar 5 memperlihatkan urutan langkah-langkah di mana data yang awalnya dipilih disajikan dan kemudian diubah sesuai dengan algoritma yang diterapkan. Dalam penelitian ini, proses transformasi data dibagi menjadi dua tahap. Tahap pertama

melibatkan konversi data dari tipe nominal menjadi numerik, mengingat format awal data bersifat nominal. Langkah ini menjadi krusial karena K-Means hanya dapat mengoperasikan data variabel numerik. Tahap kedua dilakukan dengan menggunakan operator Normalize.



Gambar 6. Operator Normalize

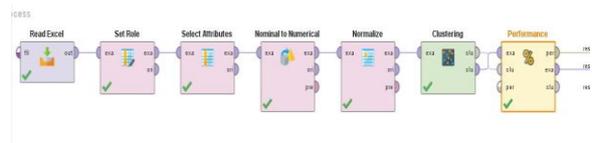
Row No.	id	spa_nama	cluster	tersepat_pu_1	tersepat_pu_2	tersepat_pu_3	tersepat_pu_4
1	1	WALID	cluster_1	0	0	0	0
2	2	CLEUDS	cluster_1	0	0	0	0
3	3	LODUN	cluster_2	1	0	0	0
4	4	PABESUAN	cluster_2	1	0	0	0
5	5	BABAKAN	cluster_2	0	0	1	0
6	6	KARAWASE	cluster_2	0	0	0	0
7	7	LESIKMBANG	cluster_2	1	0	0	0
8	8	BUSUNAN L.	cluster_2	1	0	0	0
9	9	SEKONG	cluster_2	1	0	0	0
10	10	ASTANJALAP	cluster_2	1	0	0	0
11	11	PANGSEMAN	cluster_2	1	0	0	0
12	12	MUNDU	cluster_2	1	1	1	0
13	13	BEDEP	cluster_3	0	0	0	0

Gambar 7. Hasil Normalize

Gambar 6 menggambarkan tahapan kedua dari proses transformasi data, yakni normalisasi data menggunakan operator Normalize. Pendekatan yang diterapkan dalam normalisasi ini adalah metode range transformation, yang berguna untuk menghasilkan nilai jarak yang diinginkan. Parameter yang diterapkan dalam penelitian ini melibatkan nilai minimum sebesar 0,0 dan nilai maksimum sebesar 1,0. Setelah itu, klik tombol Run untuk menampilkan hasil transformasi normalisasi.

4.1.4. Data Mining

Pada Tahap Data mining pencarian pola atau informasi menarik dari data terpilih dengan menggunakan metode k-means clustering dalam pengelompokan data tindak Badan usaha milik desa dapat digunakan untuk membentuk data set yang sudah ada. sehingga dapat membantu mengidentifikasi pola atau hubungan antar variabel dalam data tersebut. K-Means merupakan sebuah metode non-hierarki yang menggunakan pusat centroid awal sebagai representasi dari jumlah komponen populasi. metode clustering merupakan teknik pengelompokan data yang dilakukan pemisahan data kedalam kelompok berdasarkan karakteristik tertentu. Penentuan jumlah cluster merupakan langkah yang harus dilakukan. Gambar berikut ini merupakan model data mining pada algoritma K-means.



Gambar 8. Model Data Mining

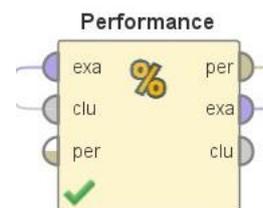
Inti dari proses Knowledge Discovery in Databases (KDD) adalah penerapan model pada data mining dengan memanfaatkan Metode K-Means clustering dalam mengelompokkan data badan usaha milik desa untuk menentukan pengelompokan data ke dalam jumlah cluster terbaik dengan nilai Davies Bouldin Index (DBI) teroptimal.

Gambar 9. Operator K-Means Clustering

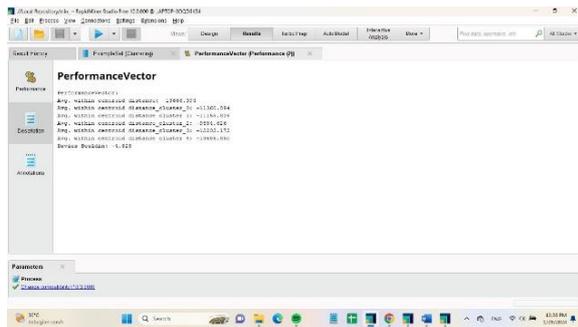
Peneliti menggunakan algoritma K-Means. dengan menggunakan operator cluster distance performance Untuk menentukan K yang paling performance menghitung DBBI.

4.1.5. Interpretation / Evaluation

Dalam penelitian ini, peneliti menggunakan operator Cluster Distance Performance. Operator K-means merupakan operatoe performance untuk mengevaluasi kerja metode clustering berbasis centroid berperan dalam mengamati nilai Davies Bouldin Index (DBI) dan rata rata jarak dalam cluster centroid berdarkan jarak terdekat yang telah dikelompokkan oleh algoritma K-Means.berikut ini gambar performance. dikelompokkan oleh algoritma K-Means.



Gambar 10. Operator Cluster Distance Performance



Gambar 11. Hasil Performance Vector

Dalam proses evaluasi digunakan operator Performance (Cluster Distance Performance). Dapat digunakan untuk membentuk data yang sudah ada dengan cara membagi atau mengelompokkan data menjadi beberapa cluster, tujuannya adalah untuk menentukan jumlah cluster yang dihasilkan dengan menggunakan nilai Davies Bouldin Index (DBI) terendah. Hasil rekapitulasi dari nilai DBI tersebut ditunjukkan dalam Tabel 2 sebagai berikut.

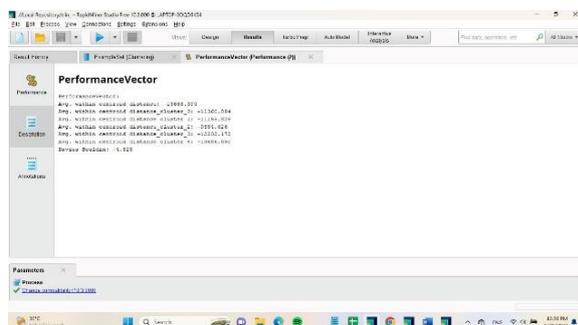
Tabel 2. Hasil percobaan menentukan Nilai K

Nilai K	DBI
2	0.289
3	0.229
...	...
9	0.122
10	0.134

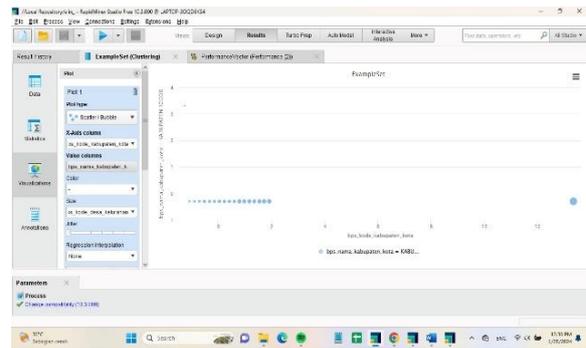
4.2. Hasil Analisis Badan Usaha Milik Desa Bisnis Menggunakan Metode K-Means Clustering Penelitian Dalam Permasalahan Ini Menggunakan K-Means Clustering

4.2.1. Hasil Visualizations

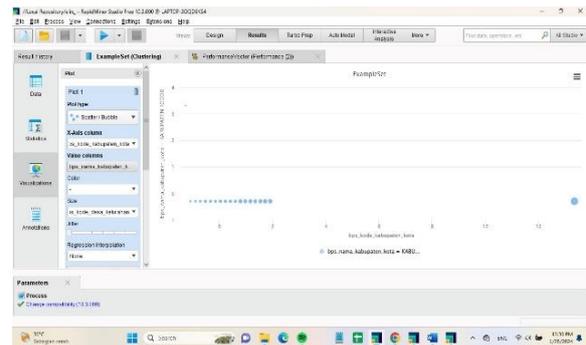
Visualisasi hasil dalam proses data mining di RapidMiner merujuk pada cara data direpresentasikan secara visual, seperti dalam bentuk grafik, plot, dan tabel. Penggunaan scatter/bubble pada penelitian ini bertujuan untuk meningkatkan aksesibilitas dan pemahaman terhadap data, seperti yang ditunjukkan pada gambar dibawah ini.



Gambar 12. Tampilan Folder View



Gambar 13. Hasil Clustering dengan 4 cluster



Gambar 14. Hasil Visualizations

Berdasarkan Gambar 14 bentuk visualisasi scatter/bubble yang merupakan titik-titik atau “bubble” yang masing-masing mempresentasikan satu atau lebih variabel dalam suatu dataset.

Tabel 3. Hasil Analisis Pengelompokan Cluster BUMDES

Cluster	2020	2021	2022	Badan usaha milik desa
Cluster 1	1	3	1	Tinggi 1
Cluster 2	2	2	24	Tinggi 2
Cluster 3	3	10	5	Tinggi 3
Cluster 4	4	4	5	Sedang 4
Cluster 5	5	5	6	Sedang 2
Cluster 6	6	4	10	Sdang 1
Cluster 7	7	3	4	Rendah 1
Cluster 8	8	8	6	Rendah 2
Cluster 9	9	3	1	Rendah 3

5. KESIMPULAN DAN SARAN

Setelah melakukan eksplorasi data menggunakan algoritma K-Means Clustering dan menganalisis informasi dari Badan Usaha Milik Desa (BUMDes), pengelompokan wilayah di Kabupaten Cirebon dilakukan dengan mengolah data tersebut melalui RapidMiner menggunakan algoritma K-Means Clustering. Kesimpulan yang dapat ditarik dari penelitian ini dapat diuraikan sebagai berikut.

Pertama, teknik pengelompokan dengan memanfaatkan algoritma K-Means Clustering berhasil membagi data menjadi 9 kelompok setelah diolah melalui RapidMiner. Badan Usaha Milik Desa Bisnis di Jawa Barat memproses dan menganalisis data ini

setelah melaksanakan eksplorasi data dengan algoritma K-Means Clustering pada 123 data BUMDes. Kedua, hasil data diproses melalui RapidMiner dan menghasilkan 9 kelompok berdasarkan atribut Nama Kecamatan, Terdapat Penelitian, dan Tahun. Dataset ini terdiri dari 123 rekaman, dan tahap awal dalam proses pemilihan data menggunakan RapidMiner melibatkan impor data dalam format Excel (xlsx).

Clustering digunakan sebagai metode untuk mempartisi data ke dalam beberapa kelas atau kelompok berdasarkan kesamaan atau karakteristik data. Data yang memiliki kesamaan dikelompokkan dalam kelompok yang sama, sedangkan data dengan karakteristik berbeda dikelompokkan dalam kelompok yang berbeda. Sebagai saran untuk penelitian mendatang, disarankan untuk terus mengeksplorasi penggunaan K-Means Clustering dengan RapidMiner untuk memprediksi, menganalisis, dan mengambil keputusan yang lebih akurat terkait data BUMDes.

Pengelompokan data ini dapat memberikan nilai yang lebih akurat bagi pemerintah desa dan masyarakat, membantu mengatasi kehilangan nilai pada data, serta memastikan ketepatan dalam proses pengolahan data.

DAFTAR PUSTAKA

- [1] J. Hutasuhut *et al.*, "MENINGKATKAN PERAN BADAN USAHA MILIK DESA DALAM MEMASARKAN PRODUK UNGGULAN LOKAL," *JABB*, vol. 4, no. 1, p. 2023, doi: 10.46306/jabb.v4i1.
- [2] M. R. Nahjan, N. Heryana, and A. Voutama, "IMPLEMENTASI RAPIDMINER DENGAN METODE CLUSTERING K-MEANS UNTUK ANALISA PENJUALAN PADA TOKO OJ CELL," 2023.
- [3] L. Kando Sihombing, U. Fatimah Sari Sitorus Pane, P. Studi Sistem Informasi, and S. Triguna Dharma, "Implementasi Data Mining Dalam Menganalisa Pola Penjualan Roti Menggunakan Algoritma Fp-Growth," *JURNAL SISTEM INFORMASI TGD*, vol. 1, pp. 228–238, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jsi>
- [4] B. Anugrah, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA FP-GROWTH PADA DATA TRANSAKSI PENJUALAN DI INDOVAPING PALEMBANG," *Bina Darma Conference on Computer Science*.
- [5] H. Jurnal, A. P. Sandi, and V. W. Ningsih, "JURNAL PUBLIKASI ILMU KOMPUTER DAN MULTIMEDIA IMPLEMENTASI DATA MINING SEBAGAI PENENTU PERSEDIAAN PRODUK DENGAN ALGORITMA FP-GROWTH PADA DATA PENJUALAN SINARMART," *JUPIKOM*, vol. 1, no. 2, 2022.
- [6] H. Agustian Siregar, N. Yanti Lumban Gaol, S. Informasi, and S. Triguna Dharma, "Penerapan Data Mining Pada Penjualan Rumah Makan Kasih Ibu Menggunakan Metode K-Means Clustering", [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jsi>
- [7] J. Hutagalung and F. Sonata, "Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 3, p. 1187, Jul. 2021, doi: 10.30865/mib.v5i3.3113.
- [8] P. Alkhairi and A. P. Windarto, *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara*. [Online]. Available: <https://seminar-id.com/semnas-sainteks2019.html>
- [9] E. Nurarofah, R. Herdiana, and N. D. Nuris, "PENERAPAN ASOSIASI MENGGUNAKAN ALGORITMA FP-GROWTH PADA POLA TRANSAKSI PENJUALAN DI TOKO ROTI," 2023.
- [10] M. Triyana, R. Juita, and C. D. Suhendra, "Penerapan Metode K-Means dalam Pengelompokan Data Penduduk Tidak Mampu di Distrik Oransbari," 2022.