

## Implementasi Algoritma K-Means Clustering Dalam Pengelompokan Data Jumlah Kerusakan Rumah Berdasarkan Kondisi Di Jawa Barat

**Fauziah Noor Musid**

STMIK IKMI Cirebon, Indonesia

**Arif Rinaldi Dikananda**

STMIK IKMI Cirebon, Indonesia

**Fathurrohman**

STMIK IKMI Cirebon, Indonesia

*Korespondensi penulis: fauziahnmusid96@gmail.com*

**Abstract.** *Based on data published by the Regional Disaster Management Agency (BPBD), West Java is the province with the highest number of natural disaster events in Indonesia with 3006 incidents for the 2015-2021 period. This condition requires BPBD (Regional Disaster Management Agency) and the Regional Government of West Java Province to pay attention to disaster management and handling the impact of disasters. Disasters that occur can have a devastating impact on various fields. One impact that is very influential is the impact of damage to the house. Damage to houses due to disasters is an impact that involves damage to the economic, social and environmental fields. Therefore, the handling of damage caused by damage to houses must be done in a mature and precise manner and the method of handling it must evolve over time. In order to be able to carry out the treatment as desired in the future, it is necessary to know the disaster clusters/groups based on the path of the number of damaged houses based on the conditions resulting from the disaster that has occurred in West Java, by implementing the K-Means clustering algorithm to classify data on the number of damaged houses based on the existing conditions. taken from the official open data portal owned by the Regional Government of West Java Province, namely Open Data Jabar. In relation to disaster impact data, grouping techniques in data mining are very useful in grouping disaster impact data in the form of damage to houses based on their similarities. This final project implements the k-means algorithm to cluster data on the amount of damage to houses due to disasters based on the conditions that occurred in West Java and produce 2 clusters/groups with the optimal dbi value of 0.118 where cluster 0 contains data from 16 regencies in West Java and cluster 1 which contains data from 9 cities in West Java.*

**Keywords:** *Disaster, Data Mining, K-means Clustering*

**Abstrak.** Berdasarkan data yang dipublikasikan oleh Badan Penanggulangan Bencana Daerah (BPBD), Jawa Barat merupakan provinsi dengan jumlah kejadian bencana alam tertinggi di Indonesia sebanyak 3006 kejadian untuk periode 2015-2021. Kondisi ini mengharuskan BPBD (Badan Penanggulangan Bencana Daerah) dan Pemda Provinsi Jawa Barat untuk memperhatikan penanggulangan bencana serta penanganan dampak bencana. Bencana-bencana yang terjadi dapat mengakibatkan dampak yang merusak berbagai bidang. Salah satu dampak yang sangat berpengaruh adalah dampak kerusakan rumah. Kerusakan rumah akibat bencana merupakan dampak yang menyangkut kerusakan pada bidang ekonomi, sosial dan lingkungan. Karena itu, penanganan dampak

kerusakan rumah harus dilakukan secara matang, tepat serta cara penanganannya harus berkembang setiap saat. Agar kedepannya bisa melakukan penanganan seperti yang diinginkan, maka perlu diketahui klaster/kelompok bencana berdasarkan dampak jumlah kerusakan rumah berdasarkan kondisi akibat bencana yang telah terjadi di Jawa Barat, dengan cara melakukan pengimplementasian algoritma K-Means clustering untuk mengklusterisasikan data jumlah kerusakan rumah berdasarkan kondisi yang diambil dari portal resmi data terbuka milik Pemda Provinsi Jawa Barat yaitu Open Data Jabar. Dalam kaitannya dengan data dampak bencana, teknik pengelompokan pada data mining sangat berguna dalam mengelompokkan data dampak bencana berupa kerusakan rumah berdasarkan kemiripannya. Proyek tugas akhir ini mengimplementasikan algoritma k-means untuk mengklusterisasi data jumlah kerusakan rumah akibat bencana berdasarkan kondisinya yang terjadi di Jawa Barat dan menghasilkan 2 klaster/kelompok dengan nilai dbi teroptimal sebesar 0,118 dimana klaster 0 berisi data yang berasal dari 16 Kabupaten di Jawa Barat dan klaster 1 yang berisi data yang berasal dari 9 Kota di Jawa Barat.

**Kata kunci:** Bencana, *Data Mining*, *K-means Clustering*

## LATAR BELAKANG

Indonesia berdasarkan letak geologisnya merupakan negara kepulauan yang terletak di tengah pertemuan empat lempeng utama bumi yaitu, Eurasia, Indo Australia, Filipina, dan Pasifik. Selain itu, terdapat pula sabuk vulkanik yang memanjang melalui pulau Jawa, Sumatera, Sulawesi, hingga Nusa Tenggara. Kondisi ini menjadikan Indonesia sebagai negara rawan bencana alam berupa gempa bumi, tsunami, dan letusan gunung berapi [1]. Suatu kejadian alam yang tidak mungkin dihindari antara lain bencana alam dengan dampaknya yang dapat berupa korban jiwa, gangguan terhadap masyarakat, kerugian harta benda, dan lingkungan sosial, sulit diantisipasi dan sulit dicegah serta kemungkinan terjadinya yang sangat tinggi. [2]. Jenis bencana berdasarkan penyebabnya dikategorikan menjadi tiga, yaitu bencana yang disebabkan oleh alam (*natural disaster*), bencana akibat teknologi (*technological-cause disaster*) dan bencana yang disebabkan oleh manusia atau *human-caused disaster* [3]. Terjadinya bencana alam sudah tentu akan mengakibatkan berbagai dampak yang merusak, baik di bidang ekonomi, sosial maupun lingkungan. Kerusakan tempat tinggal sendiri dapat digolongkan ke dalam salah satu dampak yang berpengaruh besar terhadap rusaknya kondisi ekonomi, sosial dan lingkungan masyarakat.

Open Data Jabar ialah portal data terbuka yang di dalamnya tersedia berbagai data akurat dari seluruh organisasi perangkat daerah di provinsi Jawa Barat. Seluruh data yang terdapat di Open Data Jabar merupakan data yang bersifat publik dengan satu standar

data, format, dan metadata baku. Data yang digunakan dalam proyek ini merupakan dataset yang berisi data jumlah kerusakan rumah dari tahun 2012 s.d. 2021 akibat bencana berdasarkan kondisi di Provinsi Jawa Barat. Badan Penanggulangan Bencana Daerah merupakan pembuat dataset ini yang dikeluarkan dalam periode 1 tahun sekali. Karena data ini sudah memiliki jenis, format, dan struktur data yang ditetapkan, maka data ini adalah sebuah data terstruktur [4].



**Gambar 1 Halaman Utama Website Open Data Jabar**

*Data mining* memiliki definisi yaitu serangkaian proses yang dilakukan untuk mencari pengetahuan dari suatu kumpulan data yang selama ini tidak dapat diketahui melalui cara biasa [5]. Beberapa nama atau istilah terkenal lain dari *data mining* adalah *knowledge extraction* (ekstraksi pengetahuan), KDD (*Knowledge Discovery (mining) in Databases*), analisa data/pola dan *business intelligence*[6]. KDD (*Knowledge Discovery in Databases*) ialah metode yang digunakan untuk mengidentifikasi pola dalam sebuah kumpulan data yang bertujuan untuk dapat menghasilkan sebuah pengetahuan [7]. Terdapat 5 fase/tahapan pada proses *Knowledge Discovery in Databases* (KDD [8]): (1) Seleksi Data, (2) Pemilihan Data, (3) Transformasi Data, (4) *Data Mining*, dan (5) Evaluasi. Terdapat 5 fungsi pada *data mining*: *Classification*, fungsi yang digunakan untuk menyimpulkan definisi-definisi karakteristik tertentu dari sebuah grup. *Clustering*, yaitu fungsi yang berguna untuk mengidentifikasi kelompok (klaster) dari suatu data yang memiliki karakter khusus. *Association*, fungsi yang digunakan untuk mengidentifikasi hubungan antara kejadian-kejadian yang terjadi pada waktu tertentu. *Sequencing*, fungsi ini hampir sama dengan fungsi *association* dimana *sequencing* bekerja dengan cara mengidentifikasi hubungan-hubungan yang berbeda pada periode waktu tertentu. Fungsi yang terakhir adalah *forecasting*, yaitu fungsi yang berguna untuk memperkirakan nilai pada masa depan berdasarkan pola tertentu dari sekumpulan data yang besar [9].

## KAJIAN TEORITIS

### Algoritma K-Means *Clustering*

*Clustering* ialah metode yang digunakan dalam pencarian serta pengelompokan data yang memiliki kemiripan karakteristik (*similarity*) antara data satu dengan data lainnya [10]. K-Means *Clustering* adalah bagian dari kelompok *Unsupervised Learning* karena sistem partisi digunakan untuk membagi data menjadi beberapa kelompok menggunakan prosedur ini. *Unsupervised Learning* adalah algoritma *data mining* yang digunakan untuk mencari pola yang terdapat pada atribut. Data akan dibagi menjadi beberapa kluster dengan menggunakan metode clustering berbasis jarak yang dikenal dengan algoritma K-Means. Algoritma K-Means hanya bekerja pada atribut yang didalamnya berisi nilai yang berupa angka/*numeric* [11]. Data yang ada dibagi menjadi dua atau lebih kelompok menggunakan teknik pengelompokan data non-hierarkis yang dikenal sebagai K-means *Clustering*. K-Means *Clustering* bekerja dengan mempartisi data ke dalam kelompok sehingga data yang memiliki kesamaan karakteristik akan berada pada kelompok yang sama dan data dengan perbedaan karakteristik akan dikelompokkan pada kelompok yang lain [12]. Tahapan pada algoritma K-Means *clustering* menurut [11] adalah:

1. Menentukan nilai k (jumlah kelompok/kluster) pada dataset
2. Menentukan nilai pusat (*centroid*), pada awalnya ditentukan dengan acak (*Random*), sedangkan pada tahap iterasi ditentukan dengan menggunakan rumus berikut:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X$$

Ket:

$V_{ij}$  = Centroid rata – rata pada kluster ke-i untuk variabel ke-j

$N_i$  = Jumlah anggota pada kluster ke-i

$i, k$  = *index* dari kluster

$j$  = *index* dari variabel

$X_{kj}$  = nilai dari data ke-k variabel ke-j untuk kluster tersebut.

3. Menggunakan rumus berikut untuk menentukan jarak antara setiap titik centroid massa benda dan titik-titiknya dengan menggunakan *Euclidean Distance*, yaitu jarak garis lurus yang biasa antara dua titik dalam ruang *Euclidean*.

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2}$$

Ket:

$D_e$  = Euclidean Distance (Jarak)

$i$  = Jumlah objek

$(x,y)$  = Koordinat objek

$(s, t)$  = Koordinat *centroid*

- Objek dikelompokkan berdasarkan jaraknya ke *centroid* yang terdekat
- Ulangi tahapan ke-2 hingga tahapan ke-4, iterasi dilakukan hingga nilai *centroid* bernilai optimal.

### ***Davies Bouldin Index***

*Davies Bouldin Index* (DBI) menilai evaluasi kluster dalam metode klusterisasi berdasarkan nilai kohesi dan separasi (pemisahan), DBI merupakan salah satu teknik evaluasi internal yang digunakan untuk algoritma K-Means *Clustering*. Kohesi pada *clustering* dapat dimaknakan sebagai jumlah dari kedekatan data terhadap centroid dari kluster yang diikuti, sedangkan separasi pada *clustering* merupakan pemisahan yang dilakukan pada jarak antar centroid dari klasternya [11]. Persamaan yang dipakai untuk mencari matrik kohesi dalam sebuah kluster ke- $i$  disebut *Sum of Square Within Cluster* (SSW) yang dirumuskan sebagai berikut:

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_i, c_i)$$

Ket:

$m$  = Jumlah data dalam kluster ke- $i$

$c$  = *Centroid* kluster ke- $i$

$d$  = Jarak

Sedangkan persamaan yang dipakai untuk mencari nilai separasi antar kluster disebut *Sum of Square Between Cluster* (SSB) yang dirumuskan sebagai berikut:

$$SSB_{ij} = d(c_i, c_j)$$

Keterangan:

$d$  = Jarak

$c$  = *Centroid* kluster ke- $i$

Setelah memperoleh nilai kohesi dan nilai separasi, kemudian dilakukan pengukuran rasio ( $R_{ij}$ ) yang berfungsi untuk mengetahui nilai perbandingan antara kluster ke- $i$  dan kluster ke- $j$ . Kluster terbaik merupakan kluster yang memiliki nilai kohesi terkecil dan nilai separasi terbesar. Rasio ( $R_{ij}$ ) dapat dihitung dengan menggunakan rumus berikut:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}}$$

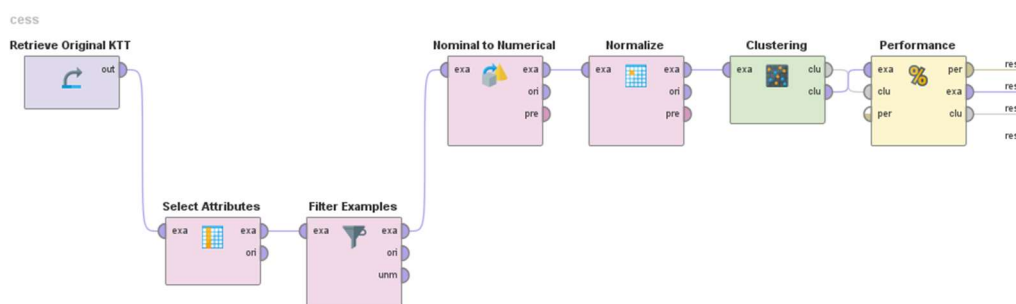
Nilai rasio yang telah didapat akan digunakan untuk menghitung nilai DBI (*Davies Bouldin Index*) melalui persamaan berikut:

$$DBI = \sum_{i=1}^k \max_{i \neq j}$$

$K$  adalah jumlah kluster yang ditentukan sesuai dengan persamaan tersebut. Kluster yang dihasilkan menggunakan *K-Means Clustering* semakin baik jika semakin kecil nilai DBI yang dicapai (bukan negatif  $\geq 0$ ). [11].

## METODE PENELITIAN

Metode perancangan yang digunakan pada proyek tugas akhir ini adalah *Knowledge Discovery in Databases* (KDD). Klusterisasi data jumlah kerusakan rumah berdasarkan kondisi akibat bencana di Jawa Barat dilakukan dengan menggunakan aplikasi Rapidminer. Proses klusterisasi menggunakan beberapa operator yang terdapat pada Rapidminer. Operator yang digunakan tersebut memiliki fungsi dan tujuan masing-masing serta berada pada tahapan yang berbeda. Tahapan pada proses klusterisasi dengan menggunakan metode *Knowledge discovery in Databases* (KDD) adalah sebagai berikut:



**Gambar 2** Proses Klusterisasi Pada Rapidminer

### Seleksi Data (*Data Selection*)

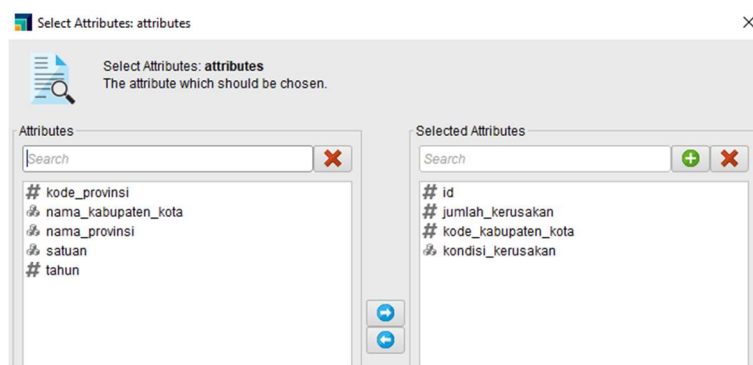
Tujuan dari proses seleksi data adalah untuk memilih atribut yang berpengaruh pada proses selanjutnya sehingga dipilih pada proses seleksi data. Proses seleksi data pada proyek tugas akhir ini dilakukan dengan menggunakan operator *Select Attributes* pada

aplikasi Rapidminer. Operator *Select Attributes* berfungsi untuk membantu proses pemilihan atribut pada dataset. Untuk dapat memilih sebagian atribut, diterapkan *attribute filter type* berupa *subset* pada parameter dari operator *Select Attributes*. Dataset jumlah kerusakan rumah berdasarkan kondisi akibat bencana yang terjadi di Jawa Barat memiliki total 9 atribut. Proses seleksi data akan menyeleksi atribut pada data sehingga nantinya hanya 4 atribut yang digunakan untuk proses selanjutnya yaitu *pre-processing data*.

**Tabel 1 Atribut Pada Dataset**

No.	Nama Atribut	Keterangan
1	id	Digunakan
2	kode_provinsi	Tidak Digunakan
3	nama_provinsi	Tidak Digunakan
4	kode_kabupaten_kota	Tidak Digunakan
5	nama_kabupaten_kota	Digunakan
6	kondisi_kerusakan	Digunakan
7	jumlah_kerusakan	Digunakan
8	satuan	Tidak Digunakan
9	tahun	Tidak Digunakan

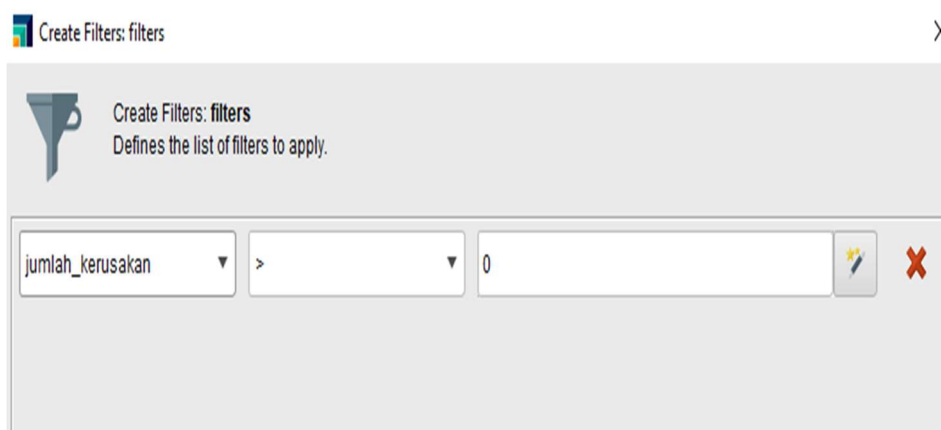
Pada Tabel 1 telah diketahui bahwa dari jumlah total 9 atribut pada dataset hanya digunakan 4 atribut yaitu atribut id, nama\_kabupaten\_kota, kondisi\_kerusakan, dan jumlah\_kerusakan. Berdasarkan informasi dari Tabel 3.1, dilakukan penerapan untuk seleksi data dengan menggunakan operator *Select Attributes* pada Rapidminer seperti yang terdapat pada Gambar 3 berikut:



**Gambar 3 Proses *Select Attributes* pada Rapidminer**

### Pemilihan Data (*Data Pre-Processing*)

Tahapan *pre-processing* dilakukan melalui Rapidminer dengan menggunakan operator *Filter Examples*. Operator *Filter Examples* digunakan untuk memilih data yang digunakan pada setiap atributnya. Pada data yang digunakan untuk proyek tugas akhir ini, atribut yang membutuhkan proses *filtering* adalah atribut jumlah\_kerusakan dengan tujuan menghapus data berupa 0 (nol) yang terdapat pada atribut jumlah\_kerusakan karena dianggap sebagai *missing value*.



Gambar 4 Proses *Filter Examples* pada Rapidminer

### Transformasi Data (*Data Transformation*)

Digunakan dua operator yang terdapat pada Rapidminer untuk tahap transformasi data, yaitu operator *Nominal to Numerical* dan juga operator *Normalize*. Operator *Nominal to Numerical* berfungsi untuk menginisialisasi atribut kondisi\_kerusakan dengan tipe data *non-numeric* menjadi tipe data *numeric*.

Tabel 2 Inisialisasi Pada Atribut Kondisi Kerusakan

No	Kondisi Kerusakan	Inisial
1	Hancur	0
2	Rusak Berat	1
3	Rusak Sedang	2
4	Rusak Ringan	3
5	Terancam	4
6	Terendam/Tertimbun	5

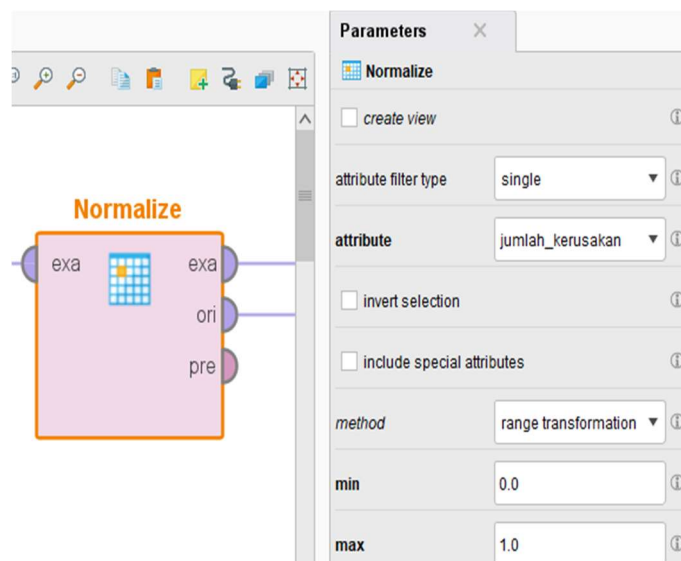


Setelah dilakukan transformasi data untuk merubah atribut *non-numeric* menjadi *numeric* dengan menggunakan operator *Nominal to Numerical*, transformasi data selanjutnya adalah menormalisasi atribut dengan data yang memiliki rentang luas.

**Tabel 3 Rentang Nilai Atribut Pada Dataset**

No	Nama Atribut	Nilai		Ket
		Terendah	Tertinggi	
1	Id	1	1071	Tidak Perlu Normalisasi
2	kode_kabupaten_kota	3201	3279	Tidak Perlu Normalisasi
3	jumlah_kerusakan	1	104536	Perlu Normalisasi
4	kondisi_kerusakan	5	0	Tidak Perlu Normalisasi

Berdasarkan Tabel 3, atribut id dan kode\_kabupaten\_kota tidak diperlukan normalisasi karena dua atribut itu merupakan atribut identitas, sedangkan untuk atribut kondisi\_kerusakan tidak dilakukan normalisasi karena rentangnya yang tidak luas. Jadi, pada dataset yang digunakan untuk proyek ini hanya terdapat satu atribut yang memiliki nilai dengan rentang luas yaitu atribut jumlah\_kerusakan sehingga diperlukan proses normalisasi atribut ini untuk mempermudah pengolahan data dengan menggunakan operator *Normalize* yang terdapat pada Rapidminer.

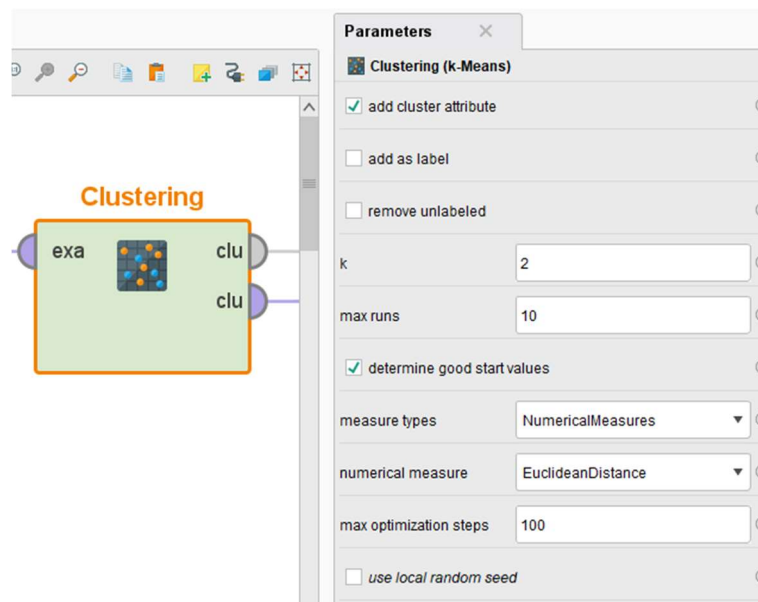


**Gambar 5 Operator *Normalize* Pada Rapidminer**

Operator *Normalize* dan parameterinya dapat dilihat pada Gambar 3.13. Pada parameter untuk operator *Normalize* diterapkan *single* untuk *attribute filter type* dikarenakan hanya ada satu atribut yang akan dinormalisasikan. Metode normalisasi yang digunakan adalah *range transformation* dengan menerapkan nilai minimal sebesar 0.0 dan nilai maksimal sebesar 1.0.

### Data Mining

Algoritma K-Means *Clustering* digunakan pada tahap *data mining* pada proyek tugas akhir ini. Algoritma K-Means *Clustering* digunakan untuk melakukan pengelompokan pada dataset jumlah kerusakan rumah ke dalam jumlah kluster terbaik dengan nilai DBI (*Davies Bouldin Index*) teroptimal.



Gambar 6 Operator *K-Means Clustering* Pada Rapidminer

Berdasarkan Gambar 5 penerapan parameter jumlah kluster sebanyak 2 ( $k = 2$ ) dengan 10 kali iterasi mengubah jumlah kluster ( $k$ ). Pada parameter *measure types* diterapkan *Numerical Measure* dan menggunakan *Euclidean Distance*.

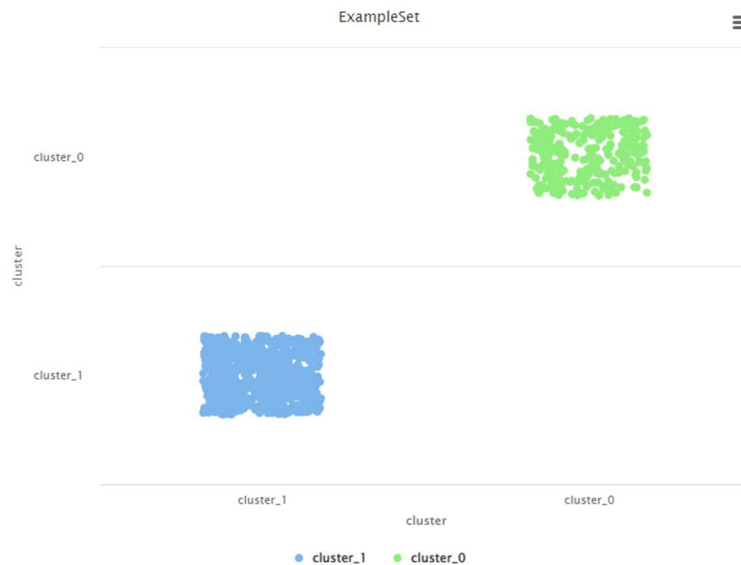
### Interpretasi/Evaluasi (*Interpretation/Evaluation*)

Proses evaluasi pada model proyek tugas akhir ini digunakan operator Performance (*Cluster Distance Performance*). Operator ini memiliki 2 jenis kriteria utama pada parameterinya, yaitu *average within centroid distance* dan juga *Davies Bouldin Index* (DBI). Pada tahap evaluasi untuk model proyek tugas akhir ini digunakan

kriteria utama Davies Bouldin *Index* (DBI) dan juga menerapkan parameter *maximize* agar dapat dihasilkan nilai DBI non-negatif ( $> = 0$ ).

### HASIL DAN PEMBAHASAN

Setelah proses seleksi data, pemilihan data dan transformasi data pada dataset, data tersebut sudah dapat diolah pada proses *data mining* dengan menggunakan algoritma *K-Means Clustering*.



Gambar 7 Visualisasi Hasil Klasterisasi

Pada parameter operator K-Means telah diterapkan jumlah klaster (*k*) sebanyak 2. Penerapan ini menghasilkan hasil klasterisasi dataset jumlah kerusakan rumah terbagi ke dalam 2 klaster sengan jumlah anggota yang berbeda.

Tabel 4 Jumlah Anggota Setiap Klaster

No	Klaster	Jumlah Anggota
1	Cluster 0	285
2	Cluster 1	786
<b>Total Data</b>		1071

Penerapan operator Performance (*Cluster Distance Performance*) untuk tahap evaluasi pada Rapidminer memiliki salah satu *output* berupa indeks Davies Bouldin. Dilakukan 10 kali percobaan untuk menemukan nilai DBI teroptimal, dimulai dari

IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING DALAM PENGELOMPOKKAN DATA  
JUMLAH KERUSAKAN RUMAH BERDASARKAN KONDISI DI JAWA BARAT

penerapan ke dalam kluster berjumlah 2 hingga penerapan ke dalam kluster berjumlah 10. Percobaan ini dilakukan untuk mencari jumlah kluster terbaik berdasarkan nilai DBInya.

Tabel 5 Nilai DBI Untuk Jumlah Kluster Berbeda

No	Jumlah Kluster	Nilai DBI
1	K = 2	0.118
2	K = 3	0.444
3	K = 4	0.562
4	K = 5	0.709
5	K = 6	0.795
6	K = 7	0.879
7	K = 8	0.900
8	K = 8	0.920
9	K = 9	0.985
10	K = 10	0.849

Berdasarkan informasi pada Tabel 5 diketahui bahwa nilai DBI teroptimal dimiliki oleh percobaan dengan jumlah kluster sebanyak 2 (dua) kluster dengan nilai DBI sebesar 0.118.

## **KESIMPULAN**

Dataset jumlah kerusakan rumah berdasarkan kondisi di Jawa Barat dapat diklasterisasi ke dalam 2 klaster dengan nilai DBI (*Davies Bouldin Index*) teroptimal yaitu sebesar 0.118 dimana anggota dari klaster 0 adalah sebanyak 285 data, dan anggota klaster 1 sebanyak 786 dengan total 1071 data. Klaster 0 beranggotakan data jumlah kerusakan rumah berdasarkan kondisi yang berasal dari 18 Kabupaten di Jawa Barat sedangkan Klaster 1 memiliki anggota data jumlah kerusakan rumah berdasarkan kondisi di Jawa Barat yang berasal dari 9 Kota yang ada di Jawa Barat. Penentuan klaster terbaik ditentukan dengan melakukan 10 kali percobaan mengubah jumlah klaster yang dapat menghasilkan nilai DBI yang berbeda sehingga dapat ditemukan jumlah klaster terbaik. Pada proyek tugas akhir ini, nilai DBI teroptimal dimiliki oleh percobaan dengan jumlah klaster 2 ( $k = 2$ ) yang mempunyai nilai DBI terkecil yaitu 0.118.

## **DAFTAR REFERENSI**

- [1] D. Ramadhan Ismana, S. Baehera, A. Fitrianto, B. Sartono, and S. Desta Oktarina, "Pengerombolan Desa di Jawa Barat Berdasarkan Daerah Rawan Bencana," *Jurnal Statistika dan Aplikasinya*, vol. 6, no. 2, 2022.
- [2] D. S. Lumbansiantar, "Analisa Data Bencana Alam Untuk Prediksi Dampak Yang Ditimbulkan Dengan Algoritma J48 (Studi Kasus : Palang Merah Indonesia)," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, pp. 25–29, Nov. 2019, doi: 10.30865/komik.v3i1.1562.
- [3] A. Heryana, "PENGERTIAN DAN JENIS-JENIS BENCANA," Jan. 2020. Accessed: Jan. 19, 2023. [Online]. Available: <https://www.researchgate.net/publication/338537206>
- [4] R. C. Prihandari, "DATAMINING: KONSEP DAN APIKASI MENGGUNAKAN RAPIDMINER (SERIES: SUPERVISED LEARNING DAN UNSUPERVISED LEARNING)," 2022.
- [5] A. H. Nasrullah, "IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI PRODUK LARIS," *Jurnal Ilmiah Ilmu Komputer*, vol. 7, no. 2, pp. 45–51, 2021, [Online]. Available: <http://ejournal.fikom-unasman.ac.id>
- [6] H. Hafizan and A. N. Putri, "Penerapan Metode Klasifikasi Decision Tree Pada Status Gizi Balita Di Kabupaten Simalungun," 2020.
- [7] R. D. Syah, "METODE DECISION TREE UNTUK KLASIFIKASI HASIL SELEKSI KOMPETENSI DASAR PADA CPNS 2019 DI ARSIP NASIONAL REPUBLIK INDONESIA," *Jurnal Ilmiah Informatika Komputer*, vol. 25, no. 2, pp. 107–114, 2020, doi: 10.35760/ik.2020.v25i2.2750.

- [8] D. Fadma Ristianti, *Komparasi Algoritma Klasifikasi pada Data Mining*, vol. 1, no. 1. 2019.
- [9] C. Zai, “IMPLEMENTASI DATA MINING SEBAGAI PENGOLAHAN DATA,” 2022. Accessed: Jan. 15, 2023. [Online]. Available: [Portaldata.org](http://Portaldata.org)
- [10] M. Iqbal, “KLAUSTERISASI DATA JAMAAH UMROH PADA AULIYA TOUR & TRAVEL MENGGUNAKAN METODE K-MEANS CLUSTERING,” *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 5, no. 2, pp. 97–104, Jun. 2019, doi: 10.33330/jurteks.v5i2.352.
- [11] Z. Nabila, A. Rahman Isnain, and Z. Abidin, “ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS,” *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [12] I. Ali, A. Rinaldi Dikananda, F. Ali Ma’ruf, and M. Abdurohman, “PENGELOMPOKAN JUMLAH PENDUDUK BERDASARKAN KATEGORI USIA 0-18 TAHUN DENGAN MENGGUNAKAN ALGORITMA K-MEANS UNTUK MENENTUKAN PENGEMBANGAN POTENSI DESA WISATA DI KABUPATEN CIREBON,” *JUMIKA*, vol. 8, no. 1, pp. 25–31, 2021.