

CLUSTERING JENIS SUMBER AIR DI PROVINSI JAWA BARAT MENGUNAKAN ALORITMA K-MEDOIDS

Ela Apriliyani¹, Bani Nurhakim², Sandy Eka Permana³, Gifthera Dwilestari⁴, Mulyawan⁵

¹²³ Program Studi Manajemen Informatika, STMIK IKMI Cirebon

⁴⁵ Program Studi Sistem Informasi, STMIK IKMI Cirebon

Jalan Perjuangan No. 10 B Majasem, Kota Cirebon

elaapriliani683@gmail.com

ABSTRAK

Air sangat penting bagi seluruh makhluk hidup di Bumi karena merupakan kebutuhan dasar yang harus dipenuhi dan dikonsumsi. Sumber air merupakan fondasi ekosistem dan keanekaragaman hayati. Terbatasnya ketersediaan air menunjukkan pentingnya melindungi dan mengelola penggunaannya secara bijaksana. Perubahan iklim dapat mempengaruhi siklus air di bumi. Dengan data sumber air yang di keluarkan dalam periode 1 tahun sekali oleh Dinas Pemberdayaan Masyarakat Tujuan penelitian ini adalah mengelompokkan jenis sumber air dengan metode *K-Medoids* dan menganalisis hasilnya dengan implementasinya menggunakan *RapidMiner*. Penelitian ini menghasilkan sebanyak 3 *cluster*, pada masing-masing *cluster* memiliki anggota yang berbeda. *Cluster* 0 memiliki 200 item, *cluster* 1 memiliki 96 items, dan *cluster* 2 memiliki 160 items yang, dengan total dari ketiga *cluster* tersebut sebanyak 456 items

Kata kunci : sumber air, *K-Medoids*, *RapidMiner*, *Clustering*

1. PENDAHULUAN

Air memiliki peranan yang sangat penting bagi setiap makhluk hidup yang ada di muka bumi ini. Bisa dikatakan bahwa, jika tidak ada air di suatu tempat di muka bumi ini maka tidak akan ada harapan untuk suatu kehidupan. Dikarenakan air merupakan kebutuhan pokok yang harus dipenuhi dan memiliki sifat wajib untuk dikonsumsi, sumber air menjadi pondasi bagi ekosistem dan keragaman hayati ketersediaan air yang terbatas menunjukkan pentingnya konservasi dan manajemen yang bijaksana dalam penggunaannya. Perubahan iklim dapat mempengaruhi siklus air di Bumi. Hal ini dapat menyebabkan pola curah hujan yang berubah, dan mengakibatkan perubahan ketersediaan air, dan dampak terhadap ekosistem, pertanian, serta kehidupan manusia. Penggunaan metode *K-Medoids* untuk *clustering* sumber air di Provinsi Jawa Barat berdasarkan kabupaten memiliki manfaat praktis yang signifikan dalam mendukung pengelolaan sumber daya air yang lebih efektif dan berkelanjutan serta meningkatkan kualitas hidup masyarakat.

Terdapat macam macam sumber air, diantaranya Sumber air permukaan yang terdapat di permukaan Bumi, seperti sungai, danau, waduk, dan embung. Lalu Sumber Air Tanah, Air tanah dapat diakses melalui sumur atau mata air dan sering menjadi sumber air yang penting bagi penggunaan domestik, irigasi, dan industri. kemudian air hujan, air laut Namun air laut tidak dapat dikonsumsi secara langsung karena kandungan salinitasnya yang cukup tinggi serta tidak memenuhi baku mutu air minum. Desalinasi merupakan cara yang banyak digunakan untuk memanfaatkan air laut menjadi air layak konsumsi. [1] dan lain sebagainya. Perlindungan dan pelestarian sumberdaya air harus menjadi salah satu

prioritas utama manusia. Pemanfaatan air untuk berbagai kebutuhan harus memperhatikan parameter-parameter kualitas air sesuai baku mutu yang sudah ditetapkan. [2]

2. TINJAUAN PUSTAKA

2.1. Penelitian terkait

Sumber daya air adalah sumber daya air yang dapat digunakan atau dimanfaatkan manusia. Air digunakan dalam bidang pertanian, industri, rumah tangga, rekreasi, dan aktivitas lingkungan. Sebagian besar air tawar yang tidak membeku ditemukan di dalam tanah dalam bentuk air tanah, dan hanya sebagian kecil berada di udara dan di atas permukaan tanah. Sebagian besar air asin di Bumi berjumlah 97%, dan sisanya 3% merupakan air tawar. [3]

2.2. Data Mining

Data mining merupakan suatu proses yang dapat digunakan untuk menemukan hubungan yang berarti dari sekumpulan data dengan memeriksa data yang tersimpan dalam media penyimpanan dengan menggunakan teknik clustering, [4]. Pembuatan model dilakukan melalui proses yang menggunakan algoritma atau rumus tertentu. Data mining menggunakan teknik seperti statistik, matematika, dan pembelajaran mesin, yang digunakan untuk mengidentifikasi dan mengolah berbagai data menjadi informasi berguna. (Arhami & Nasir & al., 2020).

2.3. Clustering

Clustering adalah proses membagi sekumpulan objek data ke dalam himpunan bagian yang disebut cluster. Cluster memiliki objek yang berbeda dan memiliki karakteristik yang sama. Melalui penggunaan algoritma clustering, partisipasi

dilakukan secara otomatis. Akibatnya, clustering sangat bermanfaat karena dapat menemukan kelompok atau kelompok yang tidak dikenal dalam data. [5].

Clustering adalah proses pembagian data ke dalam kelas atau cluster berdasarkan tingkat kesamaannya. Dalam clustering, data yang memiliki kesamaan dimasukkan ke dalam cluster yang sama, sedangkan data yang tidak memiliki kesamaan dimasukkan dalam cluster yang berbeda [6].

Kesamaan dalam kelas tinggi dan rendah ditunjukkan oleh hasil pengelompokan yang baik. Dalam masalah, kesamaan adalah ukuran numerik dari dua objek. Jika ada kemiripan yang lebih besar antara dua objek yang dibandingkan, maka skor kemiripan mereka tinggi. serta sebaliknya. Kualitas hasil clustering dipengaruhi oleh teknik yang digunakan. Untuk pengelompokan, ada empat jenis data. Keempat kategori data ini adalah

1. Variabel berukuran interval
2. Variabel biner
3. Variabel nominal, ordinal, dan rasio
4. Variabel dengan tipe lainnya.

Kemampuan metode clustering untuk menemukan pola tersembunyi dalam data yang diselidiki harus diukur. Nilai kemiripan antara objek yang dibandingkan dapat dihitung dengan berbagai cara. Salah satunya adalah jarak Euclidean tertimbang. Jarak Euclidean menghitung jarak antara dua titik dengan mengetahui nilai setiap atribut dari dua titik tersebut. Untuk menghitung jarak menggunakan jarak Euclidean, gunakan rumus berikut:

$$Distance(p, q) = \left(\sum_k^n \mu_k |pk - qk|^r \right)^{1/r} \tag{1}$$

Keterangan:

N = Jumlah data record

K= Urutan field data

r= 2

μ_k = Bidang Bobot yang diberikan pengguna

Metode umum untuk menemukan persamaan atau ketidaksamaan dua vektor ciri, yang ditunjukkan dengan peringkat, adalah jarak. Semakin kecil nilai rangking yang dihasilkan, semakin dekat kedua vektor dan lebih mirip. Metode pengukuran jarak Euclidean adalah salah satu yang paling umum digunakan. Pengukuran jarak dapat dijelaskan dengan persamaan berikut:

$$j(v_1, v_2) = \sqrt{\sum_{k=1}^N (v_1(k) - v_2(k))^2} \tag{2}$$

dimana v_1 dan v_2 adalah dua vektor yang jaraknya akan dihitung dan N menyatakan panjang vektor.

2.4. Flowchart K-Medoids



Gambar 1. Flowchart

Sistem Clustering, yang digambarkan dalam flowchart tersebut, adalah metode penulisan algoritma menggunakan notasi garis. Flowchart adalah gambar atau bagan yang menunjukkan urutan atau langkah-langkah suatu program, serta hubungan antar proses dan pernyataannya. Setiap simbol menggambarkan proses tertentu karena simbol-simbol yang digunakan untuk menyampaikan gambaran ini. sementara garis penghubung menunjukkan jarak antara proses. Proses flowchart akan memudahkan pengecekan elemen analisis masalah yang terlupakan. Sebelum proses pembersihan dan transformasi data, data pada flowchart K-medoid ini masih tidak relevan atau valid. Data tidak ada. Namun, setelah mengakses sistem melalui kuesioner online. Data kosong tersebut harus diisi sepenuhnya atau tidak kosong sama sekali,

sehingga data tersebut lengkap dan siap untuk proses klasterisasi. [8]

2.5. K-Medoids

Algoritma *K-Medoids* atau sering disebut juga dengan algoritma PAM (*Partitioning Around Medoid*) dikembangkan oleh *Leonard Kaufman dan Peter J. Rousseeuw*, yang merupakan algoritma partitional yang memecah dataset menjadi kelompok-kelompok. Algoritma *K-Medoids* menggunakan nilai rata-rata (means) dari setiap cluster untuk menentukan pusat kelompok, sedangkan algoritma *K-Medoids* menggunakan nilai rata-rata dari setiap cluster [7]

Algoritma ini bekerja dengan meminimalkan jumlah kesamaan yang ada antara setiap objek dan titik referensi yang tepat. Algoritma *K-Medoids* dapat dilakukan dalam beberapa langkah [7]

- Inisialisasi pusat cluster sebanyak k (jumlah cluster).
- Untuk menghitung jarak Euclidian, hitung setiap objek ke cluster terdekat dengan menggunakan persamaan (2.1).
- Sebagai kandidat non-medoids, mulai pusat cluster baru pada setiap objek secara acak setelah menghitung jarak Euclidian.
- Hitung jarak setiap objek pada masing-masing cluster dengan kandidat non-medoids.
- Hitung total simpangan (S) dengan menghitung jarak total baru dibagi dengan jarak total lama. Jika S kurang dari 0, maka data cluster non-medoids akan digunakan untuk membentuk sekumpulan k objek baru sebagai medoids.
- Untuk mendapatkan cluster beserta anggota cluster masing-masing, ulangi langkah c-e hingga tidak ada perubahan pada medoid.

2.6. RapidMiner

Lebih dari 500 karyawan RapidMiner menangani input, output, pra-pemrosesan data, dan visualisasi personal. RapidMiner adalah mesin penambangan data yang dapat diintegrasikan ke dalam produknya sendiri dan juga tersedia sebagai perangkat lunak analisis data mandiri. Dibangun di Java, ia bekerja dengan semua sistem operasi, [9]

3. METODE PENELITIAN

3.1. Sumber Data

Sumber data dalam penelitian ini diperoleh dari website, <https://opendata.jabarprov.go.id/id/dataset/jumlah-des-a-berdasarkan-sumber-air-untuk-mandi-dan-cuci-di-jawa-barat>. Dataset terkait topik Pemerintah dan Desa ini dihasilkan oleh Dinas Pemberdayaan Masyarakat yang dikeluarkan dalam periode 1 tahun sekali. dataset ini berjumlah 457 data. Dataset ini diambil pada 10 Desember 2023. Jenis dataset ini adalah data sekunder Data sekunder merujuk kepada informasi atau data yang dikumpulkan, diolah, atau disusun oleh pihak lain

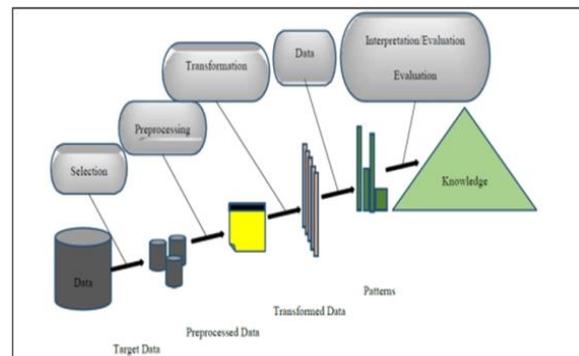
atau orang lain sebelumnya. Data sekunder dapat berasal dari berbagai sumber, seperti publikasi ilmiah, studi penelitian, laporan pemerintah, database online, dan sumber informasi lainnya.

3.2. Teknik Pengumpulan

Penelitian Dokumen Penelitian dokumen merupakan suatu metode pengumpulan data yang diperoleh dari dokumen-dokumen atau catatan-catatan yang ada. Dalam penelitian ini data diperoleh dari situs resmi Badan Pusat Statistik. Data yang diperoleh akan diperiksa dan diolah oleh peneliti dan digunakan untuk memperoleh informasi untuk penelitian ini.

3.3. Tahap Perancangan

Tahap perancangan penelitian ini dilaksanakan secara bertahap *Knowledge Discovery in Databases* (KDD) merupakan sekumpulan proses untuk menemukan pengetahuan yang bermanfaat dari data. KDD terdiri dari serangkaian langkah perubahan, termasuk data preprocessing dan juga post processing [10] berdasarkan proses penemuan pengetahuan dalam database dan disusun untuk dijalankan secara berurutan hingga tahap akhir. Tahapan perancangan tugas akhir ini dapat dilihat pada Gambar 2. Berikut :



Gambar 2. *Knowledge Discovery Database* (KDD)

Berikut penjelasan lebih lanjut mengenai proses *Knowledge Discovery Database* (KDD) berdasarkan gambar 2. diatas:

- Data selection**
Pada proses ini dilakukan pemilihan data dan pengecekan data pada dataset yang akan digunakan.
- Pemilihan Data (*Pre-processing*)**
Pengayaan, atau "peningkatan" data yang ada dengan data atau informasi eksternal, dan penghapusan data duplikat atau tidak konsisten adalah semua contoh proses pre-processing.
- Data Transformation***
Transformasi data, data kolom B (jenis sumber air) harus dilakukan proses inisialisasi data terlebih dahulu ke dalam bentuk angka/numerikal. dengan metode *K-medoids*.
- Data Transformati***

Pada tahap transformasi data, data diubah ke dalam format yang dapat digunakan oleh alat sehingga dapat digunakan dalam proses data mining. Kumpulan data tersebut diolah dan digabungkan menjadi satu sehingga proses data mining dapat dilakukan.

5. *Data Mining*

Penambangan data melibatkan pencarian informasi dan pola menarik dalam potongan data yang dipilih dengan menggunakan teknik dan metode tertentu. Pemilihan metode atau algoritma sangat penting karena sangat mempengaruhi tujuan dan proses KDD secara keseluruhan.

6. *Evaluation/Knowledge Interpretation*

Tahap ini penting untuk interaksi KDD yang mencakup pemeriksaan apakah contoh atau data yang ditemukan bertentangan dengan kenyataan saat ini atau spekulasi sebelumnya, [11] Pada tahap ini hasil *data mining* akan dievaluasi dan diinterpretasikan sehingga dapat dijadikan sebagai informasi. Hasil *data mining* akan dievaluasi menggunakan *Davies-Bouldin Index*. *Davies-Bouldin Index* akan digunakan untuk menilai hasil data mining. *DBI* mengumpulkan algoritma clustering pada kumpulan data berdasarkan tingkat densitas centroid data, rasio dan keterpisahan antara cluster, dan hasil perhitungan ketiga nilai. Jika keadaan maksimal terjadi pada jarak cluster, karakter cluster akan sangat kecil sehingga perbedaan menjadi sangat jelas [12]. Kemudian akan dilakukan penarikan kesimpulan berdasarkan hasil analisis dan evaluasi.

4. HASIL DAN PEMBAHASAN

4.1. Mengelompokan Sumber Air Menggunakan Algoritma K-Medoids

4.1.1. Pengumpulan data (*Daya Collecting*)

Data dalam penelitian ini berjenis sekunder, Berikut merupakan tabel dari data sumber air yang diambil dari website open data tahun 2021-2023 :

Tabel 1. Data sumber air

id	Nama kabupaten	Jumlah desa	satuan	tahun
1	KABUPATEN BOGOR	99	DESA	2020
2	KABUPATEN BOGOR	180	DESA	2020
3	KABUPATEN BOGOR	206	DESA	2020
4	KABUPATEN BOGOR	30	DESA	2020
5	KABUPATEN BOGOR	338	DESA	2020
6	KABUPATEN BOGOR	408	DESA	2020
7	KABUPATEN BOGOR	357	DESA	2020
...
454	KOTA BANJAR	16	DESA	2023
455	KOTA BANJAR	16	DESA	2023
456	KOTA BANJAR	3	DESA	2023

4.1.2. Data sebelum Pre-processing

Data yang telah diproses sebelumnya mengacu pada data mentah atau asli sebelum diubah atau dimodifikasi untuk memenuhi kebutuhan analisis atau

pemrosesan data tertentu. Pemrosesan awal data merupakan langkah penting dalam analisis data yang bertujuan untuk mengatur, mengatur, dan menyiapkan data untuk digunakan secara efektif dalam model dan analisis lebih lanjut.

4.1.3. Data Reduction

Data Reduction merupakan tahapan untuk mengurangi data yang tidak dibutuhkan ketika akan di proses ke sebuah pemodelan. Atribut dari data mentah yaitu id, nama provinsi, kode provinsi, nama kabupaten, sumber air, jumlah desa, tahun. Pada data sumber air ini ada atribut yang tidak dibutuhkan yaitu atribut nama provinsi, kode provinsi, nama kabupaten, sehingga data hasil integrasi ada 4 yaitu id, sumber air, jumlah desa, dan tahun.

4.1.4. Data Cleaning

Data Cleaning adalah proses mengisi nilai yang hilang atau salah, mengoreksi data yang tidak konsisten, dan mengidentifikasi redundansi data. Akumulasi Data atau Duplikasi Data. Penelitian ini melakukan proses pembersihan data dengan cara memfilter setiap kolom dan mencari data yang kosong atau hilang dengan menggunakan aplikasi Microsoft Excel.

4.1.5. Data Transformation

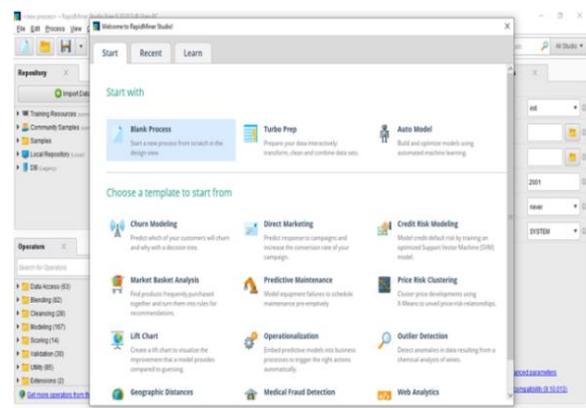
Data Transformation ialah proses melakukan transformasi data yang belum memiliki data entitas jelas untuk di ubah menjadi data yang valid atau siap menjalani proses Data Mining.

4.1.6. Data Integration

Data integration adalah suatu proses untuk menggabungkan data dari beberapa data.

4.2. Bagaimana hasil dari pengelompokan sumber air menggunakan algoritma K-Medoid

- Langkah pertama yaitu membuat dataset dengan format *.xlsx* pada *Ms.Excel*
- Buka aplikasi Rapidminer, kemudian klik *New Process* dan klik *Blank*



Gambar 3. Tampilan awal rapidminer

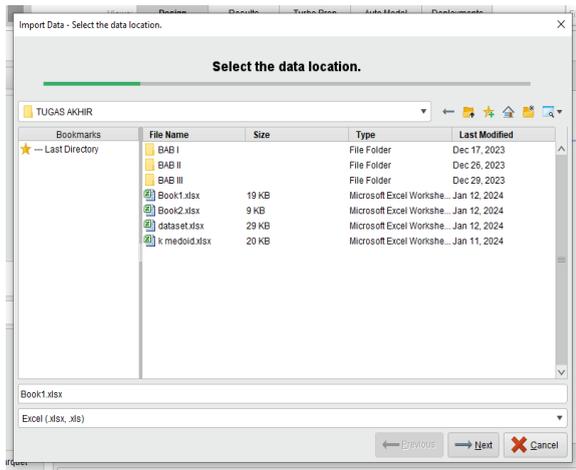
3. Operator read Excel

Untuk menambahkan data yang akan diolah, maka perlu ditambahkan *read excel* yang ada pada menu *operators*, kemudian klik menu *import configuration wizard*.



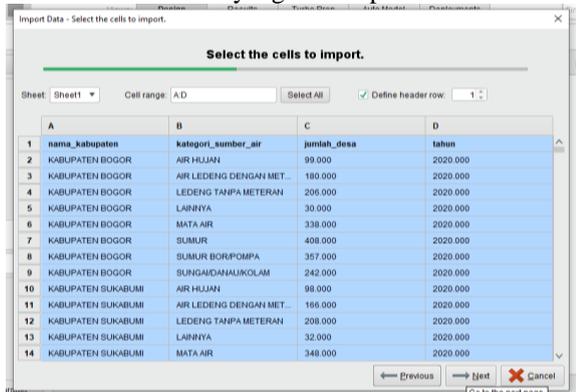
Gambar 4. Operator read Excel

4. Pilih data yang ingin ditambahkan, lalu klik Next.



Gambar 5. Mencari data yang akan digunakan

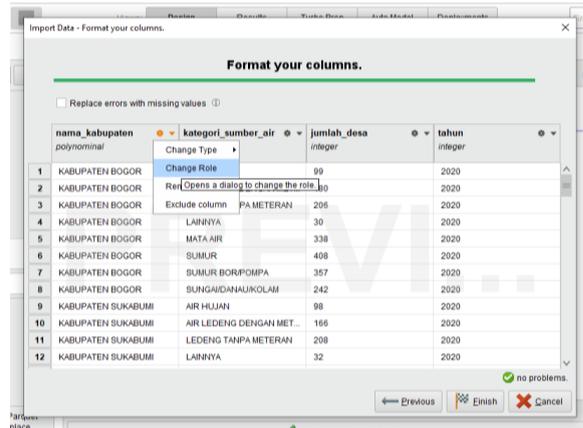
5. Pilihlah atribut yang akan diproses.



Gambar 6. Memilih atribut yang akan di proses

Pada Gambar 6. pilihlah atribut yang akan diproses pada algoritma *k-medoids clustering*, atribut yang akan di proses meliputi Nama Kabupaten, kategori jenis sumbe air, jumlah pengguna dan tahun. Setelah atributnya dipilih kemudian klik tombol *Next*.

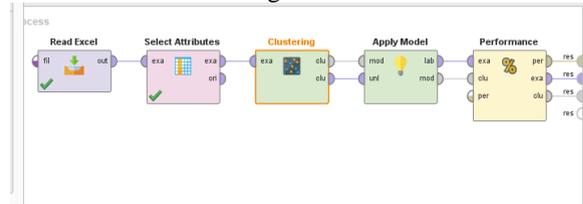
6. Merubah format kolom



Gambar 7. Merubah format kolom

Pada gambar 7. tampilkan menu *change role* kemudian rubah kolom atribut Nama Kabupaten menjadi Id, lalu klik tombol *Finish*.

7. Model mata mining



Gambar 8. Model data mining

Operator *apply model* digunakan untuk membaca kumpulan data dalam *cluster* berdasarkan data yang diselidiki. Operator *Cluster Performance Distance* digunakan untu mengukur kinerja algoritma *k-medoids clustering* pada dataset yang telah dibaca.

4.3. Hasil Clustering

4.3.1. Cluster Model

Cluster Model

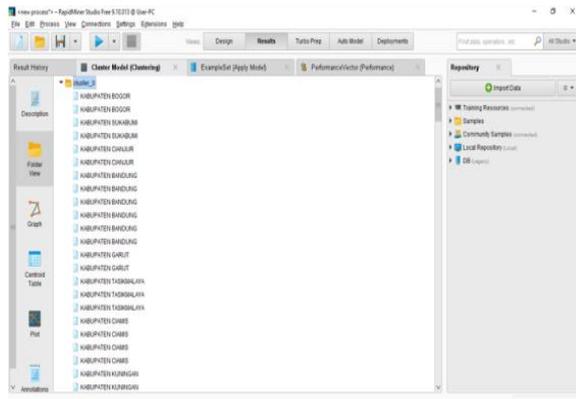
Cluster 0: 200 items
 Cluster 1: 96 items
 Cluster 2: 160 items
 Total number of items: 456

Gambar 9. Cluster model

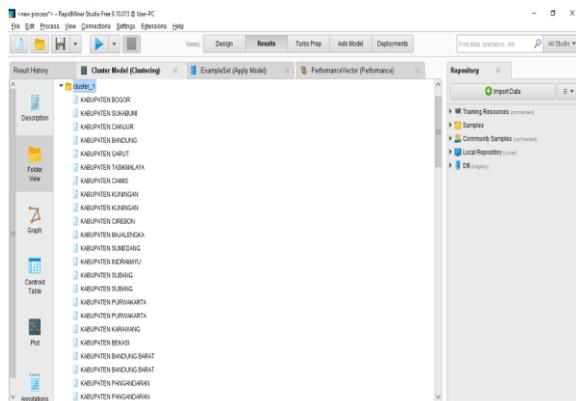
Penelitian ini menghasilkan sebanyak 3 *cluster*, pada masing-masing *cluster* memiliki anggota yang berbeda. *Cluster 0* memiliki 200 item, *cluster 1* memiliki 96 items, dan *cluster 2* memiliki 160 items yang, dengan total dari ketiga *cluster* tersebut sebanyak 456 items.

4.3.2. Folder View RapidMiner

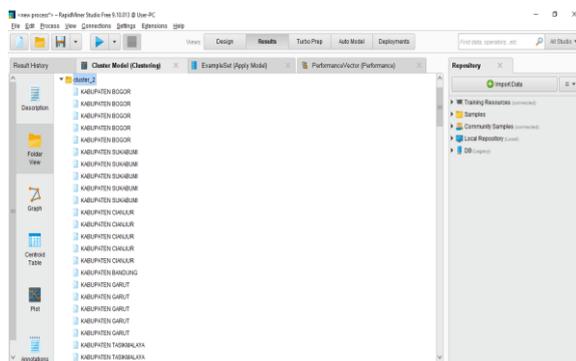
Pada gambar berikut terdapat nama nama kabupaten yang termasuk pada *cluster 0*, *cluster 1*, dan *cluster 2*.



Gambar 10. View RapidMiner Cluster 0



Gambar 11. View RapidMiner Cluster 1



Gambar 11. View RapidMiner Cluster 0

4.3.3. Performance Vector

PerformanceVector

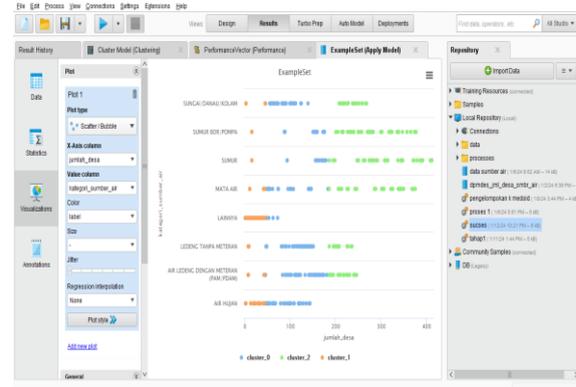
```
PerformanceVector:
Avg. within centroid distance: 1178.413
Avg. within centroid distance_cluster_0: 995.418
Avg. within centroid distance_cluster_1: 231.448
Avg. within centroid distance_cluster_2: 1975.338
Davies Bouldin: 0.295
```

Gambar 12. Performance Vector

Pada gambar 12. menunjukkan hasil performance vector, Vektor kinerja mengacu pada sekumpulan metrik atau parameter yang digunakan

untuk mengukur hasil suatu sistem, proses, atau unit.

4.3.4. Visualizations RapidMiner



Gambar 13. Visualizations RapidMiner

Pada gambar 13. Menunjukkan hasil dari cluster 0, cluster 1, dan cluster 2 dengan masing-masing penggunaan jenis sumber air yang berbeda

5. KESIMPULAN DAN SARAN

Berdasarkan data yang telah diolah, maka penelitian ini menyimpulkan bahwa dapat diketahui bagaimana Mengelompokan sumber air menggunakan algoritma *k-medoids*, yaitu menggunakan tahapan KDD yang dijalankan menggunakan RapidMiner. Dan dapat diketahui Bagaimana hasil dari pengelompokan sumber air menggunakan algoritma *k-medoids*, yaitu dengan pengelompokan jenis sumber air berdasarkan kabupaten dan desa pada tahun 2020, 2022, dan 2023 menggunakan *k-medoids clustering* dan implementasinya pada Rapidminer, menghasilkan 3 cluster, diantaranya cluster 0 memiliki tingkat pengguna jenis sumber air paling tinggi hampir semua jenis sumber air dipakai, cluster 1 memiliki tingkat pengguna jenis sumber air paling rendah dan paling banyak menggunakan air hujan, cluster 2 memiliki tingkat pengguna jenis sumber air sedang dengan penggunaan jenis sumber air sungai, danau, sumur, air ledeng (pam/pdam), mata air. Penelitian ini masih memiliki banyak kekurangan adapun saran yang dapat dikemukakan yaitu: untuk peneliti selanjutnya dalam mengelompokan jenis sumber air sebaiknya menggunakan metode lain selain metode algoritma *K-Medoids Clustering*.

DAFTAR PUSTAKA

- [1] A. M. C. Kutananda and H. S. Titah, "Kajian Desalinasi Air Laut Menggunakan Sistem Reverse Osmosis sebagai Pemenuhan," vol. 11, no. 3, 2022.
- [2] I. S. Sulistyorini, M. Edwin, and A. S. Arung, "Analisis Kualitas Air Pada Sumber Mata Air Di Kecamatan Karanggen dan Kaliorang Kabupaten Kutai Timur," *J. Hutan Trop.*, vol. 4, no. 1, p. 64, 2017, doi: 10.20527/jht.v4i1.2883.

- [3] F. Lestari, T. Susanto, and K. Kastamto, "Pemanenan Air Hujan Sebagai Penyediaan Air Bersih Pada Era New Normal Di Kelurahan Susunan Baru," *SELAPARANG J. Pengabd. Masy. Berkemajuan*, vol. 4, no. 2, p. 427, 2021, doi: 10.31764/jpmb.v4i2.4447.
- [4] R. Noviani, Y. H. Chrisnanto, and H. Ashaury, "Sistem Segmentasi Keluhan Air Bersih di PT. Suryacipta Swadaya Menggunakan K-Medoids Clustering," *Pros. SISFOTEK*, pp. 162–166, 2020, [Online]. Available: <http://www.seminar.iaii.or.id/index.php/SISFOTEK/article/view/206>
- [5] K. Annisa, B. S. Ginting, and M. A. Syar, "Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode Clustering Pada Pdam Langkat," *J. Sist. Inf. Kaputama*, vol. 6, no. 2, pp. 165–179, 2022, doi: 10.59697/jsik.v6i2.167.
- [6] D. N. Batubara, A. P. Windarto, D. Hartama, and H. Satria, "Analisis Metode K-MEANS Pada Pengelompokan Keberadaan Area Resapan Air Menurut Provinsi," *Semin. Nas. Sains Teknol. Inf.*, no. x, pp. 345–349, 2019, [Online]. Available: <https://prosiding.seminar-id.com/index.php/sensasi/article/view/324> (05 Juni 2020)
- [7] B. Wira, A. E. Budiarto, and A. S. Wiguna, "Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang," *RAINSTEK J. Terap. Sains Teknol.*, vol. 1, no. 3, pp. 53–68, 2019, doi: 10.21067/jtst.v1i3.3046.
- [8] R. Alawiyah, A. Mahmudi, and S. Achmadi, "Penerapan Metode K-Medoid Pada Analisis Respon Emosi Marah Wanita Pendalangan Berbasis Web," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 1, pp. 126–133, 2022, doi: 10.36040/jati.v6i1.4508.
- [9] D. Pascalina, R. Widhiastono, and C. Juliane, "Pengukuran Kesiapan Transformasi Digital Smart City Menggunakan Aplikasi Rapid Miner," *Technomedia J.*, vol. 7, no. 3, pp. 293–302, 2022, doi: 10.33050/tmj.v7i3.1914.
- [10] R. Bahtiar, "Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor," *J. Inform. MULTI*, vol. 1, no. 3, pp. 203–214, 2023, [Online]. Available: <https://jurnal.publikasitecno.id/index.php/jim203>
- [11] C. Menggunakan, A. Apriori, Z. D. Nugraha, H. Sibyan, and N. Mardiyantoro, "Journal of Engineering and Informatic Implementasi Data Mining untuk Menentukan Pola Penjualan di Armada," vol. 1, no. 2, pp. 60–64, 2023, doi: 10.56854/jei.v1i2.62.
- [12] S. F. Intan, W. Elvira, S. Rahayu, and ..., "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Pengeluaran Mahasiswa: Comparison of the K-Means and K-Medoids Algorithms for ...," ... *Nas. Penelit. dan ...*, pp. 35–40, 2023, [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas/article/view/543%0Ahttps://journal.irpi.or.id/index.php/sentimas/article/download/543/335>