

PENERAPAN ALGORITMA K-MEANS BERBASIS RAPIDMINER PADA ANALISIS DISTRIBUSI AIR BERSIH DI INDONESIA

Melisa Agustina✉, Nor Anisa

Fakultas Sains & Teknologi, Universitas Sari Mulia, Banjarmasin, Indonesia

Email: melisa.agustina@student.unism.ac.id

DOI: <https://doi.org/10.46880/methoda.Vol15No1.pp43-49>

ABSTRACT

This study aims to analyze the distribution pattern of clean water, identify the largest production capacity, and examine the distribution of the highest number of customers of clean water companies in several provinces in Indonesia using the K-Means algorithm. The data used includes distribution of delivery, potential production capacity (liters per second), and the number of customers of clean water companies across various provinces in Indonesia. The K-Means method is applied to cluster provinces based on the similarity of distribution patterns, production capacity, and customer numbers. The analysis results reveal significant clustering, where certain provinces have high clean water production capacity and a large number of customers, with the volume of water distributed matching the production capacity. These findings indicate a good balance between production capacity, the number of customers, and the volume of water distribution. This study focuses on the analysis of clean water distribution, the largest production capacity, and customers of clean water companies in various provinces in Indonesia. The insights gained can assist companies in designing more efficient distribution strategies to optimally meet customers needs across different regions. This contributes to the sustainable and equitable management of clean water resources.

Keyword: Clean Water Distribution, Production Capacity, Utility Customers, K-Means Algorithm.

ABSTRAK

Penelitian ini bertujuan untuk menganalisis pola distribusi air bersih, mengidentifikasi kapasitas produksi terbesar, serta mengetahui distribusi pelanggan perusahaan air bersih terbanyak pada beberapa provinsi di Indonesia dengan menerapkan algoritma K-Means. Data yang digunakan mencakup distribusi penyaluran, kapasitas produksi potensial (liter per detik), dan jumlah pelanggan perusahaan air bersih di berbagai provinsi di Indonesia. Metode K-Means diterapkan untuk mengelompokkan provinsi berdasarkan kemiripan pola distribusi, kapasitas produksi, dan jumlah pelanggan. Hasil analisis menunjukkan pengelompokan signifikan, di mana provinsi tertentu memiliki kapasitas produksi air bersih yang tinggi dan jumlah pelanggan yang juga banyak, dengan volume air yang disalurkan sesuai dengan kapasitas tersebut. Temuan ini mengindikasikan adanya keseimbangan yang baik antara kapasitas produksi, jumlah pelanggan, dan volume distribusi air. Penelitian ini berfokus pada analisis distribusi air bersih, kapasitas produksi terbesar, dan pelanggan pada perusahaan air bersih di berbagai provinsi di Indonesia. Wawasan yang dihasilkan dapat membantu perusahaan dalam merancang strategi distribusi yang lebih efisien untuk memenuhi kebutuhan pelanggan secara optimal di berbagai daerah. Hal ini memberikan kontribusi nyata terhadap pengelolaan sumber daya air bersih yang berkelanjutan dan merata.

PENDAHULUAN

Sampai saat ini bumi telah menunjukkan kemampuannya untuk memberikan kehidupan bagi makhluk hidup. Hal ini dilihat dari adanya pemanfaatan sumber daya alam yang digunakan manusia untuk memenuhi kebutuhan sehari-hari. Air bersih merupakan salah satu kebutuhan dasar yang sangat penting bagi keberlangsungan makhluk hidup dan keseimbangan lingkungan. Beragam aktivitas seperti industri, pertanian, mandi, mencuci, memasak, dan lainnya sangat bergantung pada ketersediaan air bersih (Triono, 2018). Akibat pertumbuhan ekonomi dan populasi, permintaan akan kebutuhan terhadap air bersih terus mengalami peningkatan (Kurniawan et al., n.d.). Untuk itu diperlukan adanya penyediaan air bersih yang memenuhi standar kualitas yang berlaku dan mampu mencukupi kebutuhan masyarakat secara berkelanjutan, sehingga berbagai aktivitas dapat berlangsung dengan lancar (Richardo Mampuk Tiny Mananoma & Tanudjaja, 2014).

Di Indonesia, distribusi air bersih sering kali tidak merata, dengan perbedaan yang signifikan dalam kapasitas produksi dan jumlah pelanggan air bersih antar provinsi. Beberapa daerah memiliki kapasitas produksi air yang melimpah, sementara daerah lainnya masih menghadapi kesulitan dalam memenuhi kebutuhan air bersih, karena adanya faktor-faktor yang mempengaruhi seperti keterbatasan infrastruktur, pertumbuhan jumlah penduduk dan kondisi iklim yang berubah-ubah (Sauqii & Daffa Akmal, 2023). Masalah ini sering kali disebabkan karena faktor geografis, sosial, dan ekonomi yang mempengaruhi cara distribusi yang dilakukan oleh perusahaan penyedia air. Dengan informasi yang komprehensif, pengelola sumber daya air dapat mengambil langkah-langkah yang lebih efektif dalam mengatasi kendala dalam distribusi air bersih (Fransiska et al., 2024). Dalam upaya penyediaan air bersih, sistem jaringan distribusi memegang peranan yang sangat penting. Tujuan dari sistem ini adalah untuk mengalirkan air bersih dari instalasi pengolahan ke masyarakat dengan

memenuhi standar kualitas, kuantitas, kontinuitas, serta tekanan yang memadai mencukupi (Zamzami et al., 2018).

Kapasitas produksi air bersih, yang diukur dalam liter per detik, merupakan indikator penting untuk menilai sejauh mana kebutuhan masyarakat dapat terpenuhi. Sistem pengelolaan sumber daya air yang terpadu dapat memastikan pasokan air yang diterima konsumen menerima pasokan air yang lebih efisien dan berkualitas (Kusumawardani & Astuti, 2018). Oleh karena itu, memastikan ketersediaan dan pengelolaan air bersih menjadi hal yang sangat penting untuk mendukung keberlangsungan hidup masyarakat (Harudu & Yanti, 2019). Pemenuhan produksi air bersih juga berperan penting dalam memenuhi kebutuhan domestik. Pentingnya penyediaan air untuk kelancaran aktivitas masyarakat, maka peningkatan kebutuhan terhadap pelayanan air bersih perlu ditindaklanjuti dengan penyediaan layanan yang memadai, baik oleh pemerintah, swasta maupun masyarakat itu sendiri.

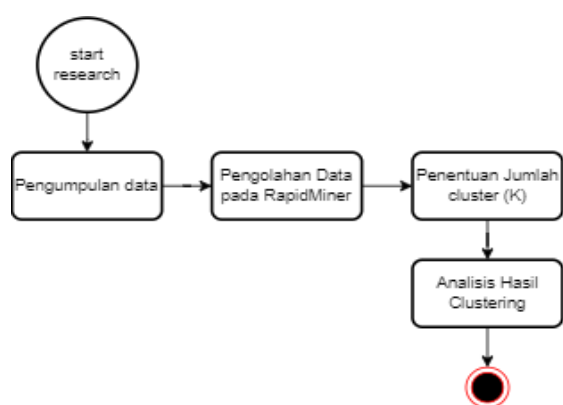
Di Indonesia, penyelenggaraan sistem penyediaan air bersih pada umumnya dilakukan oleh Perusahaan Daerah Air Minum (PDAM). Sistem penyediaan air bersih yang dikelola PDAM dalam menyediakan air bersih akan menghasilkan perbedaan kualitas dan kuantitas pelayanan yang berbeda dari satu kota atau kabupaten dengan kota atau kabupaten lainnya. Dalam mendapatkan air bersih diperlukan adanya perencanaan, desain, cara pengumpulan, dan distribusi yang baik. Sama halnya dengan yang dilakukan oleh PDAM (Agung Setiawan et al., 2020). PDAM terus berupaya melakukan peningkatan kualitas, kontinuitas, dan kuantitas debit air. Jumlah debit air yang disalurkan biasanya tergantung pada seberapa banyak jumlah penduduk dan jenis industri yang dilayani (Wardhana et al., 2013).

Penelitian ini bertujuan untuk menganalisis distribusi air bersih, kapasitas produksi air bersih (dalam liter per detik), serta jumlah pelanggan perusahaan air bersih yang disalurkan di berbagai provinsi di Indonesia. Tujuan utama adalah mengidentifikasi distribusi air bersih, kapasitas

produksi (dalam liter per detik), serta jumlah pelanggan perusahaan air bersih berdasarkan nilai tertinggi dengan menggunakan algoritma K-Means pada RapidMiner. Temuan ini diharapkan mampu memberikan wawasan yang lebih menyeluruh mengenai pola penyediaan air bersih, mendukung pengoptimalan strategi distribusi, serta memastikan terpenuhinya kebutuhan air bersih kepada masyarakat secara merata di berbagai wilayah. Beberapa penelitian terdahulu memberikan wawasan mengenai pentingnya analisis data dalam pengelolaan distribusi sumber daya (Putra et al., 2020). Dengan menggunakan pendekatan *clustering* untuk mengelompokkan daerah berdasarkan provinsi sesuai dengan tingkat ketersediaan sumber daya yang terbanyak (Sarmini et al., 2024). Pendekatan berbasis RapidMiner dengan menggunakan algoritma K-Means pada penelitian ini memberikan potensi besar untuk mengidentifikasi pola distribusi air bersih yang ada.

METODE PENELITIAN

Pada metode ini, alur penelitian dilakukan dengan beberapa tahapan yang meliputi dari pengumpulan data, pra-pemrosesan data, penerapan algoritma K-Means, dan analisis hasil. Setiap tahapan dirancang untuk memastikan data yang digunakan relevan dan mendukung tujuan dari penelitian ini. Metode penelitian ini diperjelas dengan menggunakan alur *flowchart*.



Gambar 1. Flowchart Metode Penelitian

Pengumpulan Data

Data yang digunakan bersumber dari Badan Pusat Statistik (BPS) yang merupakan lembaga pusat resmi penyedia data statistik di Indonesia

(Badan Pusat Statistik, 2022). Data tersebut meliputi distribusi air bersih, kapasitas produksi air bersih (liter per detik), serta jumlah pelanggan perusahaan air bersih di berbagai provinsi di Indonesia. Pemilihan data yang diambil dari BPS dipilih untuk memastikan informasi yang dianalisis akurat dan valid. BPS menyediakan data dari tahun 2018 hingga 2022, yang mencakup elemen penting seperti distribusi air bersih, kapasitas produksi, dan jumlah pelanggan di berbagai provinsi di Indonesia.

Pengolahan Data

Setelah data dari BPS diperoleh, kemudian dilakukannya pemrosesan data untuk memastikan data siap digunakan dalam analisis. Data yang diproses menggunakan RapidMiner untuk memastikan kualitasnya dengan melalui langkah-langkah seperti pembersihan data yang bertujuan untuk menghilangkan data yang tidak lengkap atau tidak relevan, seperti data dengan nilai yang hilang, identifikasi data kosong atau data yang tidak valid (missing values) atau data yang terduplikat. Penanganan data yang kosong atau yang tidak valid kemudian diatasi dengan metode imputasi atau penghapusan, serta seleksi atribut yang relevan untuk dianalisis.

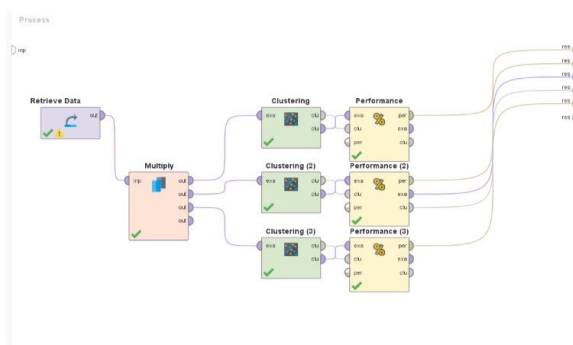
Penentuan Jumlah Cluster

Pada tahap ini, dilakukan adanya proses menentukan jumlah *cluster* (k) untuk analisis data. Algoritma K-Means diterapkan untuk mengelompokkan provinsi berdasarkan pola kesamaan dalam distribusi air bersih, kapasitas produksi, dan jumlah pelanggan (Annisa et al., 2024). Proses ini melibatkan penentuan jumlah *cluster* (k) untuk mencapai hasil yang optimal, perhitungan jarak antara data dan pusat *cluster* menggunakan metode jarak Davies Bouldin pada RapidMiner, serta pengelompokan data hingga iterasi mencapai konvergensi.

Analisis Hasil Clustering

Hasil pengelompokan ini kemudian dianalisis untuk memahami pola distribusi air bersih, mengidentifikasi provinsi dengan kapasitas produksi terbesar, serta melihat distribusi pelanggan terbanyak pada provinsi di Indonesia. Analisis ini divisualisasikan melalui *scatter plot* untuk memberikan interpretasi yang

lebih jelas dan informatif (Rayhan Informatika, 2013). Berdasarkan analisis *clustering* K-Means, dapat disimpulkan dan merekomendasikan kepada pembuat kebijakan untuk mempertimbangkan pola distribusi air bersih, kapasitas produksi, dan jumlah pelanggan dalam pengelolaan sumber daya air di Indonesia.



Gambar 2. Penentuan Jumlah Cluster

HASIL DAN PEMBAHASAN

Penelitian ini mengelompokkan provinsi di Indonesia berdasarkan pola distribusi air bersih, kapasitas produksi air bersih, dan jumlah pelanggan air bersih di berbagai provinsi di Indonesia. Dataset ini terdiri dari 34 baris data yang mempresentasikan 34 provinsi di Indonesia, dengan total 6 atribut berbeda. Hasil analisis data menggunakan metode K-Means *clustering* pada RapidMiner ini kemudian diproses agar data yang kosong dan tidak valid akan diproses sesuai dengan metode yang digunakan, seperti penghapusan data untuk menjaga kualitas dataset. Setelah itu, atribut yang sesuai dengan tujuan penelitian dipilih untuk memastikan bahwa analisis dilakukan pada informasi yang paling signifikan.

...	Provinsi	Distribusi air Bersih 2018	Distribusi air Bersih 2019	Distribusi air Bersih 2020	Distribusi air Bersih 2021	Distribusi air Bersih 2022
1	ACEH	38389	47756	55882	54456	56135
2	SUMATERA U...	316038	252088	325447	321041	331779
3	SUMATERA B...	91365	101030	106268	105405	107987
4	RIAU	16430	20710	18467	18547	20035
5	JAMBI	41755	40383	49133	48333	52647
6	SUMATERA S...	148106	153784	197242	197289	196250
7	BENGKULU	19056	25680	24721	24848	25489
8	LAMPUNG	16297	18181	17520	18232	19188
9	KEP. BANGK...	5791	8377	8725	8211	7871
10	KEP. RIAU	93872	116252	94082	95089	95179
11	DKI JAKARTA	499301	511855	494518	486417	507981
12	JAWA BARAT	395581	384202	419502	413528	429705

ExampleSet (34 examples, 6 special attributes, 6 regular attributes)

Gambar 3. Dataset awal Distribusi Penyaluran Air Bersih

...	Provinsi	Kapasitas Produksi (lit) 2018	Kapasitas Produksi (lit) 2019	Kapasitas Produksi (lit) 2020	Kapasitas Produksi (lit) 2021	Kapasitas Produksi (lit) 2022
1	ACEH	3787	4424	4426	4441	4384
2	SUMATERA U...	13437	12886	11782	14336	14270
3	SUMATERA B...	5802	6012	6631	6960	7171
4	RIAU	2315	2462	2489	2512	2583
5	JAMBI	4435	4542	4719	4655	4999
6	SUMATERA S...	10326	10705	10000	10183	10513
7	BENGKULU	1732	2067	1966	1779	1899
8	LAMPUNG	1784	1793	1790	2043	1882
9	KEP. BANGK...	933	965	1053	1118	1118
10	KEP. RIAU	5000	4955	5095	5042	4847
11	DKI JAKARTA	20085	18108	22280	22280	22067
12	JAWA BARAT	21710	21875	23742	24370	24428
13	JAWA TENGAH	30160	31019	32171	33405	35005

ExampleSet (34 examples, 6 special attributes, 6 regular attributes)

Gambar 4. Dataset Awal Kapasitas Produksi Air Bersih (Liter/Detik)

...	Provinsi	Pelanggan Tahun 2018	Pelanggan Tahun 2019	Pelanggan Tahun 2020	Pelanggan Tahun 2021	Pelanggan Tahun 2022
1	ACEH	180903	221941	211554	218367	231301
2	SUMATERA UTARA	845180	886146	894058	902348	938865
3	SUMATERA BARAT	584403	570259	578634	601629	607966
4	RIAU	86312	91389	89403	89508	92044
5	JAMBI	872881	943591	959891	974113	960149
6	SUMATERA SELATAN	874112	873312	862169	869400	876878
7	BENGKULU	88815	91641	90049	90696	92780
8	LAMPUNG	79729	85405	86220	85870	91277
9	KEP. BANGKA BELITU	30439	35008	31058	33007	34873
10	KEP. RIAU	315008	327946	332724	333382	334543
11	DKI JAKARTA	884535	885353	896782	918389	969512
12	JAWA BARAT	1623584	1831042	1870000	2005845	2034772
13	JAWA TENGAH	1834201	1755884	1864587	1867464	1941552

ExampleSet (34 examples, 6 special attributes, 6 regular attributes)

Gambar 5. Dataset Awal Pelanggan Air Bersih

Penerapan Cluster

Menentukan jumlah *cluster* adalah langkah penting dalam penerapan algoritma K-Means, karena dalam pemilihan jumlah *cluster* yang tepat akan berdampak pada kualitas hasil pengelompokan. Pada penerapan analisis ini metode Davies-Bouldin Index (DBI) digunakan untuk menentukan jumlah *cluster* agar pengelompokan menghasilkan hasil yang optimal (Uki Eka Saputri et al., 2022). Metode ini melibatkan perhitungan nilai indeks Davies-Bouldin untuk berbagai jumlah *cluster*, kemudian menganalisis nilai indeks tersebut. Nilai Davies-Bouldin (DBI) yang lebih rendah menunjukkan kualitas *cluster* yang lebih baik, dengan *cluster* yang lebih padat dan terpisah secara optimal. Pendekatan ini dipilih karena memberikan hasil kuantitatif yang jelas tentang keseimbangan antara kepadatan *cluster* dan pemisah antar-*cluster*. Dengan menentukan jumlah *cluster* yang tepat, penggunaan metode DBI menghasilkan pengelompokan yang menggambarkan pada pola distribusi air bersih, kapasitas produksi, dan jumlah pelanggan dengan lebih akurat pada masing masing provinsi.

Cluster Model

```
Cluster 0: 26 items
Cluster 1: 3 items
Cluster 2: 5 items
Total number of items: 34
```

Gambar 6. Cluster Distribusi Penyaluran Air Bersih

Model *cluster* pada distribusi penyaluran air bersih menunjukkan pembagian data ke dalam tiga klaster. *Cluster* 0 terdiri dari 26 item, *cluster* 1 terdiri dari 3 item, dan yang terakhir *cluster* 2 yang terdiri dari 5 item. Masing-masing dari *cluster* menggambarkan variasi signifikan dalam tingkat penyaluran air bersih. *Cluster* 0 yang terdiri dari 26 item merupakan *cluster* dengan jumlah terbanyak, namun tingkat penyaluran air bersih relatif lebih rendah dari *cluster* lainnya. *Cluster* 1 yang terdiri dari 3 item menunjukkan 3 provinsi dengan tingkat penyaluran air bersih yang tertinggi. Sedangkan *cluster* 2 dengan 5 item menunjukkan 5 provinsi dengan penyaluran air bersih yang sedang.

Cluster Model

```
Cluster 0: 22 items
Cluster 1: 4 items
Cluster 2: 8 items
Total number of items: 34
```

Gambar 7. Cluster Kapasitas Produksi Air Bersih (Liter/Detik)

Model *cluster* pada kapasitas produksi air (liter/detik) menunjukkan variasi kapasitas antar *cluster*. Pada *cluster* 0 yang mencakup 22 items memiliki kapasitas produksi paling rendah. *Cluster* 1 meskipun terdiri dari 4 item merupakan provinsi yang memiliki kapasitas produksi air bersih tertinggi. Sedangkan *cluster* 2 dengan 8 item yaitu 8 provinsi hanya memiliki kapasitas produksi air bersih sedang.

Cluster Model

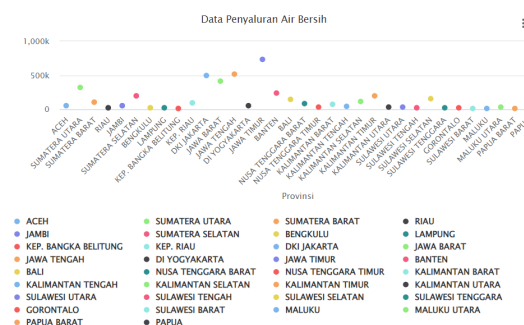
```
Cluster 0: 21 items
Cluster 1: 3 items
Cluster 2: 10 items
Total number of items: 34
```

Gambar 8. Cluster Pelanggan Air Bersih

Model *cluster* pada pelanggan air bersih menunjukkan variasi jumlah pelanggan antar *cluster*. Pada *cluster* 0 yang mencakup 21 item, yaitu 21 provinsi ini memiliki pelanggan air bersih terendah, sedangkan *cluster* 1 yang terdiri dari 3 *cluster* yaitu 3 provinsi ini mencakup pelanggan dalam jumlah terbanyak dari seluruh provinsi di Indonesia. Dan yang terakhir *cluster* 2 dengan 10 item yaitu 10 provinsi ini mempresentasikan pelanggan air bersih dalam jumlah menengah. Secara keseluruhan model ini menggambarkan distribusi jumlah pelanggan air bersih dari total 34 item yaitu 34 provinsi pelanggan air bersih yang dikelompokkan ke dalam tiga *cluster*.

Hasil Analisis

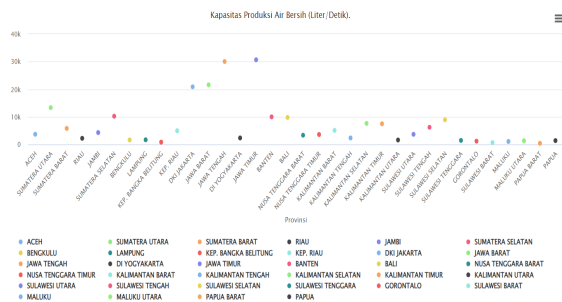
Berdasarkan hasil pengelompokan dengan menggunakan algoritma K-Means, diperoleh beberapa *cluster* yang mencerminkan pola distribusi air bersih, kapasitas produksi, dan jumlah pelanggan diberbagai provinsi di Indonesia. Hasil pengelompokan ini menunjukkan variasi antar provinsi dalam hal kebutuhan, seperti penyediaan dan pengelolaan air bersih, serta mengidentifikasi provinsi-provinsi berdasarkan nilai Davies-Bouldin (DBI) dari ke tiga atribut tersebut. Dengan berbasis RapidMiner dengan menggunakan metode K-Means, hasil analisis yang diperoleh disajikan sebagai berikut.



Gambar 9. Hasil Analisis Distribusi Penyaluran Air Bersih

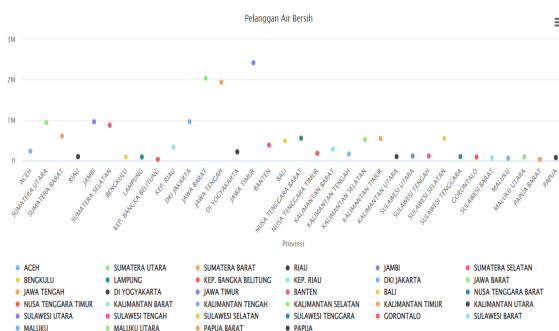
Dari hasil analisis pada RapidMiner dengan menggunakan algoritma K-Means menunjukkan pola penyaluran air bersih di berbagai provinsi di Indonesia. Berdasarkan data hasil *clustering* analisis provinsi dengan kapasitas penyaluran air bersih tertinggi adalah Jawa Timur, dengan angka hampir mendekati 1,000 ribu liter. Sedangkan

provinsi seperti Jawa Tengah, DKI Jakarta, dan Jawa Barat memiliki kapasitas penyaluran sedang yang berada pada posisi tengah pada grafik. Dan pada sisi lain, banyak dari beberapa provinsi menempati kapasitas penyaluran terendah, terlihat dari grafik banyak menempati bagian bawah. Hasil analisis ini menganalisis perbedaan yang cukup signifikan dalam kapasitas penyaluran air bersih antar provinsi, dengan Jawa Timur sebagai provinsi dengan kapasitas penyaluran tertinggi.



Gambar 10. Hasil Analisis Kapasitas Produksi Air Bersih (Liter/Detik)

Hasil analisis dari RapidMiner dengan menggunakan algoritma K-Means menunjukkan Kapabilitas Produksi Air Bersih (liter/detik) di berbagai provinsi di Indonesia. Berdasarkan data hasil *clustering* analisis provinsi dengan kapasitas produksi air bersih tertinggi adalah Jawa Timur, dan Jawa Tengah yang mencapai sekitar **30.000 L/detik**. Beberapa provinsi seperti Jawa Barat dan DKI Jakarta, memiliki kapasitas produksi yang cukup. Sedangkan provinsi lain yang memiliki kapasitas produksi yang rendah berada pada grafik paling bawah. Kapasitas produksi masih dibawah 20.000 L/detik dan juga masih banyak dari beberapa provinsi yang kapasitas produksi yang diterima masih dibawah 10.000 L/detik.



Gambar 11. Hasil Analisis Pelanggan Air Bersih

Hasil analisis pada RapidMiner dengan menggunakan algoritma K-Means menunjukkan pelanggan air bersih di berbagai provinsi di Indonesia. Berdasarkan data hasil *clustering*, analisis provinsi dengan jumlah pelanggan terbanyak terdapat pada provinsi Jawa Timur, pelanggan yang tercatat hampir memasuki 3 juta pelanggan. Sedangkan provinsi seperti DKI Jakarta dan Jawa Barat, juga menunjukkan jumlah pelanggan yang cukup tinggi yaitu masih dalam rata-rata 2 juta pelanggan. Sementara itu banyak dari beberapa provinsi dengan jumlah pelanggan yang lebih sedikit, berada pada bagian bawah grafik dengan jumlah pelanggan jauh lebih rendah. Rata-rata jumlah pelanggan dari provinsi terendah hanya mencapai sekitar 1 juta bahkan masih banyak provinsi dengan jumlah pelanggan dibawah 1 juta pelanggan air bersih. Analisis ini menunjukkan ketidakmerataan dalam distribusi pelanggan air bersih antar provinsi, antara Jawa Timur sebagai provinsi dengan jumlah pelanggan tersebut lainnya.

KESIMPULAN

Berdasarkan hasil analisis dari data distribusi air bersih, kapasitas produksi (L/detik), dan jumlah pelanggan pada perusahaan air bersih di berbagai provinsi di Indonesia, dapat disimpulkan bahwa terdapat perbedaan yang signifikan antar provinsi. Jawa Timur adalah provinsi yang distribusi dan kapasitas produksi air bersih tertinggi yang mencapai sekitar **30.000 L/detik serta memiliki jumlah pelanggan air bersih terbanyak yaitu hampir 3 juta pelanggan**. Sedangkan provinsi seperti Jawa Barat, Jawa Tengah dan DKI Jakarta menunjukkan analisis data kapasitas penyaluran dan jumlah pelanggan yang relatif tinggi, sementara itu masih banyak provinsi lain di Indonesia selain Jawa Timur, Jawa Barat, Jawa Tengah dan DKI Jakarta yang memiliki distribusi, kapasitas produksi air dan jumlah pelanggan yang jauh lebih rendah. Hal ini mencerminkan adanya ketidakmerataan dalam distribusi air bersih di Indonesia.

Penerapan algoritma K-Means berbasis RapidMiner pada analisis dalam penelitian ini berhasil mengelompokkan data yang mencakup dari provinsi-provinsi di Indonesia berdasarkan

pola distribusi air bersih, kapasitas produksi, dan jumlah pelanggan yang serupa. Analisis ini menunjukkan pentingnya pengelolaan dan perencanaan yang lebih efisien, dengan fokus pada provinsi-provinsi yang memiliki distribusi dalam kapasitas dan jumlah pelanggan yang lebih rendah, guna memastikan pemerataan distribusi air bersih diseluruh provinsi di Indonesia. Dengan demikian hasil dari analisis penelitian ini dapat menjadikan dasar untuk meningkatkan sistem distribusi air bersih yang lebih merata di seluruh wilayah Indonesia.

DAFTAR PUSTAKA

- Agung Setiawan, Eko Riyanto, & Among Prayogo. (2020). *Analisa Sistem Distribusi Air Bersih PDAM Purwodadi, Purworejo*.
- Annisa, K., Serasi Ginting, B., & Syari, M. A. (n.d.). *Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode Clustering Pada PDAM Langkat*.
- Badan Pusat Statistik. (2022). *Statistik Air Bersih Indonesia 2018-2022*.
<https://www.bps.go.id>
- Fransiska, G., Sari, A., Yolanda, D., Negeri, U., Rayi, S., Rajib, K., Kampus, A. :, Gunungpati, S., & Tengah, S. J. (2024). Krisis Air Menangani Penyediaan Air Bersih Di Dunia Yang Semakin Kekurangan Sumber Daya. *Jurnal Ilmiah Research Student*, 1(5), 334–341.
<https://doi.org/10.61722/jirs.v1i5.1373>
- Harudu, L., & Yanti, D. (2019). Analisis Kualitas Fisika Kimia Air Hujan di Desa Darawa Berdasarkan Standar Kualitas Air Bersih Kecamatan Kaledupaselatan Kabupaten Wakatobi. *Jurnal Penelitian Pendidikan Geografi*, 4.
<https://doi.org/10.36709/jppg.v4i1.5597>
- Kurniawan, M. A., Fitriani, H., & Hadinata, F. (2021). *Analisis Kebutuhan Penyediaan Air Bersih di Kota Palembang Analysis of Water Demand Supply in Palembang City*.
[https://doi.org/10.25299/saintis2021.vol21\(02\).7611](https://doi.org/10.25299/saintis2021.vol21(02).7611)
- Kusumawardani, Y., & Astuti, W. (2018). Evaluasi Pengelolaan Sistem Penyediaan Air Bersih Di Pdam Kota Madiun. In *Jurnal Neo Teknika* (Vol. 4, Issue 1).
- Putra, W. B., Indra, N., Dewi, K., & Busono, T. (2020). *Penyediaan Air Bersih Sistem Kolektif: Analisis Kebutuhan Air Bersih Domestik pada Perumahan Klaster. Jurnal Arsitektur TERRACOTTA* |, 1(2), 115–123.
- Rayhan Informatika, F. (2024). Visualisasi Data Untuk Analisis Risiko: Teknik Dan Aplikasi. In *Teknologipintar.org* (Vol. 4, Issue 5).
- Richardo Mampuk Tiny Mananoma, C., & Tanudjaja, L. (2014). Pengembangan Sistem Penyediaan Air Bersih di Kecamatan Poso Kota Sulawesi Tengah. *Jurnal Sipil Statik*, 2(5), 233–241.
- Sarmini, S., Ma'arifah, W., & Tahyudin, I. (2024). Sistem Pendukung Keputusan Berbasis K-Means untuk Evaluasi Keberhasilan Bisnis dan Nilai Perusahaan. *Jurnal Sistem Informasi Bisnis*, 14(4), 363–374.
<https://doi.org/10.21456/vol14iss4pp363-374>
- Sauqii, & Daffa Akmal. (2023). *Analisa Kualitas Air Dalam Pemenuhan Kebutuhan Air di Indonesia Pada Era 4.0*. 99, 1–6.
- Triono, M. O. (2018). Akses Air Bersih Pada Masyarakat Kota Surabaya Serta Dampak Buruknya Akses Air Bersih Terhadap Produktivitas Masyarakat Kota Surabaya. *Jurnal Ilmu Ekonomi Terapan*, 3(2).
<https://doi.org/10.20473/jiet.v3i2.10072>
- Uki Eka Saputri, D., Hidayat, T., & Masturoh, S. (2022). *Period Using the K-Means Algorithm*. 12(2), 2012–2017.
<https://doi.org/10.30700/jst.v12i2.1241>
- Wardhana, W. I., Budihardjo, Ma., & Scylla Adhesti, dan P. (2013). *Kajian Sistem Penyediaan Air Bersih Sub Sistem Bribin Kabupaten Gunungkidul*.
- Zamzami, Z., Azmeri, A., & Syamsidik, S. (2018). Sistem Jaringan Distribusi Air Bersih PDAM Tirta Tawar Kabupaten Aceh. *Jurnal Arsip Rekayasa Sipil Dan Perencanaan*, 1(1), 132–141.
<https://doi.org/10.24815/jarsp.v1i1.10330>