Pengelompokkan Sumber Air Minum Dari Air Sungai Menggunakan Metode K-Means

Sabrina Biutiqwin Sinaga¹, Solikhun², Dedi Suhendro³

¹Mahasiswa Program Studi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar

^{2,3}AMIK Tunas Bangsa, Pematangsiantar

¹sabrinabiutiqwinsinaga@gmail.com, ²solikhun@amiktunasbangsa.ac.id,

³dedi.su@amiktunasbangsa.ac.id

Abstract

River water is one of the most frequently used water by the community and has a multipurpose function for life, one of which is a source of drinking water. However, now we know that the population of river water pollution is very high and it is used as a waste disposal site which causes a lot of river water to be polluted, it can make people susceptible to disease because they consume unhealthy river water. Judging from the data obtained by province, many use river water as a source of drinking water, for this reason the authors conducted a study that aims to classify drinking water sources from river water by province using the K-means Clustering algorithm and will test it with the Rapidminer application, so that Data from 34 provinces will be divided into 3 clusters in which cluster 1 (C1) is a high group, cluster 2 (C2) is a medium group, and cluster 3 (C3) is a low group. The results obtained from this study are C1 with a total of 2 provinces, C2 with a total of 9 provinces, C3 with a total of 23 provinces and the value of the results carried out with the Rapidminer application has the same value. With this research, it is hoped that this can provide information for the government about the data on the grouping of drinking water sources and used as a consideration for overcoming polluted rivers.

Keywords: Centroid, Clustering, K-Mean, Mining.

Abstrak

Air sungai termasuk air yang paling sering digunakan oleh masyarakat dan menjadi fungsi serbaguna bagi kehidupan, salah satunya sebagai sumber air minum. Akan tetapi sekarang ini yang kita ketahui populasi pencemaran air sungai sangat tinggi dan dijadikan tempat pembuangan limbah yang menyebabkan banyak air sungai tercemar, itu dapat membuat masyarakat dengan mudah terkena penyakit karena mereka mengonsumsi air sungai yang tidak sehat. Dilihat dari data yang diperoleh berdasarkan provinsi banyak yang menggunakan air sungai sebagai sumber air minum, untuk itu penulis melakukan sebuah penelitian yang bertujuan untuk mengelompokkan sumber air minum dari air sungai berdasarkan provinsi dengan menggunakan algoritma K-means Clustering dan akan mengujinya dengan aplikasi Rapidminer, sehingga data dari 34 provinsi tersebut akan dibagi menjadi 3 cluster dimana cluster 1 (C1) kelompok yang tinggi, cluster 2 (C2) kelompok yang sedang, dan cluster 3 (C3) yaitu kelompok yang rendah. Hasil yang diperoleh dari penelitian ini yaitu C1 dengan jumlah 2 provinsi, C2 dengan jumlah 9 provinsi, C3 dengan jumlah 23 provinsi dan nilai hasil yang dilakukan dengan aplikasi Rapidminer bernilai sama. Dengan adanya penelitian ini diharapkan dapat menjadi informasi bagi pemerintah tentang data pengelompokan sumber air minum dan dijadikan sebagai bahan pertimbangan untuk menanggulangi sungai yang tercemar.

Kata Kunci: Centroid, Clustering, K-Mean, Mining.

1. PENDAHULUAN

Air merupakan sumber kehidupan dan sangat penting bagi manusia [1]. Tanpa air maka manusia juga tidak akan bertahan hidup lebih lama, karena air adalah sumber kehidupan dan tanda dari adanya kehidupan. Air sungai termasuk air yang sering digunakan oleh masyarakat dan menjadi fungsi serbaguna bagi

kehidupan salah satunya sebagai sumber air minum. Namun saat ini, populasi pencemaran air sungai sangat tinggi dan digunakan sebagai tempat pembuangan limbah yang menyebabkan banyak air sungai tercemar dan itu dapat membuat masyarakat dengan mudah terkena penyakit karena mereka mengonsumsi air sungai yang tidak sehat. Sumber pencemaran lingkungan yang terbesar yaitu sering membuang sampah kesungai, limbah rumah tangga, limbah pabrik, bahan organik, bahan anorganik, serta beberapa bahan kimia lainnya yang telah banyak ditemukan dalam air yang di gunakan.

Pemerintah membutuhkan informasi tentang data pengelompokan sumber air minum yang berasal dari sungai berdasarkan provinsi yang mana data tersebut dapat dijadikan sebagai bahan pertimbangan untuk menanggulangi sungai yang tercemar oleh limbah sehingga masyarakat dapat mengonsumsi air minum yang bersih. Dengan adanya penelitian ini diharapkan dapat menjadi informasi bagi pemerintah agar lebih menekankan dan memberi peringatan pada pabrik – pabrik dan perusahaan industri yang tidak memiliki tempat pembuangan limbah khusus, kemudian untuk masyarakat agar pemerintah dapat memberikan fasilitas air bersih dan menghimbau masyarakat agar selalu menjaga lingkungan dengan tidak mengalirkan limbah dan sampah ke sungai sehingga terhindar dari berbagai macam penyakit, reboisasi agar dapat membantu mencegah adanya banjir, efisiensi dalam penggunaan air dengan menggunakan air seperlunya saja.

2. METODOLOGI PENELITIAN

Adapun diagram alur pada penelitian ini diuraikan dalam rancangan pada gambar 1. dibawah ini:

Gambar 1. Alur Perancangan Penelitian

3. HASIL DAN PEMBAHASAN

3.1. Pengolahan Data

Pengolahan data pada penelitian ini menguraikan perhitungan yang digunakan penulis, adapun tahapan yang dilakukan yaitu: Menentukan data k yang akan di *cluster* sampai *cluster* R34, berikut adalah hasil dari nilai rata-rata untuk *cluster* pertama:

Tabel 1. Rata-Rata Sumber Air Minum

No	Nama	Nilai Rata-Rata
1	Aceh	3,608
2	Sumatera Utara	3,215
3	Sumatera Barat	3,284
4	Riau	2,116
5	Jambi	7,422
6	Sumatera Selatan	8,394
7	Bengkulu	8,394
8	Lampung	1,192
9	Kepulauan Bangka Belitung	0,983
10	Kepulauan Riau	1,075
...	...	...
33	Papua Barat	10,593
34	Papua	11,362

Dengan nilai k jumlah *cluster centroid* awal sumber air minum sebanyak 3 *cluster* (k-3). *Cluster* yang ditentukan yaitu *cluster* tinggi, *cluster* sedang, *cluster* rendah, seperti table 2 dibawah ini:

Tabel 2. Nilai Data *Centroid* Awal

Cluster	Nilai
Tinggi (Max/C1)	30,421
Sedang (Average/C2)	5,250
Rendah (Min/C3)	0,020

- Menentukan nilai pusat *cluster centroid* awal (iterasi 1) yang telah ditentukan secara random/acak berdasarkan nilai data yang telah di tentukan sebelumnya. Adapun nilai *cluster* dapat dilihat pada tabel 3.2 di atas.
- Menghitung jarak antara rata-rata dengan nilai k *centroid* awal yang telah ditentukan dengan menggunakan rumus perhitungan sebagai berikut:

$$D(x,y) = \sqrt{\sum_{i=1}^n (X_{1i} - Y_{1i})^2} \quad (1)$$

Dimana:

$D(x,y)$ = Ukuran Ketidak miripan.

$X_i = (X_1, X_2, \dots)$ yaitu variabel centroid awal.

$Y_i = (Y_1, Y_2, \dots)$ yaitu variabel data.

Adapun tabel iterasi 1 dapat dilihat pada tabel 3 dibawah ini:

Tabel 3. Iterasi 1

No.	Nama	Nilai Rata-Rata	C1	C2	C3
1	Aceh	3,608	26,813	1,642	3,588
2	Sumatera Utara	3,215	27,206	2,035	3,195
3	Sumatera Barat	3,284	27,137	1,966	3,264
4	Riau	2,116	28,305	3,134	2,096
5	Jambi	7,422	22,999	2,172	7,402
6	Sumatera Selatan	8,394	22,027	3,144	8,374
7	Bengkulu	2,291	28,130	2,959	2,271

No.	Nama	Nilai Rata-Rata	C1	C2	C3
8	Lampung	1,192	29,229	4,058	1,172
9	Kepulauan Bangka Belitung	0,983	29,438	4,267	0,963
10	Kepulauan Riau	1,075	29,346	4,175	1,055
...	...	...	...	...	...
33	Papua Barat	10,593	19,828	5,343	10,573
34	Papua	11,362	19,059	6,112	11,342

- c) Mengelompokkan berdasarkan jarak terdekat berdasarkan centroid, seperti tabel 4 dibawah ini:

Tabel 4. Jarak Terpendek Iterasi 1

C1	C2	C3	Jarak Terpendek	Hasil
26,813	1,642	3,588	1,642	C2
27,206	2,035	3,195	2,035	C2
27,137	1,966	3,264	1,966	C2
28,305	3,134	2,096	2,096	C3
22,999	2,172	7,402	2,172	C2
22,027	3,144	8,374	3,144	C2
28,130	2,959	2,271	2,271	C3
29,229	4,058	1,172	1,172	C3
29,438	4,267	0,963	0,963	C3
29,346	4,175	1,055	1,055	C3
30,401	5,230	0,000	0,000	C3
...	...	...	...	...
25,829	0,658	4,572	0,658	C2
19,828	5,343	10,573	5,343	C2
19,059	6,112	11,342	6,112	C2

Sehingga dapat diperoleh hasil *cluster* iterasi 1 pada tabel 5 seperti dibawah ini:

Tabel 5. Hasil Iterasi 1

Cluster	Provinsi	Nilai	Hasil
C1	(20,21)	6%	2
C2	(1,2,3,5, 6,19,22,23,24,26,29,30,32,33,34)	44%	15
C3	(4,7,8,9,10,11,12,13,14,15,16,17,18,25,27,28,31)	50%	17

- d) Melanjutkan ke iterasi selanjutnya hingga *centroid* bernilai sama.
 Jarak terdekat pada iterasi 2 berdasarkan centroid, seperti tabel 6 dibawah ini:

Tabel 6. Jarak Terpendek Iterasi 2

C1	C2	C3	Jarak Terpendek	Hasil
21,914	3,561	2,437	2,437	C3
22,307	3,954	2,044	2,044	C3
22,238	3,885	2,113	2,113	C3
23,406	5,053	0,945	0,945	C3
18,100	0,253	6,251	0,253	C2

C1	C2	C3	Jarak Terpendek	Hasil
17,128	1,225	7,223	1,225	C2
23,231	4,878	1,120	1,120	C3
24,330	5,977	0,021	0,021	C3
...	...	...	...	...
20,930	2,577	3,421	2,577	C2
14,929	3,424	9,422	3,424	C2
14,160	4,193	10,191	4,193	C2

Sehingga dapat diperoleh hasil *cluster* iterasi 2 pada tabel 7 seperti dibawah ini:

Tabel 7. Hasil Iterasi 2

Cluster	Provinsi	Nilai	Hasil
C1	(20,21)	6%	2
C2	(5,6,19,22,23,24,26,30,32,33,34)	32%	11
C3	(1,2,3,4,7,8,9,10,11,12,13,14,15,16,17,18,25,27,28,29,31)	62%	21

Melanjutkan ke iterasi selanjutnya hingga *centroid* bernilai sama dan untuk hasil yang bernilai sama penulis melakukan 4 iterasi yang langkah-langkahnya dilakukan seperti iterasi 2. Adapun hasil dari iterasi 3 dan 4 yaitu seperti tabel 8 dan 9 dibawah ini:

Tabel 8. Hasil Iterasi 3

Cluster	Provinsi	Nilai	Hasil
C1	(20,21)	6%	2
C2	(5,6,22,23,24,26,30,33,34)	26%	9
C3	(1,2,3,4,7,8,9,10,11,12,13,14,15,16,17,18,19,25,27,28,29,31,32)	68%	23

Tabel 9. Hasil Iterasi 4

Cluster	Provinsi	Nilai	Hasil
C1	(20,21)	6%	2
C2	(5,6,22,23,24,26,30,33,34)	32%	11
C3	(1,2,3,4,7,8,9,10,11,12,13,14,15,16,17,18,19,25,27,28,29,31,32)	62%	21

4.2. Hasil Percobaan

Setelah operator saling berhubungan maka tahap selanjutnya yaitu dengan meng klik simbol *play* sehingga hasil dari pengelompokan dalam penggunaan *Rapidminer* dapat dilihat seperti gambar 2 dibawah ini:

Cluster Model

Cluster 0: 2 items
Cluster 1: 9 items
Cluster 2: 23 items
Total number of items: 34

Gambar 2. Hasil Nilai *Cluster*

Berdasarkan dari pengolahan data di atas mengenai tahap-tahap dari pengolahan dan percobaan yang telah dilakukan maka pada gambar 2 berikut adalah hasil keterkaitan antara hasil perhitungan manual dengan hasil yang di tampilkan oleh *Rapidminer*.

Gambar 3. Tampilan Hasil Perhitungan Manual

Gambar 3 diatas menampilkan hasil dari data yang digunakan dalam perhitungan manual.

Row No.	Nama	cluster	Nilai Rata...
1	Aceh	cluster_2	3.608
2	Sumatera Utara	cluster_2	3.215
3	Sumatera Barat	cluster_2	3.284
4	Riau	cluster_2	2.116
5	Jambi	cluster_1	7.422
6	Sumatera Selatan	cluster_1	8.394
7	Bengkulu	cluster_2	2.291
8	Lampung	cluster_2	1.192
9	Kepulauan Bangka Belitung	cluster_2	0.983
10	Kepulauan Riau	cluster_2	1.075
11	DKI Jakarta	cluster_2	0.020
12	Jawa Barat	cluster_2	0.523
13	Jawa Tengah	cluster_2	0.549
14	DI Yogyakarta 4	cluster_2	0.102
15	Jawa Timur	cluster_2	0.543
16	Banten	cluster_2	1.712
17	Bali	cluster_2	1.484

Row No.	Nama	cluster	Nilai Rata...
18	Nusa Tenggara Barat	cluster_2	0.958
19	Nusa Tenggara Timur	cluster_2	4.859
20	Kalimantan Barat	cluster_0	20.622
21	Kalimantan Tengah	cluster_0	30.421
22	Kalimantan Selatan	cluster_1	16.215
23	Kalimantan Timur	cluster_1	8.975
24	Kalimantan Utara	cluster_1	6.120
25	Sulawesi Utara	cluster_2	0.324
26	Sulawesi Tengah	cluster_1	6.539
27	Sulawesi Selatan	cluster_2	1.873
28	Sulawesi Tenggara	cluster_2	2.234
29	Gorontalo	cluster_2	3.915
30	Sulawesi Barat	cluster_1	8.445
31	Maluku	cluster_2	1.929
32	Maluku Utara	cluster_2	4.592
33	Papua Barat	cluster_1	10.593
34	Papua	cluster_1	11.362

Gambar 4. Tampilan Hasil Perhitungan *Rapidminer*

Gambar 4 diatas menampilkan hasil dari data yang digunakan dalam perhitungan *Rapidminer*.

Gambar 5. Hasil Akhir Pengelompokan *Rapidminer*

Pada Gambar 5 dapat dilihat dimana *cluster 0* pada *rapidminer* adalah *cluster* tinggi yang terbagi menjadi 2 provinsi dapat dilihat pada titik merah yang berjumlah 2 titik, *cluster* sedang pada *rapidminer* yaitu *cluster 1* yang berjumlah 9 provinsi dapat dilihat pada titik hijau, sedangkan *cluster* rendah pada *rapidminer* berjumlah 23 provinsi yang dapat dilihat pada titik biru *cluster 2*. Adapun Provinsi yang termasuk kedalam *cluster* tinggi yaitu seperti tabel 10 berikut:

Tabel 10. *Cluster* Tinggi

No.	Nama Provinsi	Cluster 1
1	Kalimantan Barat	Tinggi
2	Kalimantan Tengah	Tinggi

Provinsi yang termasuk kedalam *cluster* sedang yaitu:

Tabel 11. *Cluster* Sedang

No.	Nama Provinsi	Cluster 2
1	Jambi	Sedang
2	Sumatera Selatan	Sedang
3	Kalimantan Selatan	Sedang
4	Kalimantan Timur	Sedang
5	Kalimantan Utara	Sedang
6	Sulawesi Tengah	Sedang
7	Sulawesi Barat	Sedang

No.	Nama Provinsi	Cluster 2
8	Papua Barat	Sedang
9	Papua	Sedang

Sedangkan provinsi yang termasuk kedalam *cluster* rendah yaitu:

Tabel 12. *Cluster* Rendah

No.	Nama Provinsi	Cluster 3
1	Aceh	Rendah
2	Sumatera Utara	Rendah
4	Sumatera Barat	Rendah
5	Riau	Rendah
6	Bengkulu	Rendah
7	Lampung	Rendah
8	Kepulauan Bangka Belitung	Rendah
9	Kepulauan Riau	Rendah
10	DKI Jakarta	Rendah
11	Jawa Barat	Rendah
12	Jawa Tengah	Rendah
13	DI Yogyakarta ⁴	Rendah
14	Jawa Timur	Rendah
15	Banten	Rendah
16	Bali	Rendah
17	Nusa Tenggara Barat	Rendah
18	Nusa Tenggara Timur	Rendah
19	Sulawesi Utara	Rendah
20	Sulawesi Selatan	Rendah
21	Sulawesi Tenggara	Rendah
22	Gorontalo	Rendah
23	Maluku	Rendah

4. SIMPULAN

Hasil akhir penelitian dari 34 provinsi dengan masing-masing nilai *cluster* disimpulkan dengan:

- Cluster* Tertinggi (C1) dengan jumlah sebanyak 2 provinsi yaitu: Kalimantan Barat, Kalimantan Tengah.
- Cluster* Sedang (C2) dengan jumlah sebanyak 9 provinsi yaitu: Jambi, Sumatera Selatan, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Tengah, Sulawesi Barat, Papua Barat, Papua.
- Cluster* Rendah (C3) dengan jumlah sebanyak 23 provinsi yaitu: Aceh, Sumatera Utara, Sumatera Barat, Riau, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara, Maluku, Maluku Utara.

- d) Proses pemberhentian iterasi pada pengujian data dalam penelitian ini terjadi pada iterasi ke 4.
- e) Nilai hasil yang dilakukan dengan perhitungan manual dan dengan aplikasi *rapidminer* bernilai sama

DAFTAR PUSTAKA

- [1] A. S. Suryani, "Persepsi Masyarakat Dalam Pemanfaatan Air Bersih (Studi Kasus Masyarakat Pinggir Sungai Di Palembang) Public Perception On The Use Of Clean Water Anih Sri Suryani," 2016.
- [2] R. Setiawan And N. Tes, "Penerapan Data Mining Menggunakan Algoritma K-Means *Clustering* Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik Lp3i Jakarta)," Vol. 3, No. 1, Pp. 76-92, 2016.
- [3] N. Lihayati, R. E. Pawening, And M. Furqan, "Klasifikasi Jenis Daging Berdasarkan Tekstur Menggunakan Metode Gray Level Coocurent Matrix," *Pros. Sentia*, Vol. 8, No. 1994, Pp. 305-310, 2016.
- [4] E. F. Anggara, T. W. Widodo, And D. Lelono, "Deteksi Daging Sapi Menggunakan Electronic Nose Berbasis Bidirectional Associative Memory," *Ijeis (Indonesian J. Electron. Instrum. Syst.*, Vol. 7, No. 2, P. 209, 2017, Doi: 10.22146/Ijeis.25489.
- [5] R. Novianto, "Penerapan Data Mining Menggunakan Algoritma K-Means *Clustering* Untuk Menganalisa Bisnis Perusahaan Asuransi," *Jatisi (Jurnal Tek. Inform. Dan Sist. Informasi)*, Vol. 6, No. 1, Pp. 85-95, 2019, Doi: 10.35957/Jatisi.V6i1.150.