



Analisis Metode K-MEANS Pada Pengelompokan Keberadaan Area Resapan Air Menurut Provinsi

Dinda Nabila Batubara, Agus Perdana Windarto, Dedy Hartama, Heru Satria

Program Studi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Jalan Jenderal Sudirman Blok A. No.1/2/3, Indonesia

Email: ¹dindanabilabatubara@gmail.com, ²agus.perdana@amiktunasbangsa.ac.id

Abstrak

Daerah resapan air merupakan sebuah daerah tempat masuknya air dari permukaan tanah ke dalam zona jenuh air sehingga membentuk suatu aliran air di dalam tanah. Kurangnya daerah resapan air menjadi masalah saat ini, karena daerah resapan air berfungsi untuk menampung debit air hujan yang turun di daerah tersebut agar tidak terjadi banjir dan masalah kekeringan. Sumber data diambil dari situs Badan Pusat Statistik Indonesia. Penelitian ini bertujuan untuk mengelompokkan daerah resapan air menggunakan algoritma *datamining* dengan metode K-Means. Data yang digunakan dalam penelitian ini adalah data persentase rumah tangga berdasarkan provinsi dan keberadaan area resapan air pada tahun 2017 yang terdiri dari 34 provinsi di Indonesia. Berdasarkan data tersebut diperoleh pengelompokan area resapan air berdasarkan provinsi menjadi 2 *cluster* yaitu *cluster* tinggi (C1) dan *cluster* rendah menggunakan K-means *clustering*. Dari hasil penelitian diperoleh pengelompokan menjadi 16 provinsi dengan *cluster* tinggi (C1) dan 18 provinsi dengan *cluster* rendah (C2). Hal ini dapat menjadi masukan pada pemerintah untuk provinsi yang memiliki daerah resapan air rendah menjadi perhatian lebih berdasarkan *cluster* yang telah dilakukan.

Kata Kunci: K-Means, Area resapan air, Pengelompokan, *Clustering*, Data Mining

1. PENDAHULUAN

Daerah resapan air merupakan daerah masuknya air permukaan tanah ke dalam zona jenuh air sehingga membentuk suatu aliran air tanah yang mengalir ke daerah yang lebih rendah [1]. Jenis jenis area resapan air antara lain: 1) sumur resapan, 2) lubang resapan biopori, dan 3) tanaman atau tanah berumput. Fungsi dari daerah resapan air tersebut untuk menampung debit air hujan yang turun di daerah tersebut. Curah hujan yang cukup tinggi dan sedikitnya area resapan air dapat menyebabkan banjir. Secara tidak langsung daerah resapan air memiliki peran penting sebagai pengendali banjir dan kekeringan di musim kemarau, dapat membantu menjaga keseimbangan tata air dan menyelamatkan sumberdaya air untuk jangka panjang. Saat ini daerah resapan air di daerah perkotaan sudah semakin berkurang karena tingginya aktivitas pembangunan di berbagai daerah yang tidak memperhatikan aspek lingkungan. Dengan menggunakan data Badan Pusat Statistik Indonesia (<https://www.bps.go.id>), peneliti ingin mengelompokkan wilayah dengan keberadaan area resapan air dengan teknik ilmu komputer. Tujuan penelitian ini untuk menganalisa apakah teknik ilmu komputer dapat diterapkan dalam mengelompokkan wilayah dengan keberadaan area resapan air sehingga hasil dari penelitian tersebut dapat menjadi informasi bagi pihak yang berkepentingan dalam mengambil kebijakan terhadap daerah yang memiliki area resapan air yang rendah. Dalam mengatasi hal tersebut, teknik ilmu komputer dapat diterapkan untuk menjawab permasalahan diatas. Salah satunya adalah teknik *datamining* [2] dengan metode *clustering* [3].

Analisis klaster merupakan suatu teknik multivariat dengan tujuan utama mengelompokkan objek-objek berdasarkan karakteristik yang dimiliki. Sekarang ini analisis klaster telah banyak diaplikasikan dalam berbagai bidang yang ditulis dalam berbagai penelitian dan jurnal [4]. Pada *Clustering* terdapat beberapa algoritma pengelompokan untuk mengelompokkan data secara mudah [5]. Salah satu algoritma pengelompokan adalah K-Means. K-Means merupakan salah satu metode pengelompokan data nonhierarki (sekatan) yang berusaha mempartisi data ke dalam kelompok sehingga data yang memiliki karakteristik yang sama dimasukkan ke dalam satu kelompok yang sama dan data yang memiliki karakteristik berbeda dikelompokkan kedalam kelompok yang lain [6].

Algoritma K-Means merupakan algoritma pembelajaran yang sederhana yang dapat menyelesaikan suatu permasalahan untuk meminimalkan kesalahan ganda [7]. Hal ini dibuktikan dengan penelitian terdahulu yang dilakukan [8]. Hasil penelitian tersebut menyimpulkan dengan mengimplementasi Algoritma K-means mampu mengelompokkan provinsi berdasarkan tingkat sarana kesehatan yang dimiliki. Sehingga diperoleh penilaian berdasarkan indeks Desa/Kelurahan Yang Memiliki Sarana Kesehatan dengan 4 provinsi tingkat sarana kesehatan tinggi yaitu Sumatera Utara, Jawa Barat, Jawa Tengah, Jawa Timur, 14 provinsi tingkat sarana kesehatan sedang, dan 16 provinsi lainnya termasuk tingkat sarana kesehatan rendah.

2. METODOLOGI PENELITIAN

2.1 Tahap Pengumpulan Data

Sumber data penelitian diperoleh dari data yang dikumpulkan berdasarkan dokumen-dokumen yang dihasilkan oleh Badan Pusat Statistik Indonesia melalui situs <https://www.bps.go.id>. Data yang digunakan dalam penelitian ini adalah data persentase rumah tangga menurut provinsi dan area resapan air pada tahun 2017 yang terdiri dari 34 provinsi di Indonesia. Variabel yang digunakan : sumur resapan (x), lubang resapan biopori (y), dan sumur resapan (z).

2.2 Tahap Pengolahan Data

Data diolah menggunakan 2 *cluster* yaitu *cluster* area resapan tinggi (C1) dan *cluster* area resapan air rendah (C2) sehingga pada tahapan ini sudah diperoleh perhitungan nilai yang akan diproses pada tahap selanjutnya. Proses pengolahan menggunakan bantuan aplikasi *RapidMiner*.

2.3 Tahap Clustering

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[9].

3. ANALISA DAN PEMBAHASAN

Dalam melakukan *clustering*, data yang diperoleh akan dihitung terlebih dahulu berdasarkan jumlah area resapan air pada tahun 2017 berdasarkan provinsi di Indonesia yang ditunjukkan pada tabel 1.

Tabel 1. Persentase Rumah Tangga Menurut Provinsi dan Keberadaan Area Resapan Air, 2017

Provinsi	Jenis Area Resapan Air 2017		
	Sumur Resapan	Lubang Resapan Biopori	Taman/Tanah Berumput
Aceh	2.42	1.21	27.93
Sumatera Utara	4.01	1.47	18.67
Sumatera Barat	1.95	1.33	30.09
Riau	.34	0.49	39.56
Jambi	7.99	0.80	30.39
Sumatera Selatan	4.13	1.13	32.58
Bengkulu	1.34	0.00	30.39
Lampung	2.88	0.83	38.79
Kepulauan Bangka Belitung	3.05	1.89	34.13
Kepulauan Riau	1.56	0.43	21.71
DKI Jakarta	7.85	4.58	7.99
Jawa Barat	5.48	3.23	18.05
Jawa Tengah	3.98	2.12	26.81
DI Yogyakarta	21.38	2.53	20.40
Jawa Timur	8.10	2.61	19.43
Banten	3.61	1.10	18.93
Bali	4.22	1.64	33.59
Nusa Tenggara Barat	6.23	2.94	13.34
Nusa Tenggara Timur	1.92	0.54	23.42
Kalimantan Barat	4.47	2.15	52.24
Kalimantan Tengah	.47	0.08	33.93
Kalimantan Selatan	4.36	1.27	21.91
Kalimantan Timur	.58	0.38	27.49
Kalimantan Utara	0.18	0.00	27.04
Sulawesi Utara	1.72	1.65	30.73
Sulawesi Tengah	.66	0.04	35.08
Sulawesi Selatan	1.78	1.66	26.66
Sulawesi Tenggara	3.45	0.00	37.77
Gorontalo	1.12	0.09	43.31
Sulawesi Barat	2.34	0.78	28.39
Maluku	3.67	4.60	25.11
Maluku Utara	.07	0.00	24.04
Papua Barat	.24	0.00	36.00
Papua	2.97	2.13	37.50

3.1 Centroid Data

Dalam penerapan algoritma K-means dihasilkan nilai centroid dari data yang didapat dengan ketentuan klasterisasi yang diinginkan adalah 2. Penentuan *cluster* dibagi atas 2 *cluster* yaitu *cluster* tingkat populasi tinggi (C1), *cluster* tingkat populasi rendah (C2). Penentuan titik *cluster* dilakukan dengan mengambil nilai terbesar (maksimum) untuk *cluster* tingkat populasi tinggi (C1), dan nilai terkecil (minimum) untuk *cluster* tingkat populasi rendah (C2) yang ditunjukkan pada tabel 2.

Tabel 2. Nilai centroid awal

Cluster Tinggi	21.38	4.60	52.24
----------------	-------	------	-------

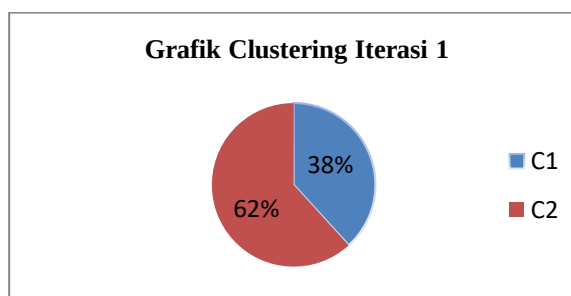
Cluster Rendah 0.07 0.00 7.99

3.2 Clustering Data

Dengan menggunakan centroid tersebut maka di *cluster* data yang telah didapat menjadi 2 *cluster*. Proses *cluster* dilakukan dengan mengambil jarak terdekat dari setiap data yang diolah. Hasil Pengelompokan ditunjukkan pada tabel 3.

Tabel 3. Penentuan jarak pusat *Cluster* iterasi 1

Provinsi	C1	C2	Jarak Terdekat	Cluster
Aceh	31.0100	20.1172	20.1172	C2
Sumatera Utara	37.9294	11.4742	11.4742	C2
Sumatera Barat	29.6501	22.2141	22.2141	C2
Riau	24.9080	31.5745	24.9080	C1
Jambi	25.9072	23.7699	23.7699	C2
Sumatera Selatan	26.3820	24.9481	24.9481	C2
Bengkulu	30.0013	22.4389	22.4389	C2
Lampung	23.1837	30.9359	23.1837	C1
Kepulauan Bangka Belitung	25.9066	26.3777	25.9066	C1
Kepulauan Riau	36.6335	13.8078	13.8078	C2
DKI Jakarta	46.2692	9.0305	9.0305	C2
Jawa Barat	37.7354	11.8654	11.8654	C2
Jawa Tengah	30.9106	19.3375	19.3375	C2
DI Yogyakarta	31.9041	24.7923	24.7923	C2
Jawa Timur	35.4445	14.2248	14.2248	C2
Banten	37.9188	11.5456	11.5456	C2
Bali	25.5129	25.9897	25.5129	C1
Nusa Tenggara Barat	41.7832	8.6679	8.6679	C2
Nusa Tenggara Timur	35.0083	15.5523	15.5523	C2
Kalimantan Barat	17.0854	44.5175	17.0854	C1
Kalimantan Tengah	28.1542	25.9470	25.9470	C2
Kalimantan Selatan	34.9378	14.6195	14.6195	C2
Kalimantan Timur	32.6026	19.5110	19.5110	C2
Kalimantan Utara	33.2539	19.0464	19.0464	C2
Sulawesi Utara	29.2872	22.8591	22.8591	C2
Sulawesi Tengah	27.2851	27.0908	27.0908	C2
Sulawesi Selatan	32.3634	18.8166	18.8166	C1
Sulawesi Tenggara	23.4949	29.9736	23.4949	C1
Gorontalo	22.5937	35.3338	22.5937	C1
Sulawesi Barat	30.7514	20.5426	20.5426	C1
Maluku	32.3983	18.0869	18.0869	C1
Maluku Utara	35.6449	16.0478	16.0478	C1
Papua Barat	27.0501	28.0082	27.0501	C1
Papua	23.7146	29.7249	23.7146	C1



Gambar 1. Grafik *Clustering* iterasi 1

Pada iterasi 1 diperoleh *cluster* data area resapan air berdasarkan provinsi yang dapat dilihat pada tabel 3. Proses *clustering* tersebut berhenti pada iterasi ke 4, pada iterasi 4 menentukan nilai titik tengah atau centroid dapat diketahui pada Tabel 4.

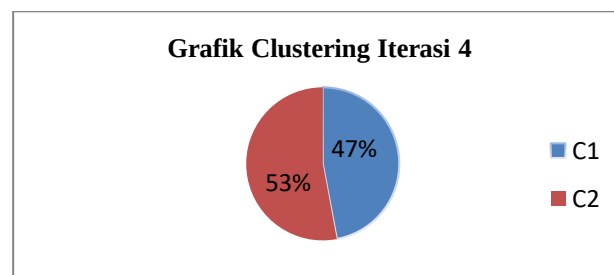
Tabel 4. Centorid pada iterasi ke 4

Cluster Tinggi	2.5629	0.8904	36.0049
Cluster Rendah	4.418	1.7479	22.0726

Setelah mendapatkan nilai centroid, proses yang sama dilakukan dengan mencari jarak terdekat. Hasil dari penentuan jarak terdekat dapat diketahui pada tabel 5.

Tabel 5. Penentuan jarak pusat *cluster* pada iterasi 4

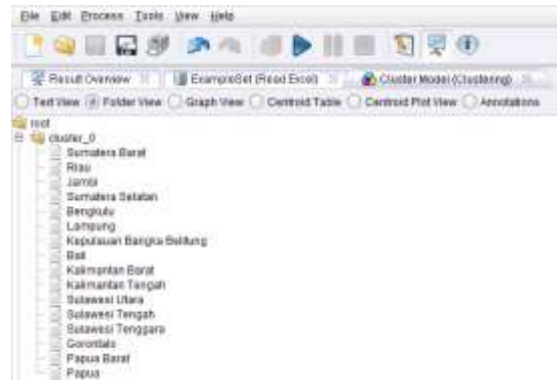
Provinsi	C1	C2	Jarak Terdekat	Kluster
Aceh	8.0796	6.2138	6.2138	C2
Sumatera Utara	17.4084	3.4417	3.4417	C2
Sumatera Barat	5.9669	8.3962	5.9669	C1
Riau	4.2132	18.0016	4.2132	C1
Jambi	7.8106	9.0996	7.8106	C1
Sumatera Selatan	3.7747	10.5297	3.7747	C1
Bengkulu	5.8119	9.0439	5.8119	C1
Lampung	2.8016	16.8113	2.8016	C1
Kepulauan Bangka Belitung	2.1813	12.1364	2.1813	C1
Kepulauan Riau	14.3364	3.1654	3.1654	C2
DKI Jakarta	28.7467	14.7683	14.7683	C2
Jawa Barat	18.3449	4.4212	4.4212	C2
Jawa Tengah	9.3861	4.7709	4.7709	C2
DI Yogyakarta	24.5021	17.0640	17.0640	C2
Jawa Timur	17.5578	4.6155	4.6155	C2
Banten	17.1130	3.3136	3.3136	C2
Bali	3.0182	11.5245	3.0182	C1
Nusa Tenggara Barat	23.0552	9.0020	9.0020	C2
Nusa Tenggara Timur	12.6027	3.0881	3.0881	C2
Kalimantan Barat	16.3934	30.1681	16.3934	C1
Kalimantan Tengah	3.0529	12.6125	3.0529	C1
Kalimantan Selatan	14.2162	0.5083	0.5083	C2
Kalimantan Timur	8.7564	6.7812	6.7812	C2
Kalimantan Utara	9.3222	6.7588	6.7588	C2
Sulawesi Utara	5.3949	9.0681	5.3949	C1
Sulawesi Tengah	2.2766	13.6403	2.2766	C1
Sulawesi Selatan	9.4133	5.2906	5.2906	C2
Sulawesi Tenggara	2.1685	15.8278	2.1685	C1
Gorontalo	7.4874	21.5542	7.4874	C1
Sulawesi Barat	7.6165	6.7218	6.7218	C2
Maluku	11.5619	4.2300	4.2300	C2
Maluku Utara	12.2557	5.0829	5.0829	C2
Papua Barat	2.4852	14.6432	2.4852	C1
Papua	1.9836	15.4970	1.9836	C1



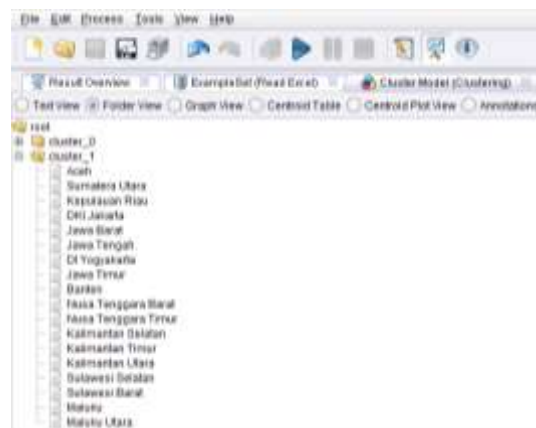
Gambar 2. Grafik *Clustering* iterasi 4



Gambar 3. Cluster model



Gambar 4. Folder View Cluster C1



Gambar 5. Folder View cluster C2

3.3 Analisis Data

Pada iterasi 4, pengelompokan data yang dilakukan terhadap 2 *cluster* dengan iterasi 3 didapatkan hasil yang sama. Dari 34 provinsi di Indonesia dapat diketahui, 16 provinsi *cluster* area resapan air tinggi yaitu Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Bali, Kalimantan Barat, Kalimantan Tengah, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Papua barat, papua, dan 18 provinsi lainnya dengan *cluster* rendah (C2).

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan maka dapat disimpulkan bahwa K-means *Clustering* pada pengelompokan keberadaan area resapan air berdasarkan provinsi dapat diterapkan. Data diolah menjadi 2 *cluster* yaitu *cluster* tinggi dan *cluster* rendah. Dimana hasil penelitian menyimpulkan dari 34 provinsi di Indonesia dan 2 *cluster* yang ada diperoleh 16 provinsi dengan *cluster* tinggi (C1) yaitu Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Bali, Kalimantan Barat, Kalimantan Tengah, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Papua barat, papua, dan 18 provinsi lainnya dengan *cluster* rendah (C2).

REFERENCES

- [1] S. C. Tendean, O. H. A. Rogi, and V. A. Kumurur, "EVALUASI KAWASAN RESAPAN AIR DI KOTA MANADO," 2017.
- [2] M. G. Sadewo, A. P. Windarto, and D. Hartama, "PENERAPAN DATAMINING PADA POPULASI DAGING AYAM RAS PEDAGING DI INDONESIA BERDASARKAN PROVINSI MENGGUNAKAN K-MEANS," pp. 60–67, 2016.
- [3] P. Alkhairi and A. P. Windarto, "Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara," pp. 762–767, 2019.
- [4] A. P. Windarto, "Penerapan Data Mining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan K-Means Clustering," *Techno.COM*, vol. 16, no. 4, pp. 348–357, 2017.
- [5] I. Parlina, A. P. Windarto, A. Wanto, and M. R. Lubis, "MEMANFAATKAN ALGORITMA K-MEANS DALAM MENENTUKAN PEGAWAI YANG LAYAK MENGIKUTI ASESSMENT CENTER," vol. 3, no. 1, pp. 87–93, 2018.
- [6] M. G. Sadewo, A. Eriza, A. P. Windarto, and D. Hartama, "Algoritma K-Means Dalam Mengelompokkan Desa / Kelurahan Menurut Keberadaan Keluarga Pengguna Listrik dan Sumber Penerangan Jalan Utama Berdasarkan Provinsi," pp. 754–761, 2019.
- [7] A. Nurzahputra, M. A. Muslim, and M. Khusniati, "Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa," vol. 16, no. 1, pp. 17–24, 2017.
- [8] M. G. Sadewo, A. P. Windarto, and S. R. Andani, "PEMANFAATAN ALGORITMA CLUSTERING DALAM MENGELOMPOKKAN JUMLAH DESA / KELURAHAN YANG MEMILIKI SARANA KESEHATAN," vol. I, pp. 124–131, 2017.
- [9] Anita, Y. Muhammi, and S. T. Siska, "Penentuan Tingkat Minat Belanja Online Melalui Media Sosial Menggunakan Metode Clustering K-Means," vol. I, no. 2, 2018.