

Penerapan Algoritma K-Means clustering Untuk Mengelompokkan Provinsi Berdasarkan Banyaknya Desa/Kelurahan Dengan Upaya Antisipasi/Mitigasi Bencana Alam

Yovi Pratama, Hendrawan, Errissya Rasywir*, Babel Tio Carenina, Dila Riski Anggraini

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dinamika Bangsa, Jambi, Indonesia

Email : ¹yovi.pratama@gmail.com, ²hendrawan.stikom@gmail.com, ^{3,*}errissya.jurnal@gmail.com,

⁴babeltiocarenina@gmail.com, ⁵dilariskianggraini@gmail.com

Email Penulis Korespondensi: ³errissya.jurnal@gmail.com

Submitted: 18/11/2022; Accepted: 17/12/2022; Published: 30/12/2022

Abstrak—Bencana Alam merupakan salah satu fenomena alam yang mengancam keberlangsungan hidup manusia. Dampak negatif yang ditimbulkan bisa berupa kerugian materi maupun non materi. Meskipun demikian, dengan kemampuan mengenali gejala-gejala awal dari sebuah bencana manusia dapat mempersiapkan diri dalam menghadapi bencana. Penerapan Algoritma K-Means clustering dalam mengelompokkan provinsi merupakan salah satu usaha berbasis matematis untuk usaha antisipasi/Mitigasi Bencana Alam. Sehingga, perlu dilakukan penggalian data cluster provinsi mana saja yang sudah baik dalam mitigasi bencana. Sumber data penelitian ini dikumpulkan berdasarkan dokumen-dokumen keterangan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/Mitigasi Bencana Alam yang dihasilkan oleh Badan Pusat Statistik Nasional. Data yang digunakan dalam penelitian ini adalah data provinsi yang terdiri dari 34 provinsi. Variable yang digunakan ada 4 yaitu Sistem Peringatan Dini Bencana Alam, Sistem Peringatan Dini Tsunami, Perlengkapan Keselamatan, Jalur Evakuasi. Data akan diolah dengan melakukan clustering dalam 2 cluster yaitu cluster tingkat antisipasi/mitigasi tinggi dan tingkat antisipasi/mitigasi rendah. Hasil yang diperoleh dari proses penilaian yaitu didapat 5 atau 14.71% provinsi dengan tingkat antisipasi/mitigasi tinggi dan 29 atau 85.29% provinsi lainnya termasuk tingkat antisipasi/mitigasi rendah. Hal ini dapat menjadi masukan kepada pemerintah untuk lebih memperhatikan Desa/Kelurahan tersebut berdasarkan cluster yang telah dilakukan.

Kata Kunci: Bencana Alam; K-Means Clustering; Weka; Antisipasi; Provinsi

Abstract—Natural disasters are one of the natural phenomena that threaten human survival. The negative impacts can be in the form of material or non-material losses. However, with the ability to recognize the early symptoms of a disaster, humans can prepare themselves for disaster. Application of the K-Means clustering Algorithm in Grouping Provinces Based on the Number of Villages / Sub-districts with Anticipation / Mitigation Efforts for Natural Disasters Using the WEKA Application. The data sources for this research were collected based on documents describing the Number of Villages/ Urban According to Natural Disaster Anticipation/Mitigation Efforts produced by the National Statistics Agency. The data used in this study is provincial data which consists of 34 provinces. There are 4 variables used, namely Natural Disaster Early Warning System, Tsunami Early Warning System, Safety Equipment, Evacuation Path. The data will be processed by clustering in 2 clusters, namely clusters with high anticipation/mitigation levels and low anticipation/mitigation levels. The results obtained from the assessment process are that there are 5 (14.71 %) provinces with a high level of anticipation/mitigation and 29 (85.29%) other provinces including a low level of anticipation/mitigation. This can be an input for the government to pay more attention to the Village/Kelurahan based on the clusters that have been carried out.

Keywords: Natural Disaster; K-Means Clustering; Weka; Anticipation; Province

1. PENDAHULUAN

Bencana alam adalah suatu kejadian atau kerusakan yang disebabkan oleh alam dan biasanya datangnya secara tiba-tiba atau mendadak, dan kemungkinan besar akan menimbulkan kerugian materi bahkan sampai jiwa manusia [1]. Bencana Alam merupakan salah satu fenomena alam yang mengancam keberlangsungan hidup manusia. Dampak negatif yang ditimbulkan bisa berupa kerugian materi maupun non materi. Bencana tersebut bisa dicontohkan seperti banjir, tanah longsor, gempa bumi ada pula bencana non alam seperti kebakaran gagal teknologi, gagal modernisasi, konflik sosial antar kelompok dan teror. Bencana merupakan sebuah fenomena kehidupan manusia yang tidak dapat diketahui secara pasti kapan terjadinya. Manusia hanya mampu mengenali gejala-gejala awal dan memprediksi terjadinya [2].

Kecanggihan teknologi yang diciptakan manusia terkadang hanya mampu menjelaskan gejala awal ini, sehingga kejadian detil dari bencana itu hanya dalam prediksi manusia. Meskipun demikian, dengan kemampuan mengenali gejala-gejala awal dari sebuah bencana manusia dapat mempersiapkan diri dalam menghadapi bencana. Persiapan itu meliputi persiapan sebelum terjadinya bencana, ketika terjadi bencana, dan pasca terjadinya bencana. Artinya, kesiapan yang dilakukan oleh manusia dapat dilakukan ketika dapat mengenali gejala awal, tingkat resikonya dan lain sebagainya. Penanggulangan bencana alam atau mitigasi adalah upaya berkelanjutan untuk mengurangi dampak bencana terhadap manusia dan harta benda. Persiapan menghadapi bencana alam termasuk semua aktivitas yang dilakukan sebelum terdeteksinya tanda-tanda bencana agar bisa memfasilitasi pemakaian sumber daya alam yang tersedia, meminta bantuan dan serta rencana rehabilitasi dalam cara dan kemungkinan yang paling baik. Kesiapan menghadapi bencana alam dimulai dari level komunikasi local. Jika sumber daya lokal kurang mencukupi, maka daerah tersebut dapat meminta bantuan ke tingkat nasional dan internasional.

Pada penelitian lain yang berjudul “Implementasi Data Mining Untuk Menentukan Persediaan Stok Barang Di Mini Market Menggunakan Metode *K-Means clustering*” oleh Hani Prastiri et al.[3] menerangkan tentang penerapan algoritma K-Means untuk merancang strategi persediaan barang yang efektif menggunakan data transaksi penjualan dari MM Glory agar mempermudah pekerjaan. Penelitian sejenis pernah dilakukan dengan judul “*Clustering Menggunakan Metode K-Means Untuk Menentukan Status Gizi Balita*” oleh Windha Mega P. D [4] menjelaskan tentang jumlah balita yang meninggal dunia akibat kekurangan gizi, maka diperlukan pemantauan untuk gizi balita dengan dilakukan pengelompokan kedalam 5 *cluster* status gizi melalui beberapa tahapan. Penelitian lain juga pernah dilakukan oleh Ai Rohman et al. [5] dengan judul “implementasi Algoritma *K-Means clustering* Analisis Untuk Menentukan Hambatan Pembelajaran Daring” menerangkan upaya menunjang proses kegiatan belajar mengajar daring dengan menggunakan *K-Means clustering* untuk menentukan tingkat hambatan pembelajaran daring dimana didapat 9 kelas *cluster* rendah dan 6 kelas *cluster* tinggi. Penelitian lain juga dilakukan dengan judul Klasterisasi Menggunakan Algoritma *K-Means clustering* Untuk Memprediksi Waktu Kelulusan Mahasiswa” oleh Hendro Priyatman et al.[6] menerangkan bahwa untuk mengetahui perkiraan waktu lulus mahasiswa dengan melihat mahasiswa-mahasiswa mana saja yang masuk kedalam suatu *cluster* tertentu berdasarkan parameter IPK dan kehadiran dengan algoritma K-Means model centroid. Penelitian sejenis pernah dilakukan dengan judul “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru” oleh Gustientiedina et al.[7] menjelaskan untuk menganalisa pemakaian obat-obatan, perencanaan dan pengendalian obat-obatan di rumah sakit diperlukan metode data clustering non hirarki yang mempartisi data ke dalam *cluster* sehingga data memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok lain. Clustering sendiri adalah proses pengelompokan satu set objek data (*into multiple groups*) atau *cluster* sehingga benda-benda dalam suatu kelompok memiliki kesamaan yang tinggi [8]–[11], tetapi sangat berbeda dengan objek di kelompok lain [12]. Data yang digunakan dalam penelitian ini berdasarkan dokumen- dokumen yang dihasilkan oleh Badan Pusat Statistik Nasional.

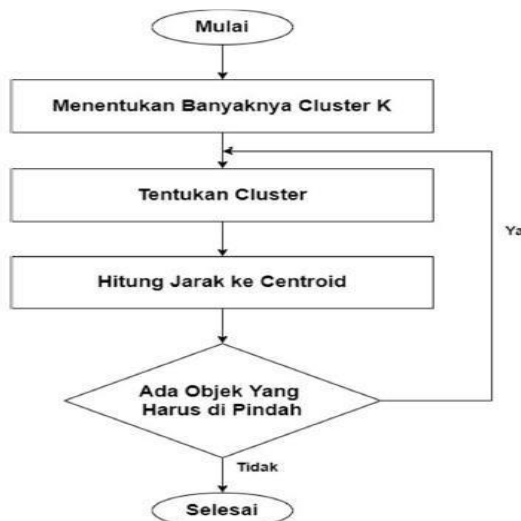
Berdasarkan uraian diatas peneliti melakukan penelitian dengan judul Penerapan Algoritma *K-Means clustering* Untuk Mengelompokkan Provinsi Berdasarkan Banyaknya Desa/Kelurahan dengan Upaya Antisipasi/Mitigasi Bencana Alam, dengan harapan hasil *cluster* dapat dijadikan masukan bagi pemerintah untuk lebih memperhatikan desa/kelurahan dengan *cluster* rendah. Selain itu dengan menerapkan clustering, dapat diketahui pendataan provinsi dengan Sistem Peringatan Dini Bencana Alam, Sistem Peringatan Dini Tsunami, Perlengkapan Keselamatan, jumlah Jalur Evakuasi serta *cluster* mitigasi bencana alam. Penerapan Algoritma *K-Means clustering* dalam mengelompokkan provinsi perlu dilakukan untuk penggalan data *cluster* provinsi mana saja yang sudah baik dalam mitigasi bencana.

2. METODOLOGI PENELITIAN

Pada metodologi penelitian ini terdapat beberapa tahapan untuk pengelompokan provinsi berdasarkan banyaknya desa/ kelurahan dengan upaya antisipasi atau mitigasi bencana alam dengan menggunakan metode *K-Means clustering*. Berdasarkan uraian diatas peneliti melakukan penelitian dengan judul Penerapan Algoritma *K-Means clustering* Untuk Mengelompokkan Provinsi Berdasarkan Banyaknya Desa/Kelurahan dengan Upaya Antisipasi/Mitigasi Bencana Alam, dengan harapan hasil *cluster* dapat dijadikan masukan bagi pemerintah untuk lebih memperhatikan desa/kelurahan dengan *cluster* rendah. Selain itu dengan menerapkan clustering, dapat diketahui pendataan provinsi dengan Sistem Peringatan Dini Bencana Alam, Sistem Peringatan Dini Tsunami, Perlengkapan Keselamatan, jumlah Jalur Evakuasi serta *cluster* mitigasi bencana alam. Penerapan Algoritma *K-Means clustering* dalam mengelompokkan provinsi perlu dilakukan untuk penggalan data *cluster* provinsi mana saja yang sudah baik dalam mitigasi bencana.

2.1 Penelitian Awal *Clustering* Data Desa/ Kelurahan Dengan Antisipasi/ Mitigasi Bencana Alam

Sumber data yang dipergunakan dalam penelitian ini merupakan data berdasarkan dokumen- dokumen yang dihasilkan oleh Badan Pusat Statistik Nasional. Data yang digunakan dalam penelitian ini adalah data Banyaknya Desa /Kelurahan Dengan Upaya Antisipasi/Mitigasi Bencana Alam di Setiap Provinsi, yang terdiri dari 34 provinsi. Dalam menentukan *cluster* berdasarkan data yang telah tersedia, dibutuhkan sebuah flowchart untuk memudahkan dalam menentukan alur perhitungan sebagai alur untuk menemukan hasil dari penerapan *cluster* terhadap data yang akan diproses. Berikut adalah flowchart dalam menentukan *cluster* dengan K-Means [13].



Gambar 1. Tahapan Algoritma K-Means clustering

Pada tahapan ini dilakukan analisis data banyaknya desa/kelurahan dengan upaya antisipasi / mitigasi bencana alam. Data yang diperoleh diolah dengan menggunakan perhitungan bobot dari tiap indeks. Pada tahapan ini ditentukan akan dicluster ke dalam 2 cluster yakni cluster tingkat antisipasi tinggi dan cluster tingkat antisipasi rendah. Pada tahapan inilah akan dianalisis hasilnya.

Tabel 1. Data Desa/ Kelurahan dengan Upaya Antisipasi/Mitigasi Bencana Alam

No	Provinsi	Sistem Peringatan Dini Bencana Alam	Sistem Peringatan Dini Tsunami	Perlengkapan Keselamatan	Jalur Evakuasi
1	Aceh	208	67	46	511
2	Sumatera Utara	185	27	59	198
3	Sumatera Barat	186	79	71	263
4	Riau	63	2	62	52
5	Jambi	34	0	22	36
6	Sumatera Selatan	92	1	25	31
7	Bengkulu	108	26	22	236
8	Lampung	171	26	12	120
9	Bangka Belitung	2	1	10	13
10	Kepulauan Riau	10	0	12	17
11	DKI Jakarta	64	0	166	123
12	Jawa Barat	761	44	184	560
13	Jawa Tengah	128	40	252	743
14	DI Yogyakarta	191	17	72	148
15	Jawa Timur	904	39	158	624
16	Banten	70	12	24	71
17	Bali	511	16	8	38
18	Nusa Tenggara Barat	56	12	16	63
19	Nusa Tenggara Timur	174	9	17	63
20	Kalimantan Barat	19	2	24	49
21	Kalimantan Tengah	26	0	21	69
22	Kalimantan Selatan	26	0	28	51
23	Kalimantan Timur	42	2	49	68
24	Kalimantan Utara	13	0	9	9
25	Sulawesi Utara	266	49	38	242
26	Sulawesi Tengah	100	7	11	54
27	Sulawesi Selatan	100	7	35	61
28	Sulawesi Tenggara	29	2	8	44
29	Gorontalo	31	2	8	14
30	Sulawesi Barat	14	6	5	15
31	Maluku	45	22	9	78
32	Maluku Utara	63	18	35	133
33	Papua Barat	49	16	14	66
34	Papua	42	8	16	48

2.2 Analisis K-Means clustering

Analisis *cluster* ialah algoritma yang dipakai untuk membagi rangkaian data menjadi beberapa grup berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya [14]

K-Means clustering merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan cluster/kelompok yang lain sehingga data yang berada dalam satu cluster/kelompok memiliki tingkat variasi yang kecil [15]

Analisis Pengelompokan / Clustering merupakan proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan datanya dalam suatu kelompok lebih besar daripada kesamaan data tersebut dengan data dalam kelompok lain. Potensi clustering adalah dapat digunakan untuk mengetahui struktur dalam data yang dapat dipakai lebih lanjut dalam berbagai aplikasi secara luas seperti klasifikasi, pengolahan gambar, dan pengenalan pola [16].

Algoritma clustering diklasifikasikan ke dalam 5 kategori [17] yaitu:

- a. Partitioning methods
Merupakan metode dengan pengelompokkan objek dimana tiap objek dimiliki oleh 1 cluster. Algoritma partisi dapat meminimalkan mean kuadrat jarak dari setiap titik data ke pusat terdekatnya. Yang termasuk ke dalam metode ini adalah algoritma *K-Means*, *K-Medoid* atau *PAM*, *CLARA*, dan *CLARANS*.
- b. Hierarchical methods
Merupakan proses clustering dimana pengelompokkan objek dapat dilakukan dengan 2 cara yaitu *agglomerative* dan *disvisive*. *Agglomerative* dimulai dengan menggabungkan beberapa *cluster* hingga menjadi satu, atau *divisive* yang dimulai dengan *cluster* yang sama kemudian dipecah menjadi beberapa *cluster* yang lebih kecil. Yang termasuk ke dalam metode ini adalah algoritma *CURE*, *BIRCH*, dan *Chameleon*.
- c. Density-based methods
Merupakan pengelompokkan objek berdasarkan tingkat kerapatan objek atau densitas. Yang termasuk dalam metode ini adalah algoritma *DBSCAN*, *DENCLUE*, dan *OPTICS*.
- d. Grid-based methods
Adalah pengelompokkan objek dengan memodelkan tiap cluster, dan mencoba mengoptimasikan kesesuaian data dengan model matematika. Yang termasuk ke dalam metode ini adalah algoritma *COBWEB*.

2.3 Penerapan Algoritma K-Means clustering Desa/ Kelurahan Dengan Upaya Antisipasi/ Mitigasi Bencana Alam

K-means merupakan algoritma yang menetapkan nilai-nilai *cluster* (*k*) secara random, untuk sementara nilai tersebut menjadi pusat dari *cluster* yang biasa disebut centroid. Kemudian menghitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid hingga nilai centroid tidak berubah (stabil). Sehingga dalam penggunaan algoritma k-means terdapat aturan sebagai berikut :

- a. Berapa jumlah *cluster* yang perlu dimasukkan.
- b. Hanya memiliki atribut bertipe numeric.

Pada dasarnya algoritma k-means hanya mengambil sebagian dari banyaknya komponen yang didapatkan untuk kemudian dijadikan pusat *cluster* awal, pada penentuan pusat *cluster* ini dipilih secara acak dari populasi data. Kemudian algoritma k-means akan menguji masing – masing dari setiap komponen dalam populasi data tersebut dan menandai komponen tersebut ke dalam salah satu pusat *cluster* yang telah didefinisikan sebelumnya tergantung dari jarak minimum antar komponen dengan tiap – tiap pusat cluster. Selanjutnya posisi pusat cluster akan dihitung kembali sampai semua komponen data digolongkan ke dalam tiap – tiap *cluster* dan terakhir akan terbentuk *cluster* baru [18].

2.4 Analisis evaluasi penerapan K-Means clustering Desa/Kelurahan Dengan Upaya Antisipasi/ Mitigasi Bencana Alam

Dalam penerapan algoritma K-means dihasilkan nilai titik tengah atau centroid dari data yang didapat dengan ketentuan bahwa clusterisasi yang diinginkan adalah 2, Penentuan *cluster* dibagi atas dua bagian yakni *cluster* tingkat antisipasi tinggi (C1) dan *cluster* tingkat antisipasi rendah (C2). Maka nilai titik tengah atau centroid juga terdapat 2 titik. Penentuan titik *cluster* ini dilakukan dengan mengambil nilai secara random. Nilai titik tersebut dapat diketahui pada Tabel 2 berikut:

Tabel 2. Centroid Data Awal

	A	B	C	D
C1	63	2	62	52
C2	42	8	16	48

Keterangan :

A = Sistem Peringatan Dini Bencana Alam

B = Sistem Peringatan Dini Tsunami

C = Perlengkapan Keselamatan



D = Jalur Evakuasi

2.5 Clustering data

Pada penelitian lain [19] , Dengan menggunakan centroid tersebut maka dapat dicluster data yang telah didapat menjadi 2 cluster. Proses cluster dengan mengambil jarak terdekat dari setiap data yang diolah dengan menggunakan rumus berikut:

$$d(x, y) = \sqrt{(x1 - y1)^2 + (x2 - y2)^2 + \dots + (xm - ym)^2} \tag{1}$$

Keterangan :

d(x,y) adalah Euclidean Distance

(x) merupakan koordinat object dan

(y) merupakan koordinat centroid.

Proses pencarian jarak terdekat, pengelompokan data pada iterasi 2 dan Clustering atas dapat digambarkan pada tabel dan gambar berikut:

Tabel 3. Perhitungan Jarak Pusat Cluster Iterasi 2

Data	Provinsi	C1	C2	Jarak Terpendek
1	Aceh	454.5612	472.2147	454.5612
2	Sumatera Utara	599.3254	158.5806	158.5806
3	Sumatera Barat	225.2492	562.1624	225.2492
4	Riau	784.1766	49.5102	49.5102
5	Jambi	821.6077	69.6121	69.6121
6	Sumatera Selatan	780.3641	46.3222	46.3222
7	Bengkulu	650.1569	161.7978	161.7978
8	Lampung	665.2038	95.3814	95.3814
9	Bangka Belitung	862.1835	109.9598	109.9598
10	Kepulauan Riau	853.3050	101.0583	101.0583
11	DKI Jakarta	737.4170	147.4705	147.4705
12	Jawa Barat	103.2597	842.7150	103.2597
13	Jawa Tengah	659.5841	1388.9811	659.5841
14	DI Yogyakarta	623.5209	131.7683	131.7683
15	Jawa Timur	250.2442	990.4429	250.2442
16	Banten	771.3026	20.8728	20.8728
17	Bali	544.6747	423.7673	423.7673
18	Nusa Tenggara Barat	788.2942	38.3048	38.3048
19	Nusa Tenggara Timur	700.4463	86.3335	86.3335
20	Kalimantan Barat	825.0406	76.1613	76.1613
21	Kalimantan Tengah	808.2819	65.1782	65.1782
22	Kalimantan Selatan	817.8991	69.0693	69.0693
23	Kalimantan Timur	792.2892	52.8657	52.8657
24	Kalimantan Utara	856.3928	104.2409	104.2409
25	Sulawesi Utara	512.2247	245.4343	245.4343
26	Sulawesi Tengah	761.4347	30.5137	30.5137
27	Sulawesi Selatan	753.1954	19.6929	19.6929
28	Sulawesi Tenggara	822.6563	72.1030	72.1030
29	Gorontalo	852.1917	100.1725	100.1725
30	Sulawesi Barat	839.5705	88.2791	88.2791
31	Maluku	788.6748	50.2345	50.2345
32	Maluku Utara	738.9613	63.4846	63.4846
33	Papua Barat	792.0718	44.7516	44.7516
34	Papua	808.4954	56.7710	56.7710

Tabel 3 di atas merupakan hasil perhitungan cluster dari data desa/kelurahan menurut upaya antisipasi/mitigasi bencana alam menurut provinsi didapatkan pengelompokan pada iterasi 2 untuk 2 cluster tersebut. Cluster tingkat antisipasi tinggi (C1) yakni Jawa Barat, Jawa Tengah, Jawa Timur, Sumatera Barat, Aceh dan tingkat antisipasi rendah (C2) yakni 29 provinsi lainnya .

Tabel 4. Hasil Pengelompokan Data ke Jenis Cluster

DATA	C1	C2
1	1	0
2	0	1
3	1	0
4	0	1

5	0	1
6	0	1
7	0	1
8	0	1
9	0	1
10	0	1
11	0	1
12	1	0
13	1	0
14	0	1
15	1	0
16	0	1
17	0	1
18	0	1
19	0	1
20	0	1
21	0	1
22	0	1
23	0	1
24	0	1
25	0	1
26	0	1
27	0	1
28	0	1
29	0	1
30	0	1
31	0	1
32	0	1
33	0	1
34	0	1

Pada tabel 3 ditampilkan Perhitungan Jarak Pusat *Cluster* ke- 2 dan tabel 4 ditampilkan Hasil Pengelompokan Data ke Jenis Cluster. Perhitungan yang dilakukan pada tabel ke 3 telah mengalami nilai yang konvergen pada iterasi ke-2. Oleh karena itu, nilai yang dihitung sebagai jarak terpendek dari setiap data ke pusat *cluster* digunakan untuk menjadi kategori *cluster* yang diambil. Kelompok *cluster* 0 adalah provinsi-provinsi yang mempunyai nilai antisipasi yang rendah terhadap bencana alam, dan pada kelompok kedua atau *cluster* 1 adalah kumpulan provinsi yang mempunyai antisipasi yang cukup tinggi terhadap bencana alam. Dari hasil *cluster* ditemukan bahwa terdapat provinsi yang antisipasi atau mitigasi terhadap bencana alam sebesar 85.29%.

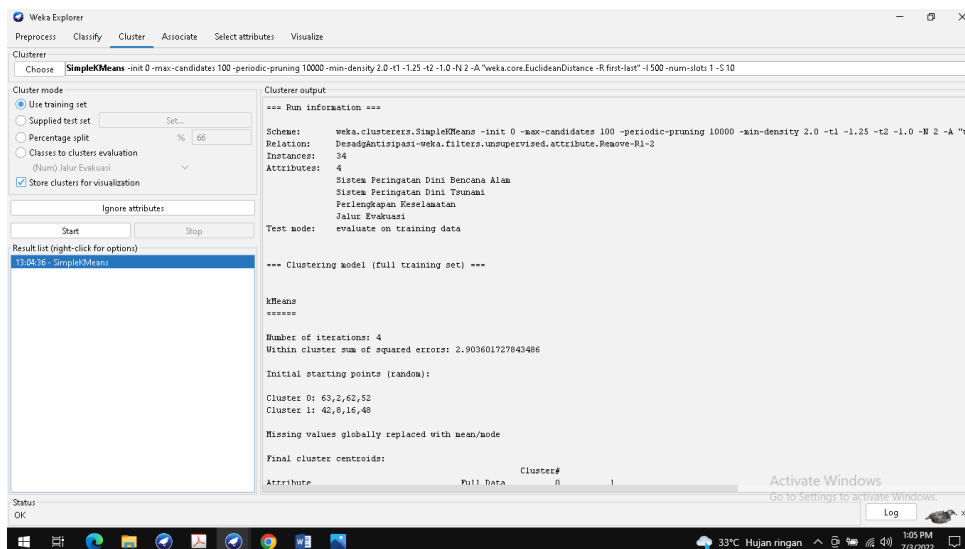
3. HASIL DAN PEMBAHASAN

3.1 Perhitungan Metode *K-Means clustering* Menggunakan aplikasi Weka

WEKA (*Waikato Environment for Knowledge Analysis*) [6] merupakan perangkat lunak data mining yang dikembangkan oleh Universitas Waikato, New Zealand. Diimplementasikan pertama kali pada tahun 1997 dan mulai menjadi open source pada tahun 1999. Hingga saat ini Weka sudah mencapai versi 3.6.11 dengan berbagai pengembangan dari versi pertama 3.3. Ditulis dalam bahasa pemrograman Java, Weka juga didukung oleh GUI yang sangat baik dan *user friendly*, dapat mengolah berbagai file data seperti *.csv dan *.arff serta memiliki fitur utama seperti data *preprocessing tools*, *learning algorithms* dan berbagai metode evaluasi [20]. Weka terdiri dari beberapa tools yang dapat digunakan untuk melakukan tugas *pre-processing data*, *classification*, *regression*, *klustering*, *association rules*, dan visualisasi [21]

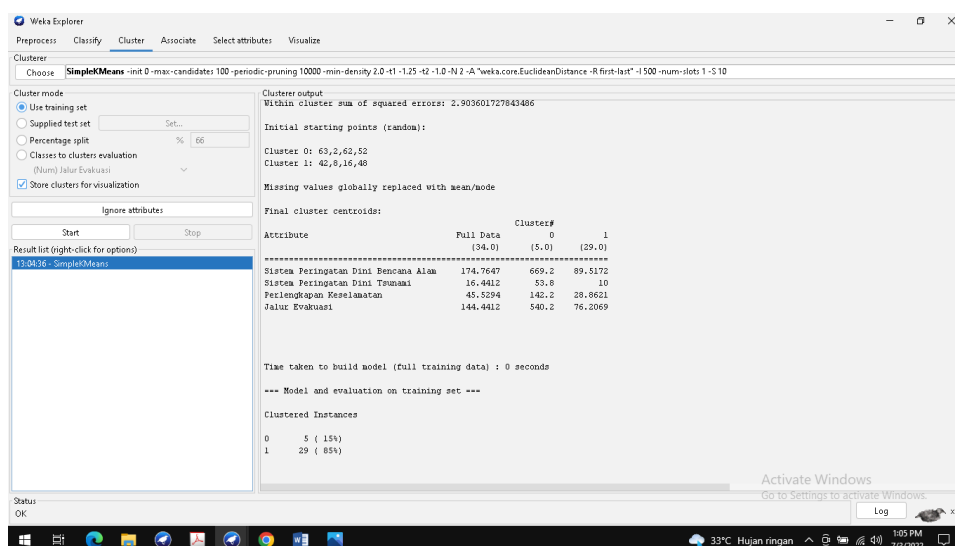
Data pengujian yang digunakan adalah berupa tabel yang memiliki komponen penyusun sebagai berikut:

- Memiliki 5 *attribute* yaitu Provinsi, Sistem Peringatan Dini Bencana Alam, Sistem Peringatan Dini Tsunami, Perlengkapan keselamatan dan jalur evakuasi.
- Jumlah *instance* adalah 34

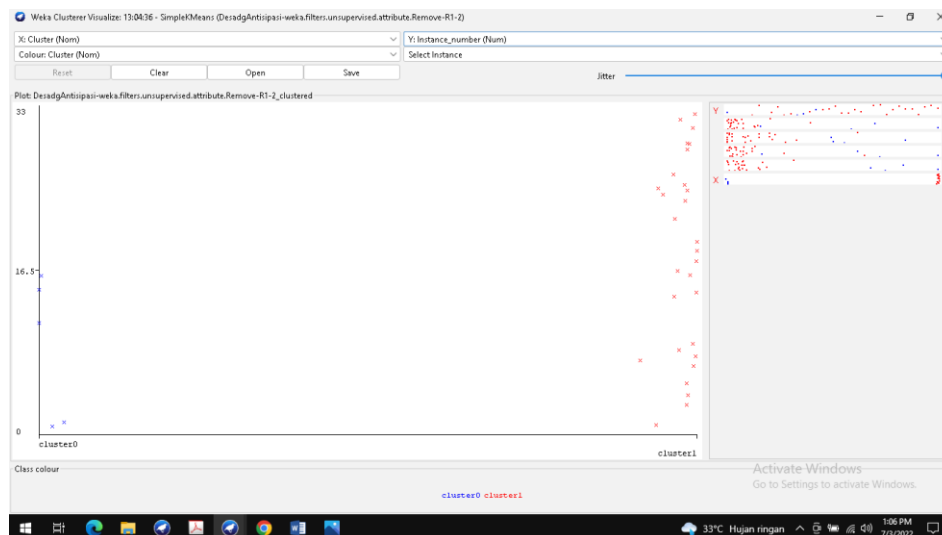


Gambar 2. Hasil Perhitungan Jarak Data dengan Centroid

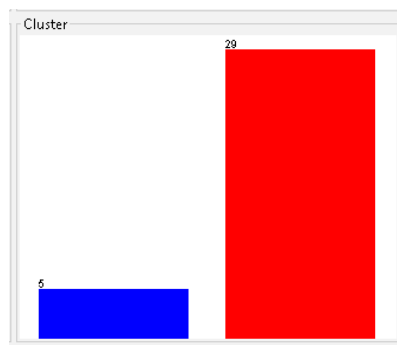
Hasil perhitungan jarak data dengan setiap *centroid* dilakukan secara otomatis menggunakan WEKA. Data yang digunakan Memiliki 5 *attribute* yaitu Provinsi, Sistem Peringatan Dini Bencana Alam, Sistem Peringatan Dini Tsunami, Perlengkapan keselamatan dan jalur evakuasi dengan jumlah *instance* adalah 34.



Gambar 3. Hasil cluster centroids dan clustered Instances dengan Weka



Gambar 4. Hasil Grafik clustering dengan Weka



Gambar 5. Grafik Hasil Cluster

Grafik yang ditampilkan dalam gambar 5 adalah 2 cluster yakni 0 (berwarna biru) untuk daerah kurang tanggap mitigasi bencana alam atau antisipasinya rendah sebesar 14.71%. Untuk cluster 1 yang berwarna merah adalah daerah yang lebih baik dalam mitigasi bencana alam dengan persentasi sebesar 85.29%. Artinya secara keseluruhan terdapat jumlah daerah provinsi yang lebih baik dalam melakukan mitigasi bencana alam.

Tabel 5. Hasil Cluster Dari Aplikasi Weka

No	Provinsi	Sistem Peringatan Dini Bencana Alam	Sistem Peringatan Dini Tsunami	Perlengkapan Keselamatan	Jalur Evakuasi	Cluster
1	Aceh	208	67	46	511	Cluster 0
2	Sumatera Barat	186	79	71	263	Cluster 0
3	Jawa Barat	761	44	184	560	Cluster 0
4	Jawa Tengah	1,28	40	252	743	Cluster 0
5	Jawa Timur	904	39	158	624	Cluster 0
6	Sumatera Utara	185	27	59	198	Cluster 1
7	Riau	63	2	62	52	Cluster 1
8	Jambi	34	0	22	36	Cluster 1
9	Sumatera Selatan	92	1	25	31	Cluster 1
10	Bengkulu	108	26	22	236	Cluster 1
11	Lampung	171	26	12	120	Cluster 1
12	Bangka Belitung	2	1	10	13	Cluster 1
13	Kepulauan Riau	10	0	12	17	Cluster 1
14	DKI Jakarta	64	0	166	123	Cluster 1
15	DI Yogyakarta	191	17	72	148	Cluster 1
16	Banten	70	12	24	71	Cluster 1
17	Bali	511	16	8	38	Cluster 1
18	NTB	56	12	16	63	Cluster 1
19	NTT	174	9	17	63	Cluster 1
20	Kalimantan Barat	19	2	24	49	Cluster 1
21	Kalimantan Tengah	26	0	21	69	Cluster 1
22	Kalimantan Selatan	26	0	28	51	Cluster 1
23	Kalimantan Timur	42	2	49	68	Cluster 1
24	Kalimantan Utara	13	0	9	9	Cluster 1
25	Sulawesi Utara	266	49	38	242	Cluster 1
26	Sulawesi Tengah	100	7	11	54	Cluster 1
27	Sulawesi Selatan	100	7	35	61	Cluster 1
28	Sulawesi Tenggara	29	2	8	44	Cluster 1
29	Gorontalo	31	2	8	14	Cluster 1
30	Sulawesi Barat	14	6	5	15	Cluster 1
31	Maluku	45	22	9	78	Cluster 1
32	Maluku Utara	63	18	35	133	Cluster 1
33	Papua Barat	49	16	14	66	Cluster 1
34	Papua	42	8	16	48	Cluster 1

Tabel 5 mengenai hasil cluster dari aplikasi Weka berisi data Provinsi, Sistem Peringatan Dini, Bencana Alam, Sistem Peringatan Dini Tsunami, Perlengkapan Keselamatan, Jalur Evakuasi dan Cluster. Data cluster yang dikelompokkan antara lain adalah cluster 0 dan cluster 1. Hasil dari perhitungan maka diperoleh bahwa Cluster 0 dengan tingkat antisipasi tinggi, sebanyak 5 Provinsi dari 34 Provinsi (15%) dan Cluster 1 dengan tingkat antisipasi rendah, sebanyak 29 Provinsi dari 34 Provinsi (85%).

4. KESIMPULAN

Berdasarkan penelitian yang dilakukan, dapat disimpulkan bahwa algoritma K-Means bisa digunakan untuk mengelompokkan Provinsi berdasarkan jumlah desa/kelurahan dengan upaya antisipasi/mitigasi. Dari data yang di atas, didapatkan 2 kelompok dimana, pada kelompok *Cluster* 0 dengan tingkat antisipasi tinggi, sebanyak 5 Provinsi dari 34 Provinsi (14.71%). *Cluster* 1 dengan tingkat antisipasi rendah, sebanyak 29 Provinsi dari 34 Provinsi (85.29%). Hasil yang didapat dari penelitian dapat menjadi masukan kepada pemerintah, provinsi yang menjadi perhatian lebih dalam peningkatan sarana dan prasana pada desa/kelurahan menurut upaya antisipasi/mitigasi bencana alam berdasarkan *cluster* yang telah dilakukan.

REFERENCES

- [1] B. Aeniah, "Meningkatkan Prestasi Belajar IPS Materi Mengenal Cara Menghadapi Bencana Alam Dengan Model Cooperative Tipe Circuit Learning Siswa Kelas VI Semester I SDN Batu Kembar Kecamatan Janapria Tahun Pelajaran 2015/2016," *JISIP (Jurnal Ilmu Sos. dan Pendidikan)*, vol. 4, no. 1, 2020, doi: 10.36312/jisip.v4i1.1043.
- [2] S. Maarif, R. Pramono, R. A. Kinseng, and E. Sunarti, "Kontestasi Pengetahuan dan Pemaknaan Tentang Ancaman Bencana Alam," *Penanggulangan Bencana*, vol. 3, no. 6, pp. 1–13, 2012.
- [3] H. Prastiwi, J. Pricilia, and E. Raswir, "Implementasi Data Mining Untuk Menentukan Persediaan Stok Barang Di Mini Market Menggunakan Metode *K-Means clustering* Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)," *J. Inform. Dan Rekayasa Komput.*, vol. 1, no. April, pp. 141–148, 2022.
- [4] W. Mega and P. Dhuhita, "CLUSTERING MENGGUNAKAN METODE K-MEANS UNTUK," *J. Informatia*, vol. 15, no. 2, 2015.
- [5] A. Rohmah, F. Sembiring, and A. Erfina, "IMPLEMENTASI ALGORITMA *K-MEANS CLUSTERING ANALYSIS* UNTUK MENENTUKAN HAMBATAN PEMBELAJARAN DARING (STUDI KASUS: SMK YASPIM GEGERBITUNG)," *SISTAMIK*, no. 8, pp. 290–298, 2021.
- [6] H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma *K-Means clustering* untuk Memprediksi Waktu Kelulusan," *JEPIN*, vol. 5, no. 1, pp. 62–66, 2019.
- [7] Gustientiedina, M. H. Adiya, and Y. Desnelita, "Jurnal Nasional Teknologi dan Sistem Informasi Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019.
- [8] H. Hendrawan, A. Haris, E. Rasywir, and Y. Pratama, "Diagnosis Penyakit Tanaman Karet dengan Metode Fuzzy Mamdani," *J. Paradig. UBSI*, vol. 22, no. 2, pp. 132–138, 2020.
- [9] Hendrawan, A. Haris, E. Rasywir, and Y. Pratama, "Sistem Pakar Diagnosis Penyakit Tanaman Karet dengan Metode Fuzzy Mamdani Berbasis Web," *J. Media Inform. Budidarma*, vol. 4, no. 4, pp. 1225–1234, 2020, doi: 10.30865/mib.v4i4.2521.
- [10] M. R. Borroek, E. Rasywir, Y. Pratama, Fachrudin, and M. Istoningtyas, "Analysis on Knowledge Layer Application for Knowledge Based System," in *Proceedings of 2018 International Conference on Electrical Engineering and Computer Science, ICECOS 2018*, 2019, pp. 177–182, doi: 10.1109/ICECOS.2018.8605262.
- [11] E. Gho, D. Z. Abidin, and E. Rasywir, "Analisis Dan Penerapan Data Mining Pada Transaksi Penjualan Obat Menggunakan Algoritma Apriori Di Apotek Persijam," *Tek. Inform. STIKOM Din. Bangsa*, pp. 56–64, 2013.
- [12] T. Connolly and B. Carolyn, *Database Systems A Practical Approach to Design, Implementation, and Management 6th Edition-Global Edition*. Harlow: Pearson Education Limited, 2015.
- [13] A. Wanto and A. P. Windarto, "Analisis Prediksi Indeks Harga Konsumen Berdasarkan Kelompok Kesehatan Dengan Menggunakan Metode Backpropagation," *J. Penelit. Tek. Inform. Sink.*, vol. 2, no. 2, pp. 37–43, 2017.
- [14] and B. Sumijan, A. P. Windarto, A. Muhammad, "Implementation of Neural Networks in Predicting the Understanding Level of Students Subject.," *Eng: Int. J. Softw.*, 2016.
- [15] Yudi Agusta, "K-Means – Penerapan, Permasalahan dan Metode Terkait," *J. Sist. dan Inform.*, vol. 3, no. Februari, pp. 47–60, 2007.
- [16] S. K. and S. K. Rathi, "Performance Evaluation of K-Means Algorithm and Enhanced Mid-point based K-Means Algorithm on Mining Frequent Patterns," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 4, no. 10, pp. 545–548, 2014.
- [17] E. Ian H, "Practical Machine Learning Tools and Techniques," *Morgan Kaufmann*, vol. 2, p. 4, 2005.
- [18] A. P. Windarto, L. S. Dewi, and D. Hartama, "Implementation of Artificial Intelligence in Predicting the Value of Indonesian Oil and Gas Exports With BP Algorithm," *Int. J. Recent Trends Eng. Res.*, vol. 3, no. 11, pp. 1–12, 2017, doi: 10.23883/ijrter.2017.3484.j5bbs.
- [19] M. G. Sadewo, A. P. Windarto, and A. Wanto, "Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/ Mitigasi Bencana Alam Menurut Provinsi Dengan K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 311–319, 2018, doi: 10.30865/komik.v2i1.943.
- [20] F. Eibe, "Machine Learning with WEKA," *Department of Computer Science*. New Zealand: University of Waikato, 2011.
- [21] S. Jiménez, "Text Classification and Klustering with WEKA: WEKA A guided example.," Colombia: Univercidad Nacional De.