



Pengelompokkan Data Bencana Alam Berdasarkan Wilayah, Waktu, Jumlah Korban dan Kerusakan Fasilitas Dengan Algoritma K-Means

Murdiaty, Angela, Chatrine Sylvia

Program Studi Sistem Informasi, STMIK-STIE Mikroskil, Medan, Indonesia

Email: ¹ murdiaty@mikroskil.ac.id, ² angela.woen@mikroskil.ac.id, ³ chatrine.sylvia@mikroskil.ac.id

Email Penulis Korespondensi: chatrine.sylvia@mikroskil.ac.id

Abstrak—Indonesia memiliki tanah yang subur, hasil bumi dan kekayaan laut yang berlimpah. Namun Indonesia juga tidak luput dari resiko bencana alam yang merupakan serangkaian peristiwa yang mengganggu dan mengancam keselamatan hidup serta menyebabkan kerugian materi maupun non-materi. Letak geologis Indonesia yang strategis menyebabkan Indonesia sering dilanda gempa bumi, letusan gunung berapi dan berbagai bencana alam lain. Dari data yang dikumpulkan, bencana alam yang terjadi di Indonesia terdiri dari beberapa kategori yaitu gempa bumi, letusan gunung berapi, banjir, longsor, puting beliung, serta tsunami. Banyaknya peristiwa bencana alam di Indonesia ini menyebabkan timbulnya korban, baik korban jiwa maupun luka-luka, menghancurkan wilayah sekitarnya serta menghancurkan infrastruktur dan menimbulkan kerugian harta benda. Adanya tren peningkatan kejadian bencana alam perlu diteliti lebih lanjut untuk mencegah agar jumlah korban tidak semakin banyak. Informasi tersebut dapat diperoleh melalui pendekatan data mining mengingat jumlah data yang tersedia sangat banyak. Dalam kaitannya dengan data bencana alam, teknik clustering pada data mining sangat berguna untuk mengelompokkan data bencana alam berdasarkan karakteristik yang sama sehingga dapat dijadikan dasar untuk memprediksi peristiwa bencana alam kedepannya. Dengan demikian, penelitian ini diharapkan dapat mengelompokkan data bencana alam memanfaatkan teknik clustering menggunakan algoritma k-means menjadi beberapa kelompok, dilihat dari segi jenis bencana alam, waktu terjadinya bencana, jumlah korban, dan kerusakan berbagai fasilitas sebagai dampak bencana alam.

Kata Kunci: Bencana Alam, K-Means, Clustering, Data Mining

Abstract—Indonesia has fertile soil, natural resources and abundant marine resources. However, Indonesia is also not immune to the risk of natural disasters which are a series of events that disturb and threaten life safety and cause material and non-material losses. Indonesia's strategic geological location causes Indonesia to be frequently hit by earthquakes, volcanic eruptions and other natural disasters. From the data collected, natural disasters that occurred in Indonesia consisted of several categories, namely earthquakes, volcanic eruptions, floods, landslides, tornados, and tsunamis. Many natural disasters in Indonesia have caused casualties, both fatalities and injuries, destroying the surrounding area and destroying infrastructure and causing property losses. The trend of increasing incidence of natural disasters needs to be further investigated to prevent the number of victims from increasing. This information can be obtained through a data mining approach given the large amount of data available. In relation to natural disaster data, clustering techniques in data mining are very useful for grouping natural disaster data based on the same characteristics so that the data can be adopted as a groundwork for predicting natural disaster events in the future. Thus, this research is supposed to group natural disaster data using clustering techniques using the k-means algorithm into several groups, in terms of natural disaster types, time of disaster, number of victims, and damage to various facilities as a result of natural disasters.

Keywords: Data Mining, Clustering, K-Means, Natural Disasters

1. PENDAHULUAN

Posisi Indonesia sangat strategis secara geografis di mana Indonesia diapit oleh 2 benua yaitu Benua Asia dan Benua Australia. Tidak hanya itu saja, Indonesia juga diapit oleh 2 samudra yakni Samudra Pasifik dan Samudra Hindia. Indonesia terdiri atas berbagai pulau besar dan pulau kecil yang berjumlah sebanyak 16.056 pulau. Indonesia merupakan negara yang memiliki banyak pegunungan dan dikelilingi lautan. Dengan demikian, tanah Indonesia menjadi sangat subur serta berlimpah dengan hasil bumi dan kekayaan laut. Walaupun memiliki kekayaan alam yang berlimpah, namun Indonesia juga tidak luput dari resiko bencana alam. Bencana alam merupakan serangkaian peristiwa yang mengganggu dan mengancam keselamatan hidup setiap makhluk hidup, yang terjadi secara alamiah karena disebabkan oleh faktor alam sehingga menyebabkan kerugian materi maupun non-materi [1].

Salah satu penyebab terjadinya bencana alam ini adalah letak geologis Indonesia yang strategis. Secara geologis, Indonesia terletak di antara Sirkum Mediterania dan Sirkum Pasifik, dan juga berada di pertemuan tiga lempeng yaitu lempeng Eurasia, lempeng Pasifik, dan lempeng Australia. Posisi ini menyebabkan Indonesia sering dilanda bencana alam seperti gempa bumi, gunung meletus serta berbagai bencana lain. Indonesia dikelilingi banyak gunung dimana sebagian besar merupakan gunung api aktif. Gunung api aktif di Indonesia diperkirakan sebanyak 76 gunung, sementara gunung api berstatus di atas normal sebanyak 22 gunung dengan 4 gunung di antaranya berstatus siaga [2].

Selain itu, Indonesia merupakan bagian Cincin Api Pasifik, yang menyebabkan wilayah Indonesia sering mengalami berbagai aktivitas tektonik. Dengan demikian, sebagian besar wilayah di Indonesia berpotensi tinggi mengalami berbagai peristiwa bencana alam. Beberapa bencana alam bahkan berpotensi menyebabkan bencana alam lainnya seperti bencana gempa bumi dan gunung meletus yang berpotensi menyebabkan tsunami. Berdasarkan data NOAA terdapat 204 peristiwa tsunami yang melanda wilayah Indonesia [3]. Dari data peristiwa tsunami yang dikumpulkan oleh NCEI (National Centers for Environmental Information), suatu badan yang



menyediakan akses data mengenai data kelautan, atmosfer, dan geofisika, diketahui bahwa peristiwa tsunami terparah yang dialami Indonesia yaitu terjadi pada provinsi Nanggroe Aceh Darussalam yang menyebabkan 167.540 korban jiwa [4].

Dari data yang dikumpulkan, bencana alam yang terjadi di Indonesia terdiri dari beberapa kategori yaitu banjir, tanah longsor, gempa bumi, letusan gunung berapi, puting beliung, dan tsunami yang terjadi di berbagai provinsi di Indonesia. Banyaknya peristiwa bencana alam di Indonesia ini menimbulkan korban, baik korban jiwa maupun korban luka-luka, menghancurkan wilayah sekitarnya serta menghancurkan infrastruktur dan menimbulkan kerugian harta benda. Dari data yang dihimpun oleh Badan Nasional Penanggulangan Bencana (BNPB), diketahui bahwa sepanjang tahun 2017 telah terjadi 2.866 peristiwa bencana alam yang menyebabkan sebanyak 378 orang meninggal dunia. Sementara pada tahun 2018, tercatat sebanyak 3.397 peristiwa bencana alam yang menyebabkan sebanyak 4.719 orang meninggal dunia [5]. Jumlah tersebut mengalami peningkatan sebesar 0.19% dari jumlah peristiwa bencana alam di tahun lalu serta mengalami peningkatan sebesar 11.48% dari jumlah korban jiwa di tahun lalu. Adanya tren peningkatan ini perlu diteliti lebih lanjut untuk mencegah agar jumlah korban tidak semakin banyak.

Peristiwa bencana alam bisa terjadi di mana pun dan kapan pun sehingga dibutuhkan suatu informasi yang dapat memperkirakan tanda-tanda terjadinya bencana agar masyarakat di daerah yang akan terkena bencana dapat mempersiapkan segala sesuatu untuk menghindari bencana tersebut. Untuk itu, dibutuhkan informasi yang dapat digunakan untuk menyusun rencana mitigasi dan meminimalkan dampak dari bencana tersebut. Informasi yang dibutuhkan tersebut dapat diperoleh melalui pendekatan data mining mengingat jumlah data yang tersedia sangat banyak.

Pendekatan data mining ini bukan merupakan hal baru, dibuktikan dengan banyaknya penelitian terdahulu yang memanfaatkan teknik data mining. Penelitian terdahulu yang diangkat oleh berbagai peneliti di antaranya, penelitian yang dilaksanakan oleh Sadewo, Windarto, dan Wanto (2018) yang membahas mengenai penggunaan algoritma k-means pada data desa/kelurahan di Indonesia berdasarkan upaya antisipasi/mitigasi bencana alam dimana data yang diolah meliputi data Sistem Peringatan Dini Bencana Alam, Sistem Peringatan Dini Tsunami, Perlengkapan Keselamatan, dan Jalur Evakuasi [6]. Penelitian lainnya dilakukan oleh Supriyadi et al. (2018) yang meneliti mengenai pengklasifikasian daerah rawan bencana menggunakan metode clustering k-means dengan bantuan aplikasi RapidMiner [7]. Penelitian selanjutnya, yaitu penelitian oleh Yana et al. (2018) yang bertujuan mengelompokkan wilayah Indonesia menurut frekuensi kejadian bencana alam tiap provinsi di Indonesia dengan mengambil data dari tahun 2013 sampai tahun 2018 [8].

Kemudian, penelitian yang dilakukan oleh Prihandoko, Bertalya, dan Ramadhan (2017) yang menerapkan algoritma k-means dengan algoritma k-medoids untuk dibandingkan pada gabungan data banjir dan data cuaca hujan di Provinsi Jawa Barat, Jawa Tengah, dan Jawa Timur [9]. Penelitian berikutnya membandingkan algoritma k-means dan isodata untuk mengelompokkan titik api yang tersebar di wilayah Sumatera, dilakukan oleh Sirait, Setiawan, dan Ramdani pada tahun 2018 [10]. Selanjutnya penelitian milik Ramadhan dan Prihandoko (2017) yang mengelompokkan data bencana setiap provinsi dengan menerapkan algoritma k-means kemudian memprediksi data bencana yang akan terjadi ke depannya [11]. Kemudian penelitian oleh Artatia dan Hakim (2015) yang mengelompokkan data gempa bumi dari berbagai provinsi menggunakan algoritma k-means [12]. Penelitian lainnya dibahas oleh Sugianto dan Astita (2018) yang membahas mengenai penerapan teknik data mining menggunakan algoritma Fp-growth untuk mencari pola hubungan aturan asosiatif dalam data bencana tanah longsor [13].

Data mining adalah proses untuk menemukan pola atau informasi yang berharga dan terpendam dari sejumlah besar data dengan menerapkan teknik ataupun metode tertentu [14], [15]. Salah satu teknik yang termasuk dalam data mining adalah teknik clustering. Data yang telah dikelompokkan berdasarkan teknik clustering dapat mengekstraksi informasi yang tersembunyi secara lebih baik untuk kebutuhan analisis dan prediksi data daripada data yang belum dikelompokkan [16]. Dengan demikian, dalam kaitannya dengan data bencana alam, teknik clustering pada data mining sangat berguna untuk mengelompokkan data bencana alam berdasarkan karakteristik yang sama. Kluster-kluster data bencana alam tersebut juga dapat digunakan sebagai data dasar untuk memprediksi peristiwa bencana alam (pemantauan bencana alam). Penelitian ini berencana mengelompokkan data bencana alam memanfaatkan teknik clustering menggunakan algoritma k-means menjadi beberapa kelompok, dilihat dari segi jenis bencana alam, waktu terjadinya bencana, jumlah korban, dan kerusakan berbagai fasilitas sebagai dampak bencana alam. Pengelompokkan data ini dapat digunakan untuk mengidentifikasi daerah rawan bencana, menentukan jenis bencana yang sering melanda suatu wilayah, mengidentifikasi kejadian bencana alam pada bulan tertentu, jumlah korban serta kerusakan yang terjadi.

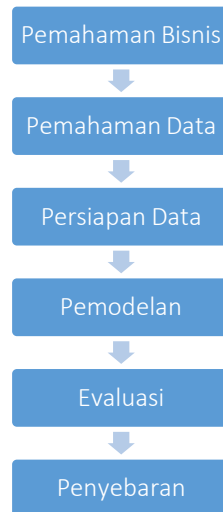
2. METODE PENELITIAN

Tahapan penelitian berlandaskan pada framework CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang merupakan framework yang dipakai pada proses data mining. Framework CRISP-DM diawali dengan penentuan permasalahan, dimana permasalahan pada penelitian ini yaitu data bencana alam yang hanya disimpan tanpa diolah untuk menghasilkan informasi yang lebih detail, misalnya berupa informasi daerah rawan bencana,



jenis bencana yang sering melanda suatu daerah, periode kejadian bencana, jumlah korban dan kerusakan yang ditimbulkan. Dari permasalahan penelitian kemudian ditentukan pendekatan yang digunakan, dalam hal ini yaitu pengelompokkan (*clustering*) data bencana alam memakai algoritma K-Means dengan aplikasi Rapidminer. Selanjutnya, yaitu di tahap implementasi, maka ditentukan objek implementasi yang digunakan yaitu Dataset Bencana dan desain eksperimen menggunakan CRISP-DM.

Adapun tahapan yang dilaksanakan dalam kegiatan penelitian meliputi:



Gambar 1. Tahapan Kegiatan Penelitian

Berikut penjelasan dari tahapan yang dilakukan dalam kegiatan penelitian seperti yang tampak pada gambar 1 yaitu:

1. Pemahaman Bisnis

Pada tahap awal ini, tim peneliti memahami tujuan pengumpulan data bencana alam yang terjadi di Indonesia, yaitu mengingat letak Indonesia yang terletak di area rawan bencana agar informasi ini dapat digunakan untuk meminimalkan dampak bencana dan memprediksi terjadinya bencana alam di kemudian hari. Untuk mengetahui informasi tersebut, perlu dilakukan pengelompokkan terhadap data bencana yang memiliki karakteristik yang sama dengan menerapkan proses data mining pada data tersebut. Untuk dapat mengidentifikasi wilayah rawan bencana maupun menentukan jenis bencana yang sering terjadi di suatu wilayah maka perlu dilakukan pengelompokkan data bencana alam terlebih dahulu. Data bencana alam akan dikelompokkan ke dalam beberapa cluster sesuai karakteristiknya masing-masing.

2. Pemahaman Data

Peneliti mengumpulkan data awal yang dipakai dalam penelitian ini yaitu berupa dataset mengenai bencana alam yang terjadi di 34 Provinsi di negara Indonesia selama tahun 2014 sampai tahun 2018 yang dihimpun dari BNPB. Data bencana alam yang dikumpulkan berupa berbagai kejadian bencana alam yang terjadi yaitu banjir, tanah longsor, abrasi, puting beliung, gempa bumi, tsunami, gempa bumi & tsunami serta letusan gunung berapi. Data yang telah terkumpul nantinya akan dideskripsikan kemudian disimpulkan pemahaman terhadap data tersebut.

3. Persiapan Data

Dari data bencana diperoleh sebanyak 10.600 kejadian bencana alam di Indonesia dalam periode waktu 2014 sampai 2018. Dalam langkah ini, peneliti menentukan struktur data untuk membantu dan memudahkan proses data mining. Persiapan data ini terdiri dari 3 tahapan antara lain (1) memilih atribut data yang akan digunakan (*data selection*) yaitu atribut Nama Provinsi, Bulan, Kejadian, Jumlah Kejadian Per Provinsi, Jumlah Kejadian Per Bulan, Jumlah Korban, Jumlah Kerusakan Fasilitas, (2) memastikan kualitas data yang dipilih (*data processing*) di mana data akan dibersihkan dari berbagai kesalahan dimana terdapat kesalahan penulisan yang mencakup penulisan nama yang tidak sama walaupun mengacu pada suatu kejadian yang sama seperti "LETUSAN GUNUNGAPI" dengan "LETUSAN GUNUNG API" serta missing value yang diatasi dengan memasukkan angka 0, dan (3) mengelompokkan atribut-atribut yang terpilih (*data transformation*) menjadi 1 tabel agar dapat diproses lebih lanjut.

4. Pemodelan

Menentukan metode *clustering* yang akan dimanfaatkan yakni *partitional clustering* menggunakan algoritma *k-means clustering*, di mana harus ditentukan jumlah kluster terlebih dahulu. Proses *data mining* terhadap data bencana alam akan menggunakan *tools* RapidMiner.

5. Evaluasi

Pada tahap ini, tim peneliti mengevaluasi hasil *data mining* apakah pengelompokan data bencana alam sudah dapat mencapai tujuan pada fase pertama (pemahaman bisnis).

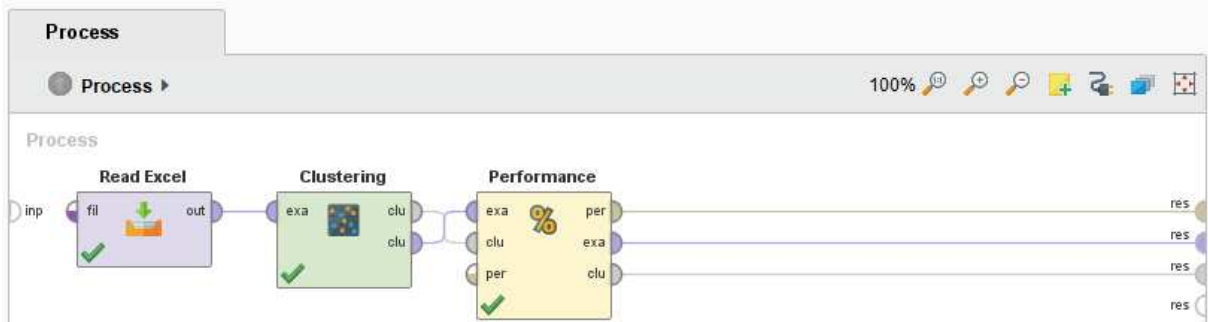
6. Penyebaran



Pada tahap akhir ini, tim peneliti menyimpulkan hasil dari proses data mining berupa pengelompokkan data bencana alam yang dapat digunakan sebagai rekomendasi untuk menentukan strategi dalam menanggulangi peristiwa bencana alam dan memprediksi kejadian bencana alam ke depannya.

3. HASIL DAN PEMBAHASAN

Setelah data selesai ditransformasi dan dibersihkan, proses clustering mulai dijalankan untuk mengelompokkan data-data yang memiliki kemiripan dengan memakai algoritma k-means. Proses data mining ini dilakukan lewat *tools* RapidMiner yang akan memudahkan proses iterasi untuk menghitung jarak data ke centroid sampai jarak datanya tidak berubah lagi. Pemodelan data yang digunakan pada RapidMiner dapat dilihat pada Gambar 2.



Gambar 2. Model pada RapidMiner

Pada gambar di atas, terlihat terdapat 3 proses yang dimasukkan ke dalam RapidMiner. Proses pertama adalah “Read Excel” yang berfungsi untuk membaca data-data bencana alam yang telah disimpan dalam *file Excel*. Pada proses ini, pengguna harus memilih *sheet* mana dalam *file Excel* yang akan dijadikan tabel data. Tabel tersebut harus diformat sedemikian rupa sehingga setiap kolom mewakili 1 atribut. Baris pertama pada tabel boleh berisikan nama atribut (bukan data). Tabel data boleh terletak pada *range* manapun dalam *sheet* Excel dan boleh mengandung baris ataupun kolom yang kosong nilainya.

Proses kedua merupakan proses “Clustering” yang berfungsi untuk mengelompokkan data ke dalam *cluster-cluster* menggunakan algoritma *k-means* dimana pada proses ini, pengguna memasukkan nilai $k=3$ (jumlah *cluster*). Algoritma *k-means* termasuk ke dalam *unsupervised machine learning*. Data-data yang memiliki kemiripan akan dikelompokkan ke dalam *cluster* yang sama sampai membentuk sejumlah k *cluster*. Kesamaan antara data didasarkan pada ukuran jarak di antara mereka. *Cluster* dalam algoritma *k-means* ditentukan oleh posisi pusat yang disebut sebagai *centroid*. Algoritma *k-means* dimulai dengan titik k yang diperlakukan sebagai pusat massa dari k *cluster* potensial. Titik awal ini adalah posisi k yang ditarik secara acak. Semua data akan diarahkan ke *cluster* terdekat. Berikutnya *centroid cluster* dihitung ulang dengan merata-ratakan semua data dalam satu *cluster*. Langkah-langkah sebelumnya diulang untuk menentukan *centroid* baru sampai *centroid* tidak lagi bergerak atau langkah optimalisasi maksimum tercapai. Prosedur ini diulangi sebanyak *max runs* yang ditentukan oleh pengguna dengan setiap kali set titik awal yang berbeda. Pada penelitian ini *max runs* diset nilainya 10. *Tools* akan menampilkan himpunan *cluster* yang memiliki jumlah minimal jarak kuadrat dari semua data ke *centroid* yang sesuai.

Pada proses ketiga terdapat proses “Performance” yang bertujuan untuk mengevaluasi kinerja metode pengelompokan berbasis *centroid*. Model *cluster centroid* memiliki informasi mengenai *clustering* yang dilakukan, seperti memberitahu data mana yang merupakan bagian dari *cluster* mana. Proses ini juga memiliki informasi mengenai *centroid* dari setiap *cluster*. Proses ini memberikan daftar nilai kriteria kinerja berdasarkan *centroid cluster* yaitu *Average within cluster distance* dan *Davies-Bouldin Index*. *Average within cluster distance* diperoleh dengan merata-ratakan jarak antara *centroid* dan semua data dalam *cluster*. Pada *cluster* dengan jarak dalam *cluster* yang rendah (kesamaan *intra-cluster* yang tinggi) dan jarak antar *cluster* yang tinggi (kesamaan antar *cluster* yang rendah) akan memiliki indeks *Davies-Bouldin* yang rendah. Algoritma *clustering* yang menghasilkan kumpulan *cluster* dengan indeks *Davies-Bouldin* terkecil dianggap sebagai algoritma terbaik berdasarkan kriteria ini.

3.1 Pengelompokkan Jumlah Kejadian Bencana Alam Per Provinsi

Data pertama yang dimasukkan adalah data jumlah kejadian bencana alam per provinsi di mana isinya adalah jumlah kejadian bencana alam yang menjadi 8 atribut yaitu banjir, gelombang pasang / abrasi, gempa bumi, gempa bumi & tsunami, letusan gunung api, puting beliung, tanah longsor, serta tsunami yang terjadi pada 34 provinsi di Indonesia. Nilai k ditentukan sebanyak 3 *cluster*. Tujuannya untuk mengelompokkan provinsi-provinsi yang sering mengalami kejadian bencana alam. Gambar 3 memperlihatkan hasil *clustering* dari RapidMiner.



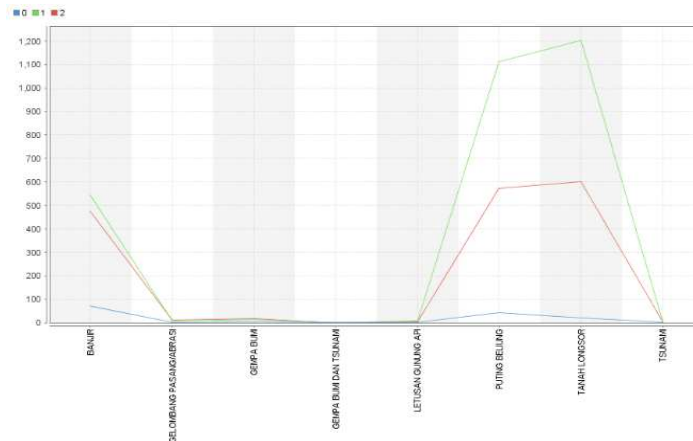
Cluster Model

Cluster 0: 31 items
 Cluster 1: 1 items
 Cluster 2: 2 items
 Total number of items: 34

Gambar 3. Hasil Clustering Jumlah Kejadian Bencana Alam Per Provinsi

Dari sebanyak 34 data tersebut, pembagiannya ke dalam cluster adalah sebagai berikut:

- Cluster 0 terdiri dari 31 data, yaitu Aceh, Bali, Banten, Bengkulu, DI Yogyakarta, DKI Jakarta, Gorontalo, Jambi, Kalimantan Utara, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Selatan, Kepulauan Riau, Kepulauan Bangka Belitung, Lampung, Maluku, Maluku Utara, Nusa Tenggara Barat, Nusa Tenggara Timur, Papua, Papua Barat, Riau, Sulawesi Utara, Sulawesi Barat, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Selatan, Sumatera Utara, Sumatera Barat, dan Sumatera Selatan. *Cluster 0* terdiri dari provinsi-provinsi yang jumlah kejadian bencana alamnya tidak terlalu tinggi jika dibandingkan dengan provinsi pada *cluster 1* dan *2*.
- Cluster 1 terdiri dari 1 data, yaitu Jawa Tengah yaitu provinsi yang paling sering mengalami kejadian bencana alam.
- Cluster 2 terdiri dari 2 data, yaitu Jawa Barat dan Jawa Timur yaitu provinsi yang sering mengalami kejadian bencana alam.



Gambar 4. Grafik Centroid Jumlah Kejadian Bencana Alam Per Provinsi

3.2 Pengelompokkan Jumlah Kejadian Bencana Alam Per Bulan

Berikutnya, data yang dimasukkan ke dalam RapidMiner adalah data jumlah kejadian bencana alam per bulan di mana isinya adalah jumlah kejadian bencana alam yang menjadi 8 atribut yaitu banjir, gelombang pasang / abrasi, gempa bumi, gempa bumi & tsunami, letusan gunung api, puting beliung, tanah longsor, dan tsunami yang terjadi pada saat bulan Januari sampai Desember. Nilai k ditentukan sebanyak 3 cluster. Tujuannya untuk mengelompokkan bulan-bulan yang rawan terjadi bencana alam. Gambar 5 menunjukkan hasil clustering dari RapidMiner.

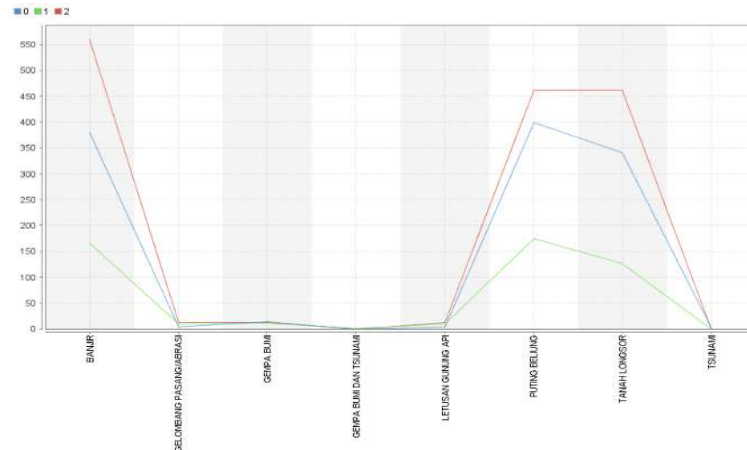
Cluster Model

Cluster 0: 4 items
 Cluster 1: 6 items
 Cluster 2: 2 items
 Total number of items: 12

Gambar 5. Hasil Clustering Jumlah Kejadian Bencana Alam Per Bulan

Berdasarkan data yang dimasukkan ke dalam RapidMiner, pembagian 12 bulan tersebut ke dalam cluster adalah sebagai berikut:

- Cluster 0 terdiri dari 4 data, yaitu Maret, April, November, dan Desember merupakan bulan-bulan yang memiliki jumlah kejadian bencana alam sedang.
- Cluster 1 terdiri dari 6 data, yaitu Mei, Juni, Juli, Agustus, September, dan Oktober merupakan bulan yang jumlah kejadian bencana alamnya relatif rendah.
- Cluster 2 terdiri dari 2 data, yaitu Januari dan Februari yaitu bulan yang paling sering terjadi bencana alam.



Gambar 6. Grafik Centroid Jumlah Kejadian Bencana Alam Per Bulan

3.3 Pengelompokkan Jumlah Korban Bencana Alam Per Provinsi

Kemudian, data yang dimasukkan ke dalam RapidMiner adalah data jumlah korban bencana alam per provinsi di mana isinya adalah jumlah korban kejadian bencana alam yang terbagi menjadi 8 atribut yaitu banjir, gelombang pasang / abrasi, gempa bumi, gempa bumi & tsunami, letusan gunung api, puting beliung, tanah longsor, serta tsunami yang berjatuh pada 34 Provinsi di seluruh Indonesia. Nilai k masih tetap ditentukan sebanyak 3 cluster. Tujuannya untuk mengelompokkan provinsi-provinsi yang paling banyak korbannya pada saat terjadi bencana alam. Gambar 7 menampilkan hasil clustering dari RapidMiner.

Cluster Model

```
Cluster 0: 27 items
Cluster 1: 6 items
Cluster 2: 1 items
Total number of items: 34
```

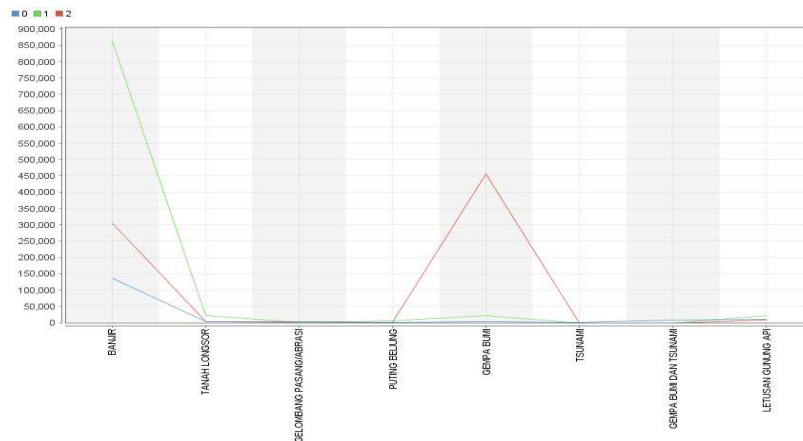
Gambar 7. Hasil Clustering Jumlah Korban Bencana Alam Per Provinsi

Hasil clustering 34 Provinsi berdasarkan jumlah korban bencana alam adalah sebagai berikut:

- Cluster 0 terdiri dari 27 data, yaitu Sumatera Utara, Sumatera Barat, Riau, Kepulauan Riau, Jambi, Sumatera Selatan, Kepulauan Bangka Belitung, Bengkulu, Lampung, DI Yogyakarta, Banten, Bali, Nusa Tenggara Timur, Kalimantan Utara, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Selatan, Sulawesi Utara, Sulawesi Barat, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Selatan, Maluku, Maluku Utara, Papua Barat, dan Papua. Provinsi-provinsi di atas memiliki jumlah korban banjir, tanah longsor, puting beliung, gempa bumi yang lebih rendah daripada provinsi dalam *cluster* lain. Untuk kejadian banjir, *centroid* sebesar 136157,815. Untuk kejadian tanah longsor, *centroid* sebesar 2458,556. Untuk kejadian puting beliung, *centroid* sebesar 1048,148. Untuk kejadian gempa bumi, *centroid* sebesar 2353,593. Kalau dilihat dari jumlah korban gelombang pasang/abrasi, tsunami, serta gempa bumi dan tsunami, jumlah korbannya bisa dikatakan lebih tinggi dibandingkan provinsi lain. Hal ini terlihat dari *centroid* gelombang pasang/abrasi dengan angka 4398,148, *centroid* tsunami dengan angka 2060,889, serta *centroid* gempa bumi dan tsunami dengan angka 8489,37. Apabila dilihat dari segi jumlah korban letusan gunung api, maka termasuk jumlah korban sedang di mana angka *centroid* adalah sebesar 10455,556.
- Cluster 1 terdiri dari 6 data, yaitu Aceh, DKI Jakarta, Gorontalo, Jawa Barat, Jawa Tengah, dan Jawa Timur merupakan provinsi dengan jumlah korban bencana alam banjir jauh lebih tinggi (*centroid* 863400) dibandingkan provinsi-provinsi lain. Lalu, jumlah korban bencana alam untuk kejadian tanah longsor, puting beliung, dan letusan gunung api juga lebih tinggi dibandingkan provinsi pada *cluster* lain. *Centroid* jumlah korban untuk kejadian tanah longsor adalah 20917,833, *centroid* untuk puting beliung adalah 5788,333, dan *centroid* untuk letusan gunung api adalah sebesar 20520. Sementara itu, jumlah korban gelombang pasang/abrasi dengan *centroid* 1757,5 dan jumlah korban gempa bumi dengan *centroid* 20770, 833 berada pada kategori sedang, dan jumlah korban tsunami serta gempa bumi dan tsunami yang memiliki *centroid* sebesar 0 masuk kategori rendah.
- Cluster 2 terdiri dari 1 data, yaitu Nusa Tenggara Barat, jumlah korban gempa bumi dengan *centroid* sebesar 456025 lebih tinggi daripada provinsi-provinsi lain. Untuk kejadian banjir, tanah longsor, puting beliung, jumlah korbannya termasuk sedang dengan *centroid* banjir adalah sebesar 304981, *centroid* tanah longsor sebesar 2516, dan *centroid* puting beliung sebesar 1686. Sedangkan kejadian gelombang pasang/abrasi yang memiliki *centroid* 823, kejadian tsunami yang memiliki *centroid* 0, kejadian gempa bumi dan tsunami yang



memiliki *centroid* 0, serta kejadian letusan gunung api yang memiliki *centroid* 7328, termasuk dalam yang jumlah korbannya relatif rendah.



Gambar 8. Grafik Centroid Jumlah Korban Bencana Alam Per Provinsi

3.4 Pengelompokkan Jumlah Kerusakan Fasilitas Per Provinsi

Terakhir, data yang akan dikelompokkan menggunakan RapidMiner adalah data jumlah kerusakan fasilitas per provinsi di mana isinya adalah jumlah kerusakan fasilitas yang ditimbulkan akibat atribut bencana alam yaitu banjir, tanah longsor, gelombang pasang / abrasi, gempa bumi, gempa bumi & tsunami, letusan gunung api, puting beliung, serta tsunami yang terletak pada 34 Provinsi Indonesia. Pengelompokkan provinsi dihasilkan ke 3 cluster. Tujuannya untuk mengelompokkan provinsi-provinsi yang mengalami paling banyak kerusakan fasilitas pada waktu terjadi bencana alam. Gambar 9 menyajikan hasil pengelompokkan RapidMiner.

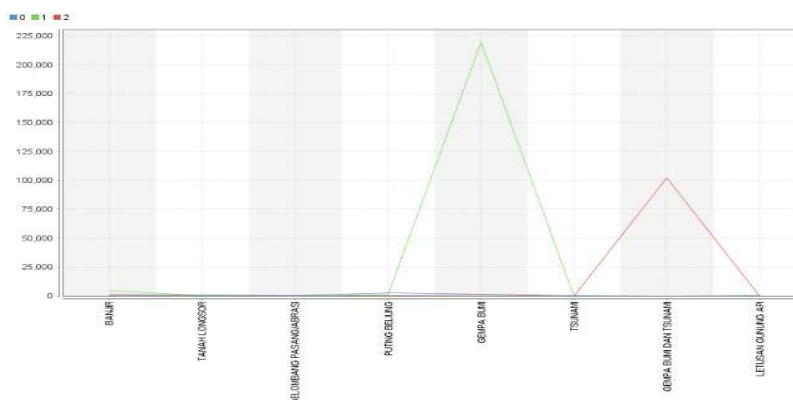
Cluster Model

```
Cluster 0: 32 items
Cluster 1: 1 items
Cluster 2: 1 items
Total number of items: 34
```

Gambar 9. Hasil Clustering Jumlah Kerusakan Fasilitas Per Provinsi

Hasil clustering 34 Provinsi berdasarkan jumlah korban bencana alam adalah sebagai berikut:

- Cluster 0 terdiri dari 32 data, yaitu Aceh, Sumatera Utara, Sumatera Barat, Riau, Kepulauan Riau, Jambi, Sumatera Selatan, Kepulauan Bangka Belitung, Bengkulu, Lampung, Banten, Jawa Barat, DKI Jakarta, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Timur, Kalimantan Utara, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Selatan, Sulawesi Utara, Gorontalo, Sulawesi Barat, Sulawesi Tenggara, Sulawesi Selatan, Maluku, Maluku Utara, Papua Barat, dan Papua. Rata-rata provinsi pada *cluster* ini memiliki jumlah kerusakan fasilitas yang lebih tinggi pada kejadian tanah longsor dengan *centroid* 613,25, gelombang pasang/abrasi dengan *centroid* 28,75, puting beliung dengan *centroid* 2694,406, tsunami dengan *centroid* 86,094, dan letusan gunung api dengan *centroid* 566,812. Lalu, jumlah kerusakan fasilitas yang sedang pada kejadian banjir dengan *centroid* 1525 dan kejadian gempa bumi dengan *centroid* 1202,625. Jumlah kerusakan fasilitas pada kejadian gempa bumi dan tsunami termasuk rendah di mana nilai *centroid* adalah 0.
- Cluster 1 terdiri dari 1 data, yaitu Nusa Tenggara Barat memiliki jumlah kerusakan fasilitas pada kejadian gempa bumi yang jauh lebih tinggi dengan nilai *centroid* sebesar 219577. Jumlah kerusakan fasilitas pada kejadian banjir juga lebih tinggi yaitu nilai *centroid* sebesar 4286. Pada kejadian tanah longsor dan kejadian puting beliung, jumlah kerusakan fasilitasnya relatif sedang di mana nilai *centroid* kejadian tanah longsor adalah 17 dan nilai *centroid* kejadian puting beliung adalah 1549. Untuk kejadian gelombang pasang/abrasi, tsunami, gempa bumi & tsunami, serta letusan gunung api, *cluster* ini memiliki jumlah kerusakan fasilitas yang rendah ditandai dengan nilai *centroid* untuk keempat kejadian tersebut adalah 0.
- Cluster 2 terdiri dari 1 data, yaitu Sulawesi Tengah yang jumlah kerusakan fasilitas kejadian gempa bumi dan tsunami yang tinggi, terlihat dari nilai *centroid* sebesar 101839. Jumlah kerusakan fasilitas gelombang pasang/abrasi sedang yaitu dengan *centroid* senilai 24. Diikuti dengan jumlah kerusakan fasilitas yang rendah pada kejadian banjir yang memiliki *centroid* 533, kejadian tanah longsor yang memiliki *centroid* 11, puting beliung yang memiliki *centroid* 67, gempa bumi yang memiliki *centroid* 597, serta kejadian tsunami dan juga letusan gunung api yang sama-sama memiliki *centroid* 0.



Gambar 10. Grafik Centroid Jumlah Kerusakan Fasilitas Per Provinsi

4. KESIMPULAN

Berdasarkan data kejadian bencana alam Indonesia selama tahun 2014-2018 yang telah dikumpulkan, ditransformasi, dan diolah menggunakan tools RapidMiner, maka didapatkan kesimpulan pengelompokan data bencana alam sebagai berikut:

- Dari segi pengelompokan data berdasarkan jumlah kejadian bencana alam per provinsi, didapatkan bahwa cluster 1 yaitu Jawa Tengah yang menjadi provinsi yang mengalami kejadian bencana alam paling sering terutama tanah longsor, puting beliung, dan banjir, cluster 2 yaitu Jawa Barat dan Jawa Timur yang jumlah kejadian bencana alamnya sedang dengan pengecualian kejadian gelombang pasang/abrasi, gempa bumi, dan tsunami yang sedikit lebih sering terjadi, dan cluster 0 merupakan provinsi-provinsi lain yang jumlah kejadian bencana alamnya rendah.
- Dari segi pengelompokan data berdasarkan jumlah kejadian bencana alam per bulan, diperoleh hasil bahwa cluster 2 yaitu bulan Januari dan Februari merupakan bulan yang paling sering mengalami kejadian bencana alam, cluster 0 yaitu bulan Maret, April, November, dan Desember merupakan bulan yang jumlah kejadian bencana alamnya sedang dengan pengecualian kejadian gempa bumi dan kejadian tsunami yang sedikit lebih sering terjadi, dan cluster 1 yang berisi bulan-bulan lain yang jumlah kejadian bencana alamnya relatif rendah, kecuali kejadian gempa bumi dan tsunami.
- Dari segi pengelompokan data berdasarkan jumlah korban bencana alam per provinsi, hasilnya berupa cluster 1 antara lain Aceh, Gorontalo, Jawa Barat, DKI Jakarta, Jawa Tengah, dan Jawa Timur merupakan provinsi yang jumlah korban banjir, tanah longsor, puting beliung, dan letusan gunung api lebih tinggi, jumlah korban gelombang pasang/abrasi dan gempa bumi sedang, dan jumlah korban tsunami serta gempa bumi dan tsunami yang rendah, cluster 2 yaitu Nusa Tenggara Barat adalah provinsi yang jumlah korban gempa bumi lebih tinggi, jumlah korban banjir, tanah longsor, puting beliung sedang, serta jumlah korban gelombang pasang/abrasi, tsunami, gempa bumi & tsunami, serta letusan gunung api yang rendah, kemudian ada cluster 0 yang terdiri dari provinsi-provinsi lain yang jumlah korban gelombang pasang/abrasi, tsunami, serta gempa bumi dan tsunami lebih tinggi, jumlah korban letusan gunung api sedang, dan jumlah korban banjir, tanah longsor, puting beliung, gempa bumi yang lebih rendah.
- Dari segi pengelompokan data berdasarkan jumlah kerusakan fasilitas per provinsi, diperoleh kesimpulan bahwa cluster 1 berisi provinsi Nusa Tenggara Barat yang jumlah kerusakan fasilitas pada kejadian gempa bumi dan banjir lebih tinggi, jumlah kerusakan fasilitas pada kejadian tanah longsor dan puting beliung sedang, serta jumlah kerusakan fasilitas pada kejadian gelombang pasang/abrasi, tsunami, gempa bumi & tsunami, serta letusan gunung api yang rendah, cluster 2 berisi provinsi Sulawesi Tengah yang jumlah kerusakan fasilitas pada kejadian gempa bumi dan tsunami lebih tinggi, jumlah kerusakan fasilitas pada kejadian gelombang pasang/abrasi sedang, serta jumlah kerusakan fasilitas pada kejadian banjir, tanah longsor, puting beliung, gempa bumi, tsunami, serta letusan gunung api yang rendah, dan juga cluster 0 yang memiliki jumlah kerusakan fasilitas pada kejadian tanah longsor, gelombang pasang/abrasi, puting beliung, tsunami, serta letusan gunung api yang lebih tinggi, jumlah kerusakan fasilitas pada kejadian banjir dan gempa bumi yang sedang, serta jumlah kerusakan fasilitas pada kejadian gempa bumi dan tsunami termasuk rendah.

Untuk penelitian selanjutnya, diharapkan agar penentuan centroid untuk keperluan clustering tidak dilakukan secara acak, tetapi melalui algoritma tertentu untuk menemukan posisi centroid awal. Selain itu, diharapkan pula agar hasil clustering data bencana alam ini dapat dilanjutkan untuk melakukan prediksi data di masa mendatang.

REFERENCES

- [1] B. Badan Pusat Statistik, "Statistik Indonesia 2018," *Badan Pusat Statistik*, 2018. [Online]. Available: <https://www.bps.go.id/publication/2018/07/03/5a963c1ea9b0fed6497d0845/statistik-indonesia-2018>. [Accessed: 25-



- Aug-2019].
- [2] P. V. dan M. B. G. Badan Geologi, "Data Dasar Gunung Api Indonesia," *Pusat Vulkanologi dan Mitigasi Bencana Geologi*, 2019. [Online]. Available: <http://www.vsi.esdm.go.id/index.php/gunungapi/data-dasar-gunungapi>. [Accessed: 25-Aug-2019].
- [3] N. O. and A. A. NOAA, "Natural Hazards Viewer," *National Oceanic and Atmospheric Administration*, 2019. [Online]. Available: <https://maps.ngdc.noaa.gov/viewers/hazards/?layers=0#>. [Accessed: 26-Aug-2019].
- [4] N. G. D. C. D. C. NGDC/WDC, "Tsunami Runup Data," *National Geophysical Data Center/World Data Center*, 2019. [Online]. Available: https://www.ngdc.noaa.gov/hazard/tsu_db.shtml. [Accessed: 26-Aug-2019].
- [5] B. N. P. B. BNPB, "Data Statistik Bencana," *Badan Nasional Penanggulangan Bencana*, 2018. [Online]. Available: <http://bnpb.cloud/dibi/tabel1b>. [Accessed: 25-Aug-2019].
- [6] 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.
- [7] B. Supriyadi, A. P. Windarto, T. Soemartono, and Mungad, "Classification of Natural Disaster Prone Areas in Indonesia Using K-Means," *Int. J. Grid Distrib. Comput.*, vol. 11, no. 8, pp. 87–98, 2018.
- [8] M. S. Yana, L. Setiawan, E. M. Ulfa, and A. Rusyana, "Penerapan Metode K-Means dalam Pengelompokan Wilayah Menurut Intensitas Kejadian Bencana Alam di Indonesia Tahun 2013-2018," *J. Data Anal.*, vol. 1, no. 2, pp. 93–102, 2018.
- [9] Prihandoko, Bertalya, and M. I. Ramadhan, "An Analysis of Natural Disaster Data By Using K-Means and K-Medoids Algorithm of Data Mining Techniques," *QiR 2017 - 2017 15th Int. Conf. Qual. Res. Int. Symp. Electr. Comput. Eng.*, vol. 2017-Decem, pp. 221–225, 2017.
- [10] E. F. Sirat, B. D. Setiawan, and F. Ramdani, "Comparative Analysis of K-Means and Isodata Algorithms for Clustering of Fire Point Data in Sumatra Region," *2018 4th Int. Symp. Geoinformatics, ISyG 2018*, pp. 1–6, 2018.
- [11] M. I. Ramadhan, "Penerapan Data Mining Untuk Analisis Data Bencana Milik BNPB Menggunakan Algoritma K-Means dan Linear Regression," *J. Ilm. Inform. Komput.*, vol. 22, no. 1, 2017.
- [12] H. Artatia and H. R.B., "Pengelompokan Dampak Gempa Bumi Dari Segi Kerusakan Fasilitas Pada Provinsi Yang Berpotensi Gempa Di Indonesia Menggunakan K-Means-Clustering," pp. 742–748, 2015.
- [13] C. A. Sugianto and M. N. Astita, "Implementasi Data Mining Dalam Data Bencana Tanah Longsor Di Jawa Barat Menggunakan Algoritma Fp-Growth," *Techno.Com*, vol. 17, no. 1, pp. 91–102, 2018.
- [14] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018.
- [15] M. A. Sembiring, M. F. L. Sibuea, and A. Sapta, "Analisa Kinerja Algoritma C.45 Dalam Memprediksi Hasil Belajar," *J. Sci. Soc. Res.*, vol. 1, no. February, pp. 73–79, 2018.
- [16] B. Depaire, G. Wets, and K. Vanhoof, "Traffic accident segmentation by means of latent class clustering," *Accid. Anal. Prev.*, 2008.