**KLASTERISASI WILAYAH TANAH LONGSOR BERDASARKAN DAMPAK WILAYAH DAN GEOGRAFIS MENGGUNAKAN METODE K-MEANS**

**(Studi Kasus : Kabupaten dan Kota di Jawa Timur)**

**Pramana Yoga Saputra, S.Kom., MMT.1 ,** **Habibie Ed Dien, S.Kom., MT.2, Moch. Rizki Eko Waluyo3**

1,2,3 Jurusan Teknologi Informasi, Program Studi Teknik Informatika, Politeknik Negeri Malang

[1pramana.yoga@polinema.ac.id](mailto:1pramana.yoga@polinema.ac.id)., [2habibie@polinema.ac.id](mailto:2habibie@polinema.ac.id) ,[3mochrizkiekow@gmail.com](mailto:3mochrizkiekow@gmail.com)

**Abstrak**

Tanah Longsor adalah suatu peristiwa geologi yang terjadi karena pergerakan massa batuan atau tanah dengan berbagai tipe dan jenis seperti jatuhnya bebatuan atau gumpalan besar tanah. Berdasarkan data yang dihimpun BPBD sepanjang tahun 2017-2018, terjadi sebanyak 212 kejadian dan mengakibatkan lebih dari 60 korban jiwa pada wilayah kabupaten & kota di Jawa Timur. Oleh karena itu, diperlukan usaha mengurangi dampak yang ditimbulkan dengan melakukan klasterisasi terhadap daerah kerawanan di wilayah kabupaten & kota di Jawa Timur.

Dalam penelitian ini memanfaatkan teknik data mining dengan menggunakan metode K-Means untuk mengelompokkan wilayah terdampak tanah longsor dan geografis. Selain itu, penelitian ini juga menggunakan metode Pearson Correlation Coefficient untuk mengetahui tingkat keterkaitan antara dampak wilayah tanah longsor dengan geografis. Parameter pada penelitian ini terdiri dari jumlah kejadian, jumlah korban dan jumlah kerusakan serta beberapa parameter lainnya seperti kemiringan lereng, jenis tanah, curah hujan tahunan.

Berdasarkan hasil klasterisasi menggunakan metode K-means terdapat tiga klaster untuk data wilayah tanah longsor dan data geografis. Hasil pengujian mendapatkan nilai baik/optimal sebesar 0.021298251540091 untuk pengujian klasterisasi data wilayah tanah longsor dan nilai sebesar 0.24021392702327 untuk pengujian klasterisasi data geografis menggunakan davies-bouldin index. Sedangkan untuk hasil pengujian menggunakan purity hasil klasterisasi data wilayah tanah longsor mendapatkan nilai cukup baik sebesar 0.92 dengan error cluster 0.08 dan hasil klasterisasi data geografis mendapatkan nilai kurang baik sebesar 0.2 dengan error cluster 0.8. Hasil pengukuran keterkaitan data menggunakan Pearson Correlation Coefficient mendapatkan nilai 2.2959588452857e-05 nilai korelasi positif yang berarti data dampak wilayah tanah longsor dan geografi memiliki korelasi secara linear.

**Kata kunci** : data mining, klasterisasi, k-means, longsor, korelasi

# **Pendahuluan**

Bencana adalah peristiwa atau rangkaian peristiwa yang mengancam dan mengganggu kehidupan dan penghidupan masyarakat yang disebabkan, baik oleh faktor alam dan atau faktor non alam maupun faktor manusia sehingga mengakibatkan timbulnya korban jiwa manusia, kerusakan lingkungan, kerugian harta benda, dan dampak psikologis (BNPB, 2019). Bencana merupakan pertemuan dari tiga unsur, yaitu ancaman bencana, kerentanan, dan kemampuan yang dipicu oleh suatu kejadian (Indonesia, 2007), dengan definisikan bencana menjadi beberapa jenis tentunya memerlukan penanganan yang berbeda pula.

Bencana tanah longsor adalah bencana yang sering terjadi di berbagai wilayah di indonesia dan dapat mengakibatkan korban jiwa karena kejadian dan waktu tiba-tiba. Jawa timur merupakan salah satu dari tiga provinsi yang sering terjadi bencana tanah longsor di indonesia. Bencana tanah longsor hampir merata di setiap daerah kabupaten dan kota di provinsi jawa timur. Berdasarkan data yang dihimpun oleh Badan Nasional Penanggulangan Bencana (BNPB), Pada Provinsi Jawa Timur kejadian tanah longsor sepanjang tahun 2017-2018 terjadi sebanyak 212 kejadian, dari total kejadian tersebut kabupaten ponorogo dan trenggalek menjadi daerah dengan jumlah kejadian tertinggi di Provinsi Jawa Timur. Kejadian tersebut juga mengakibatkan lebih dari 60 korban jiwa, sekitar 1998 rumah rusak berat dan sekitar 20 fasilitas pendidikan rusak.

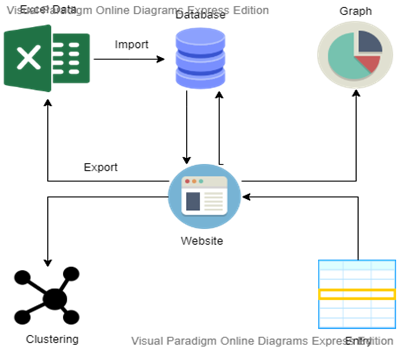
Badan Penanggulangan Bencana Daerah (BPBD) merupakan salah satu lembaga pemerintah yang menjadi pelaksana penanggulangan bencana daerah di Indonesia khususnya di Kabupaten/Kota dan Provinsi.Selama ini Badan Penanggulangan Bencana Daerah (BPBD) kurang maksimal dalam melakukan penanganan,hal ini dikarenakan kurang maksimalnya analisa data ketika pasca bencana sehingga Indeks Kerentanan Bencana masih tinggi di beberapa wilayah di Jawa Timur. Hal ini diperlukan agar dapat terbentuk penanggulangan yang tepat dan koordinasi yang baik antar dinas terkait. Sehingga BPBD dalam hal ini sebagai Pelaksana penanggulangan bencana di daerah dapat melakukan langkah awal (Pra-Bencana) guna mengurangi dampak kerusakan dari kejadian tanah longsor.

Pada Penelitian ini menggunakan analisis metode korelasional dan klasterisasi menggunakan data dari Badan Penanggulangan Bencana Daerah (BPBD) Jawa Timur Tahun 2017-2019 dan beberapa literatur sumber terkait penelitian ini. Dari data tersebut  diperoleh data dampak wilayah tanah longsor dan geografis. Dimana parameter data dampak wilayah adalah jumlah kejadian, kerusakan dan korban jika sedangkan data geografis berupa kemiringan lereng tanah,jenis tanah,dan curah hujan tahunan. Selanjutnya, hasil dari klaster tersebut dianalisis menggunakan menggunakan metode korelasional untuk menemukan korelasi dari data tersebut.

Peneliti berharap dengan adanya penelitian ini diharapkan dapat meningkatkan kinerjapihak BPBD untuk mengolah data-data wilayah tanah longsor dan menganalisis keterkaitan dampak wilayah dan geografis di Provinsi Jawa Timur. Oleh karena itu,berkaitan dengan adanya penelitian ini peneliti berencana membuat sebuah sistem pengelompokkan data wilayah tanah longsor dan menganalisis keterkaitan dampak wilayah tanah longsor dan geografis di Provinsi Jawa Timur.

# **Analisa dan Perancangan**

1. **Analisa Sistem**



Gambar 2 Gambaran Umum Sistem

Sistem akan memetakan daerah terdampak tanah longsor dan geografis berdasarkan hasil klasterisasi dari data dokumen kejadian bencana tanah longsor pada tahun 2017-2019 dan data geografis. Masukan sistem berupa data terdampak tanah longsor dan geografis di kabupaten/kota yang tercatat pihak Badan Penanggulangan Bencana Daerah. Masukan tersebut akan diolah menggunakan teknik klasterisasi dengan metode K-Means dan akan menghasilkan keluaran berupa tabel dan grafik yang sesuai dengan kluster yang terbentuk.

1. **Analisa Kebutuhan**

Analisis kebutuhan merupakan tahap untuk menganalisis segala sesuatu yang dibutuhkan untuk mengembangkan sistem, dalam hal ini terdapat beberapa hal yang harus dipenuhi antara lain meliputi dataset.

Data yang digunakan adalah data dokumen kejadian bencana tanah longsor dan geografis di kabupaten/kota yang tercatat di Badan Penanggulangan Bencana Daerah (BPBD). Adapun atribut dari data kejadian tanah longsor dan geografis yang akan dilibatkan dalam proses klusterisasi adalah sebagai berikut :

Tabel 1. Keterangan Data Wilayah Tanah Longor

|  |  |  |
| --- | --- | --- |
| No | Atribut | Keterangan |
| 1 | Jumlah Kejadian | Jumlah Kejadian tanah longsor |
| 2 | Korban Jiwa | Akumulasi jumlah korban jiwa dari korban meninggal,luka-luka,dan terdampak |
| 3 | Kerusakan | Akumulasi jumlah kerusakan dari kerusakan rumah rusak berat,ringan,sedang,fasilitas kesehatan,fasilitas peribadatan dan fasilitas pendidikan |

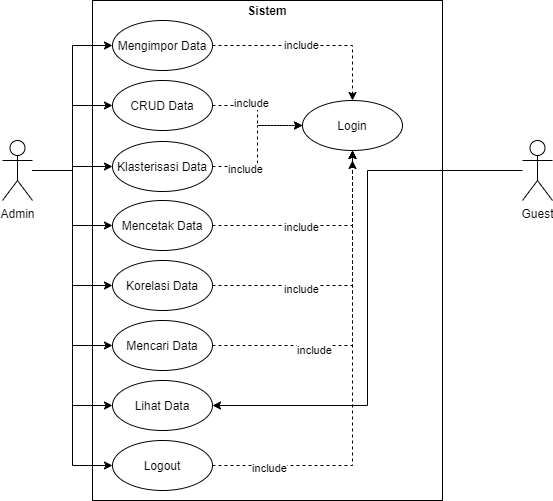
Tabel 2. Keterangan Data Geografis

|  |  |  |
| --- | --- | --- |
| No | Artibut | Keterangan |
| 1 | Curah Hujan | Rerata Curah Hujan Tahunan di daerah Kabupaten dan Kota |
| 2 | Kemiringan Tanah | Rerata Kemiringan Tanah di daerah Kabupaten dan Kota berdasarkan ketinggian topografi dibagi jarak datar sebenarnya |
| 3 | Jenis Tanah | Rerata Jenis Tanah di daerah Kabupaten dan Kota |

1. **Use Case Diagram**

Pada tahap ini yaitu berupa perancangan sistem yang digunakan untuk memberikan gambaran umum tentang sistem yang akan dibangun. Use Case Diagram.

Pengguna dalam sistem ini hanya ada satu, yang diinisialisasikan dengan “user”. Interaksi yang dapat dilakukan pengguna dengan sistem ini adalah mengelola data dampak wilayah tanah longsor dan geografis, melakukan proses klasterisasi, melihat hasil klasterisasi dan melihat pemetaan hasil klasterisasi.

****

Gambar 2. Use Case Diagram

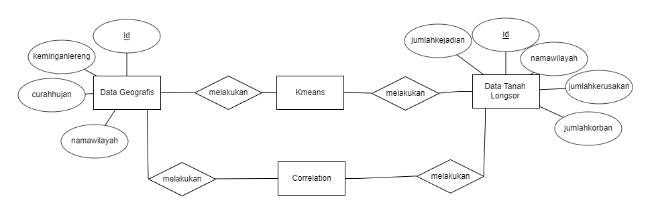
Keterangan Use Case Diagram:

Tabel 3. Keterangan Use Case Diagram

|  |  |
| --- | --- |
| Aktor | Keterangan |
| Admin | Administrator memiliki hak akses penuh. Akses tersebut meliputi manajemen data yang terdiri dari menambah, mengubah, mencetak dan menghapus data. Data data yang dimaksud adalah data wilayah terdampak tanah longsor dan geografis. Selain itu admin dapat melakukan melihat data hasil perhitungan clustering menggunakan K-Means Clustering dan data hasil perhitungan korelasi data. |
| Guest | Guest user memiliki hak akses sebagai pengguna biasa yang dapat melihat data dalam bentuk tabel. Hak akses tersebut dapat dilakukan tanpa proses login terlebih dahulu. |

1. **Entity Relationship Diagram**

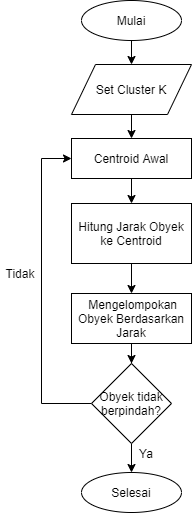
Entity Relationship diagram merupakan suatu model untuk menjelaskan hubungan antar data dalam basis data berdasarkan objek-objek dasar data yang mempunyai hubungan antar relasi. Berikut gambar dibawah untuk sistem yang akan dibangun:



Gambar 3. Entity Relationship Diagram

1. **Metode K-Means**

Pada Gambar 6, masukan untuk melakukan proses klusterisasi menggunakan metode k-means antara lain: data yang akan diolah, jumlah kluster, menentukan centroid awal,iterasi dan mengelompokkan berdasarkan kedekatan jarak euclidean. Setelah itu hitung pusat kluster, nilai objektif dan hitung ulang matriks nilai random. Perhitungan berhenti apabila memenuhi suatu kondisi, jika anggota dari cluster sebelumnya tidak berubah/telah mencapai maksimum iterasi



Gambar 4. Flowchart K-Means

1. **Alur Proses Klasterisasi**

Pada alur proses klasterisasi pada sistem yang akan dibangun, user diminta memilih data yang akan digunakan dalam proses klasterisasi. Kemudian sistem akan memeriksa jumlah data yang dipilih, apabila jumlah data tidak memenuhi kondisi untuk dilakukan proses klasterisasi, user diminta memilih data yang lain. Jika data yang dipilih memenuhi kondisi untuk dilakukan proses klasterisasi, sistem akan memeriksa data yang dipilih apabila data pernah diklusterisasi sebelumnya maka user akan diminta untuk memilih data lain. Apabila data yang dipilih telah memenuhi dua kondisi sebelumnya, sistem akan memeriksa jumlah kluster yang dimasukkan oleh user. Jumlah kluster yang dimasukkan harus tidak lebih besar dari jumlah data yang dipilih, jika jumlah kluster yang dimasukkan lebih besar dari jumlah data yang dipilih maka user diminta memasukkan jumlah kluster lain. Setelah kondisi tersebut telah dipenuhi, data dapat diklasterisasi menggunakan metode k-means. Keluaran dari sistem berupa hasil klasterisasi dan ditampilkan dalam bentuk piechart dan tabel.

1. **Davies-Bouldin Index**

Untuk pengujian hasil klaster pertama, digunakan menggunakan metode Davies-Bouldin Index (DBI). DBI merupakan metode yang digunakan untuk menguji validitas jumlah klaster. DBI ini memaksimalkan jarak inter-cluster antara cluster Ci dan Cj dan pada waktu yang sama mencoba untuk meminimalkan jarak antar titik dalam sebuah cluster. Berikut merupakan tahapan DBI yang dipergunakan

1. Data yang digunakan merupakan data hasil klasterisasi dari proses K-Means.
2. Cari nilai Sum of Square Within (SSW) untuk setiap cluster.
3. Cari nilai Sum of Square Between-Cluster (SSB).
4. Cari nilai Ratio (R).

|  |  |
| --- | --- |
|  | ( 1 ) |

Hasil akhir dari nilai DBI Cluster.

1. **Purity**

Untuk pengujian hasil klaster kedua, digunakan menggunakan metode Purity. Purity salah satu metode uji cluster dengan semua objek class yang sama berada pada cluster yang sama dikatakan murni (pure).

|  |  |
| --- | --- |
|  | ( 2 ) |

Dimana:

r = tingkat akurasi cluster purity.

k = jumlah cluster.

a = objek yang muncul.

Semakin tinggi nilai r (semakin mendekati 1), semakin baik kualitas cluster. Untuk menghitung error cluster pada purity dapat menggunakan persamaan berikut:

Dimana r adalah tingkat akurasi cluster purity.

1. **Pearson's Correlation Coefficient**

Pearson’s Correlation Coeffiecient adalah salah satu metode uji statistik Multikolinearitas yang mengukur keterkaitan antara dua atau lebih variabel. Berikut perhitungan yang digunakan untuk menghitung nilai pearson

|  |  |
| --- | --- |
|  | ( 3 ) |

Dimana:

r = koefisien korelasi pearson.

n = jumlah data.

x = variabel ke x.

y = variabel ke y.

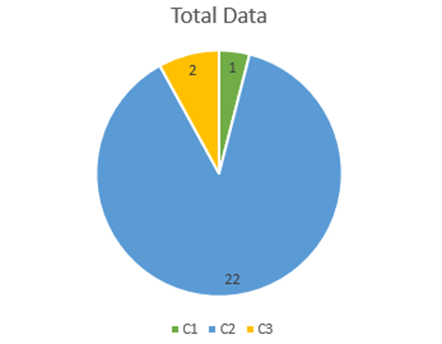
Jika r bernilai positif maka ada kedekatan korelasional secara linear.

# **Hasil dan Pengujian**

1. **Hasil**

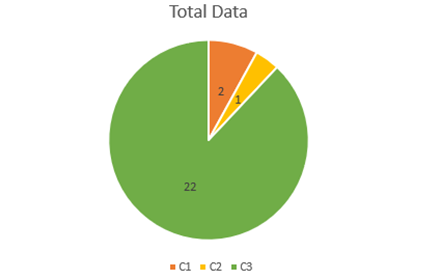
Setelah semua data wilayah terdampak tanah longsor dan geografis terkumpul , maka data-data tersebut telah dapat dikelompokan dengan menggunakan algoritma K-means Clustering. Cara penghitungannya dengan mengukur Euclidean dis-tance-nya. Pengukuran dilakukan terhadap dua ti-tik dalam satu, dua dan tiga dimensi secara beruru-tan. Untuk dapat melakukan pengelompokan data-data tersebut menjadi beberapa cluster perlu dila-kukan beberapa langkah, yaitu: 1) Menentukan jumlah cluster, dalam penelitian ini data-data yang ada akan dikelompokkan menjadi 3 cluster 2) Menentukan titik pusat awal dari setiap cluster. Dalam penelitian ini, titik pusat awal ditentukan dengan menghitung dari rata-rata data terkecil dan terbanyak dan didapat titik pusat dari setiap cluster. Pada tahap ini perlu dihitung jarak tiap data ke tiap pusat cluster. Jarak paling dekat antara satu data dengan satu cluster tertentu akan menentukan suatu data masuk dalam cluster mana. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak Euclidean,lalu dikelompokkan berdasarkan jarak terdekatnya.

Dari hasil analisa menggunakan metode K-means yang ditunjukkan gambar 9 pada data wilayah terdampak tanah longsor didapati untuk cluster 1 terdapat 1 anggota cluster yang didominasi wilayah yang sering terdampak dan kerusakan parah,untuk cluster 2 terdapat 22 anggota cluster yang didominasi dengan wilayah yang mengalami kerusakan dan korban jiwa sedang dan untuk cluster 3 terdapat 2 anggota cluster didominasi wilayah yang mengalami kerusakan,kejadian dan korban jiwa rendah.



Gambar 5. Pie Chart Hasil Klastering Data Wilayah

Hasil analisa menggunakan metode K-means ditunjukkan gambar 10 pada data geo didapati untuk cluster 1 terdapat 2 anggota cluster,untuk cluster 2 terdapat 1 anggota cluster dan untuk cluster 3 terdapat 22 anggota cluster.



Gambar 6. Pie Chart Hasil Klastering Data Geo

Sedangkan untuk hasil korelasi dari data wilayah terdampak tanah longsor dan data geografis menggunakan metode Pearson Correlation Coefficient mendapatkan nilai 0,000024080 dengan ratio positif yang berarti korelasi dari kedua data tersebut linear.

1. **Pengujian Metode K-Means pada Klasterisasi**

Untuk mengetahui hasil dari metode yang telah diterapkan pada sistem maka dilakukan proses pengujian dengan keseluruhan data yaitu sejumlah 25 data wilayah terdampak tanah longsor dan geografis di kabupaten dan kota di Jawa Timur. Dalam pengujian yang dilakukan peneliti menggunakan 2 metode yaitu Davies-Bouldin Index dan Purity.

1. **Pengujian Davies-Bouldin Index**

Uji coba ini dilakukan untuk menguji seberapa baik pengelompokan telah dilakukan dibuat menggunakan jumlah dan fitur yang melekat pada dataset setelah dilakukan proses clustering dengan menggunakan metode K-means Clustering. Dari hasil proses k-means akan terlihat hasil iterasi terakhir, selanjutnya dari iterasi tersebut akan dihitung sum of square within dan sum of square between cluster, berikutnya akan dihitung ratio menggunakan rumus (1). Dari hasil pengujian tersebut pada sistem, menunjukkan bahwa sistem dapat melakukan klasterisasi data secara baik karena nilai DBI yang diperoleh mendekati nilai nol dengan hasil nilai DBI untuk data klaster wilayah terdampak sebesar 0.021298251540091 dan nilai DBI untuk data klasterisasi geografis sebesar 0.24021392702327.

1. **Pengujian Purity**

Uji coba ini dilakukan untuk menguji seberapa baik pengelompokan telah dilakukan dinilai berdasarkan data yang tidak digunakan untuk pengelompokan, seperti label kelas yang dikenal dan tolok ukur eksternal pada dataset setelah dilakukan proses clustering dengan menggunakan metode K-means Clustering.  Dari hasil proses k-means akan terlihat hasil iterasi terakhir, selanjutnya dari iterasi tersebut akan dihitung obyek sama pada cluster, berikutnya akan dihitung ratio menggunakan rumus (2) Dari hasil pengujian pada sistem, menunjukkan bahwa sistem dapat melakukan klasterisasi data secara cukup baik untuk data klaster wilayah terdampak dengan hasil nilai Purity sebesar 0.92 dan untuk data klasterisasi geografis mendapatkan nilai Purity kurang baik sebesar 0.2 karena purity  dikatakan optimal/baik jika bernilai 1 atau mendekati 1

# **Kesimpulan dan Saran**

1. **Kesimpulan**

Berdasarkan hasil pengujian dan analisis penelitian terhadap pemetaan daerah terdampak wilayah tanah longsor dengan menggunakan metode k-means, maka dapat diambil beberapa kesimpulan yaitu:

1. Metode k-means dapat digunakan untuk melakukan pengelompokan data wilayah terdampak dan geografis di kab/kota Jawa Timur
2. Berdasarkan hasil pengujian mendapatkan nilai baik/optimal sebesar 0.021298251540091 untuk pengujian hasil klasterisasi data wilayah tanah longsor dan nilai sebesar 0.24021392702327 untuk pengujian hasil klasterisasi data geografis menggunakan davies-bouldin index.
3. Berdasarkan hasil pengujian menggunakan purity hasil klasterisasi data wilayah tanah longsor mendapatkan nilai cukup baik sebesar 0.92 dengan error cluster 0.08, Sedangkan hasil klasterisasi geografis mendapatkan nilai kurang baik sebesar 0.2 dengan error cluster 0.8.
4. Berdasarkan perhitungan korelasi yang telah dilakukan terhadap data dampak wilayah tanah longsor dan data geografis dengan metode Pearson Correlation Coefficient diperoleh nilai korelasi PCC sebesar 2.2959588452857e-05 dengan nilai korelasi positif berarti data dampak wilayah tanah longsor dan geografi memiliki korelasi secara linear.
5. **Saran**

Dalam penelitian tugas akhir ini terdapat beberapa kekurangan yang dapat diperbaiki dan ditingkatkan lagi untuk penelitian selanjutnya. Oleh karena itu, berikut beberapa saran dan masukan yang diharapkan dapat memberikan perbaikan dalam penelitian selanjutnya, antara lain:

1. Penggunaan data geografis sebaiknya didapatkan tidak hanya menggunakan data laporan BPBD sehingga hasil pengelompokan kerawanan lebih baik.
2. Dalam penelitian selanjutnya, lebih baik menggunakan menggunakan variabel pendukung yang lebih sedikit variasinya agar menghasilkan clustering yang lebih akurat dan memiliki presentase yang tinggi.
3. Untuk meningkatkan hasil akurasi clustering, dapat ditambahkan metode penentuan centroid awal dan penentuan jumlah klaster awal agar menghasilkan clustering yang lebih akurat.
4. klasterisasi dapat divisualisasikan dalam bentuk geo map dengan ditambahkannya atribut jarak atau location.

**Daftar Pustaka:**

BNPB. (2019). Definisi Bencana: bnpb. Dipetik 2019, dari bnpb: https://bnpb.go.id/definisi-bencana.

Haomiao, Z., Zhihong, D., Yuanqing, X., & Mengyin, F. (2016). A new sampling method in particle filter based on Pearson correlation coefficient.

Indonesia. (2007, April 26). Undang-Undang Nomor 24 Tahun 2007 : Tentang Penanggulangan Bencana. Jakarta.

Rohmawati, N., Sofi, D., & Mohamad, J. (2015). Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa

*.*Shreya, B., Ankit, C., & Somnath, P. (2015). *Empirical Evaluation of K-Means, Bisecting KMeans,Fuzzy C-Means and Genetic K-Means*. Diambil kembali dari https://ieeexplore.ieee.org/document/7443889

Wani, M. A., & Riyaz, R. (2017). *A novel point density based validity index for clustering gene*. 66-84. doi:10.1504/IJDMB.2017.084027