



UNIVERSITAS
MERCU BUANA

**KLASTERISASI DAMPAK BENCANA GEMPA BUMI MENGGUNAKAN
ALGORITMA K-MEANS DI PULAU JAWA**

TUGAS AKHIR

Aan Wahyu
41517120054

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA**

2022



UNIVERSITAS
MERCU BUANA

**KLASTERISASI DAMPAK BENCANA GEMPA BUMI MENGGUNAKAN
ALGORITMA K-MEANS DI PULAU JAWA**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:
Aan Wahyu
41517120054

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA

2022

MERCU BUANA

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

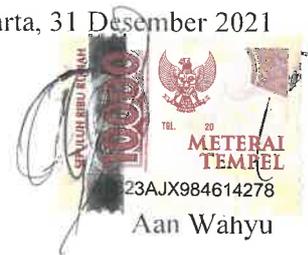
NIM : 41517120054

Nama : Aan Wahyu

Judul Tugas Akhir : Klasterisasi Dampak Bencana Gempa Bumi Menggunakan Algoritma K-Means di Pulau Jawa

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 31 Desember 2021



Aan Wahyu

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Aan Wahyu
NIM : 41517120054
Judul Tugas Akhir : Klasterisasi Dampak Bencana Gempa Bumi Menggunakan Algoritma K-Means di Pulau Jawa

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 31 Desember 2021


1DAJX984614274
Aan Wahyu

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Aan Wahyu
NIM : 41517120054
Judul Tugas Akhir : Klasterisasi Dampak Bencana Gempa Bumi Menggunakan Algoritma K-Means di Pulau Jawa

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis		Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi		Diajukan
		Jurnal Nasional Terakreditasi	Ya	
		Jurnal International Tidak Bereputasi		Diterima
		Jurnal International Bereputasi		
Disubmit/dipublikasikan di :	Nama Jurnal	: Jurnal Edukasi & Penelitian Informatika(JEPIN)		
	ISSN	: 2548-9364		
	Link Jurnal	: https://jurnal.untan.ac.id/index.php/jepin/author/submission/52260		
	Link File Jurnal Jika Sudah di Publish	:		

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Mengetahui
Dosen Pembimbing TA


Rushendra, S.Kom, MT

Jakarta, 09 Februari 2022


Aan Wahyu

LEMBAR PERSETUJUAN PENGUJI

NIM : 41517120054
Nama : Aan Wahyu
Judul Tugas Akhir : Klasterisasi Dampak Bencana Gempa Bumi
Menggunakan Algoritma K-Means di Pulau Jawa

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 8 Februari 2022



UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41517120054
Nama : Aan Wahyu
Judul Tugas Akhir : Klasterisasi Dampak Bencana Gempa Bumi
Menggunakan Algoritma K-Means di Pulau Jawa

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 05 Februari 2022



(Harni Kusniyati, M.Kom)

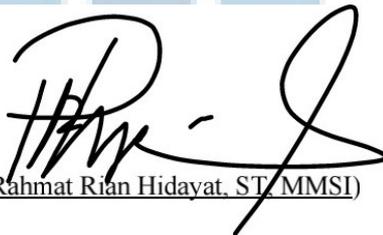
UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41517120054
Nama : Aan Wahyu
Judul Tugas Akhir : Klasterisasi Dampak Bencana Gempa Bumi
Menggunakan Algoritma K-Means di Pulau Jawa

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 04 Februari 2022



(Rahmat Rian Hidayat, ST/MMSI)

UNIVERSITAS
MERCU BUANA

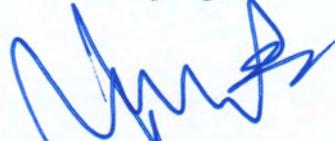
LEMBAR PENGESAHAN

NIM : 41517120054
Nama : Aan Wahyu
Judul Tugas Akhir : Klasterisasi Dampak Bencana Gempa Bumi Menggunakan Algoritma K-Means di Pulau Jawa

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 09 Februari 2022

Menyetujui,



(Rushendra S.Kom, MT)
Dosen Pembimbing

Mengetahui,



(Wawan Gunawan, S.Kom, MT)
Koord. Tugas Akhir Teknik Informatika



(Emil R. Kaburuan, Ph.D.)
Ka. Prodi Teknik Informatika

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa. Atas rahmat dan karunia-Nya, penulis dapat menyelesaikan skripsi dengan judul "Klasterisasi Dampak Bencana Gempa Bumi Menggunakan Algoritma K-Means di Pulau Jawa"

Tugas Akhir ini disusun dan diajukan untuk memenuhi syarat perolehan gelar Sarjana Komputer (S.Kom) pada Fakultas Ilmu Komputer di Universitas Mercu Buana. Disamping itu, penulisan skripsi ini juga bertujuan untuk memberikan pengetahuan kepada pembaca.

Tugas Akhir ini dapat diselesaikan semata karena penulis menerima banyak bantuan dan dukungan. Untuk itu, penulis mengucapkan terima kasih yang tak terhingga kepada:

1. (Bapak) Emil R. Kaburuan, Ph.D selaku Ka. Prodi Teknik Informatika yang telah memberikan pengarahan selama masa perkuliahan.
2. (Bapak) Rushendra, S.Kom, MT selaku Dosen Pembimbing tugas akhir yang telah memberikan arahan terkait materi dan penulisan laporan tugas akhir kepada penulis.
3. (Bapak) Sabar Rudiarto, M.Kom sebagai Dosen Pembimbing MPTI yang juga memberikan perhatian dan waktu untuk membimbing penulis menyelesaikan MPTI.
4. (Bapak) Sabar Rudiarto, M.Kom sebagai Dosen Pembimbing MPTI yang juga memberikan perhatian dan waktu untuk membimbing penulis menyelesaikan MPTI.
5. (Bapak) Hery Derajad Wijaya, S.Kom., MM sebagai Dosen Mata Kuliah MPTI yang sudah membantu membimbing penulis dalam proses perancangan proposal MPTI.
6. (Bapak) Kol. Dr. Adi Subiyanto, S.Si, M.Han yang telah membantu saya dalam proses mendapatkan data statistik gempa terkait penelitian ini, sehingga saya bisa menggunakan data BNPB maupun BMKG.
7. Orang tua yang telah memberikan dukungan serta doa selama proses penyusunan laporan tugas akhir.

8. Feby Dhiandra yang selalu menemani dan telah menjadi *supporting system* yang baik dan tiada henti membantu dalam proses pembuatan TA. Tanpa bantuan dari Feby, naskah jurnal/paper ini tidak mungkin bisa tembus skor 17% saja di Turnitin. Bantuan *paraphrase* nya sangat membantu.
9. Serta berbagai pihak yang tidak mungkin dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa laporan tugas akhir ini masih banyak kekurangan baik bentuk, isi, maupun Teknik penyajiannya. Oleh sebab itu, kritikan yang bersifat membangun dari berbagai pihak penulis terima dengan tangan terbuka dan sangat diharapkan. Semoga kehadiran jurnal tugas akhir ini dapat bermanfaat serta menjadi sumber inspirasi.



Penulis
Aan Wahyu

DAFTAR ISI

<u>HALAMAN SAMPUL</u>	i
<u>HALAMAN JUDUL</u>	i
<u>LEMBAR PERNYATAAN ORISINALITAS</u>	ii
<u>SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR</u> ...	iii
<u>SURAT PERNYATAAN LUARAN TUGAS AKHIR</u>	iv
<u>LEMBAR PERSETUJUAN PENGUJI</u>	v
<u>LEMBAR PENGESAHAN</u>	viii
<u>ABSTRAK</u>	ix
<u>ABSTRACT</u>	x
<u>KATA PENGANTAR</u>	xi
<u>DAFTAR ISI</u>	xiii
<u>NASKAH JURNAL</u>	1
<u>KERTAS KERJA</u>	10
<u>BAB 1. PENDAHULUAN</u>	11
<u>BAB 2. ANALISIS DAN PERANCANGAN</u>	14
<u>BAB 3. SOURCE CODE</u>	16
<u>BAB 4. DATASET</u>	21
<u>BAB 5. TAHAPAN EKSPERIMEN</u>	23
<u>BAB 6. HASIL SEMUA EKSPERIMEN</u>	26
<u>DAFTAR PUSTAKA</u>	29
<u>LAMPIRAN DOKUMEN</u>	31
<u>LAMPIRAN KORESPONDENSI</u>	33

Klasterisasi Dampak Bencana Gempa Bumi Menggunakan Algoritma K-Means di Pulau Jawa

Aan Wahyu ^{#1}, Rushendra ^{*2}

[#] *Fakultas Ilmu Komputer, Universitas Mercu Buana
Jl. Meruya Selatan No. 1 Kembangan Jakarta Barat*

¹41517120054@student.mercubuana.ac.id

² Rushendra@mercubuana.ac.id

Abstrak— Dari semua bencana alam yang terjadi, gempa bumi adalah yang paling sering terjadi di Indonesia. Kajian seismogenetik menunjukkan pulau Jawa merupakan salah satu daerah rawan gempa bumi. Seperti diketahui bahwa pulau Jawa memiliki populasi penduduk yang padat. Dikarenakan alasan tersebut, pemerintah diharuskan untuk memberikan perhatian lebih pada penanggulangan bencana khususnya gempa bumi. Dengan tujuan mengurangi jumlah korban jiwa. Maka dari itu dibutuhkan sebuah algoritma untuk melakukan klasterisasi data dampak gempa bumi dengan tujuan untuk membantu pihak - pihak bersangkutan dalam mengambil keputusan. Penelitian ini melakukan klasterisasi data penyebaran dampak bencana gempa bumi (2012 - 2021) dari Badan Nasional Penanggulangan Bencana (BNPB) menggunakan algoritma K-Means Clustering.

Dari penelitian ini ditemukan bahwa dampak bencana dapat dibagi menjadi 4 klaster. Klaster 1 memiliki dampak bencana paling banyak meliputi meninggal, luka, menderita, mengungsi, kerusakan rumah, fasilitas pendidikan, fasilitas kesehatan, fasilitas ibadah, kantor, jembatan, dan kios. Klaster ini memiliki Mean Absolute Error (MAE) senilai 0,017, Mean Square Error (MSE) senilai 0,002, standar deviasi senilai 0,255 dan variance senilai 0,065. Klaster 2 memiliki dampak bencana paling banyak kedua meliputi meninggal, luka, menderita, mengungsi, rumah, fasilitas pendidikan, fasilitas kesehatan, fasilitas ibadah, pada korban menderita, mengungsi, kerusakan rumah dan fasilitas ibadah. Klaster ini memperoleh MAE senilai 0,053, MSE senilai 0,011, standar deviasi senilai 0,249 dan variance senilai 0,062. Klaster 3 memiliki dampak bencana paling banyak ketiga meliputi korban luka, menderita, mengungsi, kerusakan rumah, fasilitas pendidikan, fasilitas kesehatan, fasilitas ibadah dan kantor. klaster ini memperoleh MAE senilai 0,102, MSE senilai 0,039, standar deviasi senilai 0,212 dan variance senilai 0,045. Klaster 0 memiliki dampak bencana paling sedikit meliputi kerusakan rumah, fasilitas pendidikan, fasilitas kesehatan, fasilitas ibadah, kantor, dan kios. klaster ini memperoleh MAE sebesar 0,021, MSE senilai 0,005, standar deviasi senilai 0,251 dan variance senilai 0,063.

Kata kunci— Bencana, Klaster, Dampak, K-Means, Gempa Bumi

I. PENDAHULUAN

Dari semua bencana alam yang terjadi, gempa bumi adalah yang paling sering terjadi di Indonesia. Beberapa faktor yang mendukung yaitu faktor geologi meliputi jenis tanah dan batuan, struktur dan tekstur tanah, topografi, tektonik, dan pola pengaliran sungai [1]. Lokasi geografis Indonesia terletak di lempeng tektonik yang dikenal dengan sebutan wilayah *ring of fire* (cincin api). Dimana wilayah *ring of fire* terkenal dengan banyaknya gempa bumi yang terjadi disana. Dengan pergerakan lempeng tektonik ini dapat menyebabkan terbentuknya rangkaian gunung api aktif dan jalur gempa bumi. [2].

Berdasarkan kajian seismogenetik, pulau Jawa merupakan daerah rawan gempa dengan sistem satuan seismotektonik busur sangat aktif yang meliputi Jawa Barat dan Sumatra, serta busur aktif bagian Jawa Barat, Jawa Tengah, Jawa Timur [3]. Kekuatan gempa, jarak sumber, kepadatan penduduk, kondisi geologi setempat, dan infrastruktur merupakan hal-hal penentu indeks bencana dan resiko di pulau Jawa. [3]. Berdasarkan pemetaan daerah rawan bencana, terdapat 4 provinsi yang memiliki tingkat kerawanan yang tinggi yaitu Aceh, Jawa Tengah, Jawa Timur, dan Jawa Barat [4]. Jika dilihat dari tingkatan tersebut, pulau Jawa merupakan wilayah yang mendominasi rawan. Seperti diketahui bahwa pulau Jawa memiliki populasi penduduk yang padat. Dikarenakan alasan tersebut, pemerintah diharuskan untuk memberikan perhatian lebih pada penanggulangan bencana khususnya gempa bumi. Dengan tujuan mengurangi jumlah korban jiwa. Dibutuhkan sebuah algoritma untuk melakukan klusterisasi data dampak gempa bumi.

Merujuk pada latar belakang di atas, pulau Jawa dipilih menjadi objek penelitian untuk dianalisis klusterisasi gempa buminya menggunakan algoritma klusterisasi k-means. Cara kerja klusterisasi k-means yaitu dengan memaksimalkan persamaan karakteristik dalam kluster dan memaksimalkan perbedaan antar kluster untuk mendapatkan kelompok data yang diinginkan. Algoritma K-means ini memiliki keunggulan yaitu sederhana dan mudah dipahami, cepat pemrosesannya, mudah penerapannya, dan tersedia di berbagai tools. Dengan begitu diharapkan dengan penelitian ini bisa membantu pihak-pihak bersangkutan dalam mengambil keputusan dalam menanggulangi dampak bencana gempa bumi khususnya di pulau Jawa.

II. PENELITIAN TERDAHULU

Beberapa penelitian yang terkait dengan bencana alam, data mining, dan data science pernah dilakukan. Hasil dari pengelompokan data menggunakan metode k-means dalam penelitian ini menunjukkan bahwa, provinsi Jawa Tengah dan Jawa Barat adalah dua provinsi yang paling rawan dilanda bencana gempa bumi di pulau Jawa [5]. Dari semua teknik data mining, clustering adalah teknik yang dipilih paling efektif dalam mengelompokkan data menjadi 3 kluster dimulai tahun 2013 sampai 2018. Dalam hal ini peneliti memanfaatkan metode pengembangan Waterfall, Google Maps API, dan PHP, serta metode klustering menggunakan Fuzzy Possibilistic C-Means [6]. Pengelompokan episentrum gempa di Provinsi Bengkulu menggunakan k-means clustering dengan euclidean distance. Variabelnya adalah garis lintang, garis bujur dan besaran. Jumlah cluster optimum ditentukan dengan menggunakan indeks Krzanowski dan Lai (KL) yaitu 7 [7]. Klusterisasi gempa di Provinsi Nusa Tenggara Barat menggunakan metode k-means yang bertujuan sebagai acuan terkait mitigasi bencana gempa bumi dan evaluasi wilayah dan tata ruang [8].

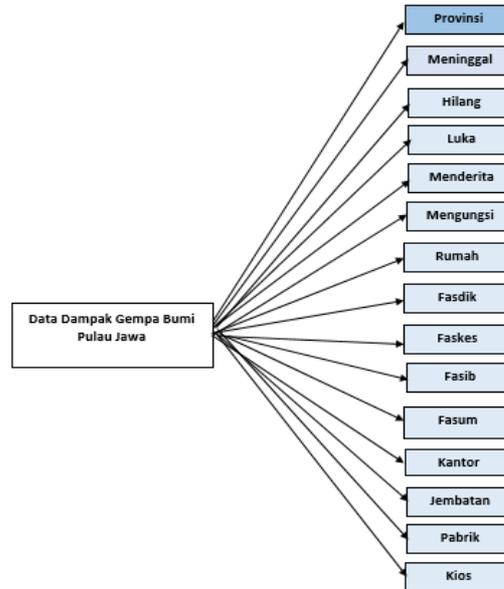
Terdapat dua algoritma yang dapat digunakan dalam proses klasifikasi data dampak bencana gempa bumi yaitu K-means dan k-medoids. Klusterisasi dilakukan berdasarkan kedalaman gempa dan magnitudonya. Dalam penelitian ini K-medoids menjadi pusatnya dan memanfaatkan Clustering Large Application (CLARA) [9]. Penggunaan metode clustering menggunakan data gempa dari seluruh wilayah di Indonesia selama tahun 2014-2018 yang tercatat oleh United State Geological Survey (USGS). Berdasarkan hasil yang didapatkan, ditemukan bahwa k-medoids lebih baik dibandingkan k-means, hal ini karena k-medoids menghasilkan siluet dengan waktu komputasi yang lebih efisien. Untuk hasil cluster k-medoids, nilai silhouette tertinggi adalah 0,4574067 dengan k = 6 [10].

Pengelompokan data gempa bumi yang dibuat secara otomatis untuk dianalisis dan mengembangkan metode klusterisasi otomatis yang tanpa memberikan parameter manual [11]. Pengembangan strategi untuk memilih standar gerakan tanah yang digunakan untuk transportasi Pusat Penelitian Teknik Gempa Pasifik [12]. Algoritma Geo K-means dengan memasukan informasi kontekstual dalam proses pengelompokan. Pendekatan ini untuk diagnosis pasca-gempa yang lebih akurat tentang perilaku geodinamika daerah yang terkena dampak [13]. Pengelompokan produk menggunakan metode k-means clustering untuk merancang ukuran kotak dan menemukan lapisan paling banyak yang dapat ditumpuk pada palet [14]. K-means sebagai sumber titik yang dapat digunakan untuk berpartisipasi katalog hiposenter gempa dan sebagai pengembangan sumber baris baru dari algoritma yang sesuai dalam seismologi [15].

III. METODOLOGI

A. Pengumpulan Dataset

Dataset yang digunakan adalah data persebaran dampak bencana gempa bumi (2012 – 2021) dari sumber gis.bnpb.go.id dan dbi.bnpb.go.id [16]. Gambar 1 menunjukkan dataset persebaran dampak gempa bumi memiliki 15 parameter : meninggal, hilang, luka, menderita, mengungsi, rumah, fasilitas pendidikan (fasdik), fasilitas kesehatan (faskes), fasilitas ibadah (fasib), fasilitas umum (fasum), kantor, jembatan, pabrik, dan kios. Dataset yang belum dinormalisasi dengan total seluruh data 147 data atau baris. Tabel 1 merupakan contoh dataset yang digunakan.



Gambar 1 Dataset Dampak Gempa

Table 1 Contoh Dataset

prov	meninggal	hilang	luka	menderita	mengungsi	rumah	fasdik	faskes	fasib	fasum	kantor	jembatan	pabrik	kios
DKI JAKARTA	0	0	1	0	0	0	0	0	0	0	3	0	0	0
JAWA BARAT	1	0	0	0	285	1,847	15	0	25	0	2	0	0	0
BANTEN	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JAWA TIMUR	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JAWA TENGAH	0	0	1	0	0	0	0	0	0	0	0	0	0	0
JAWA TIMUR	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DI YOGYAKARTA	1	0	0	0	0	0	0	0	0	0	0	0	0	0
DI YOGYAKARTA	0	0	10	0	0	1	0	0	0	0	0	0	0	0
JAWA BARAT	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JAWA TIMUR	0	0	1	0	0	77	1	135	2	0	145	0	0	0

Dari informasi dataset didapatkan data masih sangat acak, tidakurut berdasarkan tanggal, masih terdapat data float, dan masih terdapat beberapa parameter yang tidak digunakan dan tidak diperlukan dalam proses eksperimen dan pembuatan model namun tidak ada data yang bernilai null atau hilang.

Dataset ini merupakan gambaran dari beberapa provinsi di pulau Jawa, Dengan jumlah terbanyak berasal dari Jawa Barat sebanyak 58 dan terbanyak kedua yaitu berasal dari Jawa Timur. Table 2 menunjukkan jumlah tiap provinsi yang memiliki catatan gempa bumi dari tahun 2012 hingga tahun 2021.

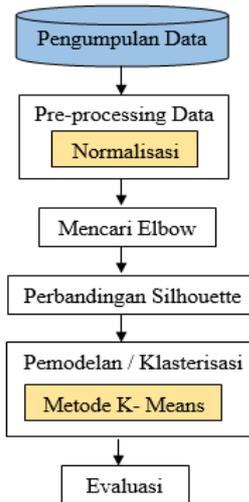
Table 2 Data Jumlah Tiap Provinsi

Provinsi	Jumlah
Jawa Barat	58
Jawa Timur	48

Jawa Tengah	22
Banten	11
DI Yogyakarta	7
DKI Jakarta	1

B. Metode Penelitian

Gambar 1 menunjukkan cara kerja sistem atau alur eksperimen penelitian, terlihat bahwa terdapat beberapa tahapan system sebagai berikut :



Gambar 2 Metode Penelitian

Langkah pertama adalah mengumpulkan Data Pesebaran Dampak Bencana Gempa Bumi (2012 – 2021) yang didapat dari BNPB (Badan Nasional Penanggulangan Bencana) Republik Indonesia. Namun data ini masih mentah, sehingga perlu dilakukan preprocessing berupa uji data hilang dan normalisasi untuk mempersiapkan data sebelum digunakan untuk clustering. Sebelum ke tahap pemodelan, dilakukan pencarian kluster terbaik menggunakan metode elbow, serta melakukan perbandingan silhouette menggunakan PCA (*Principal Component Analysis*) dan tanpa PCA. Tahap selanjutnya adalah tahap pemodelan klusterisasi menggunakan k-means untuk mencari pola sebaran dampak bencana dan kemudian di evaluasi menggunakan MSE, MAE, Standar Deviasi dan Analysis of Variance. Berikut penjelasan tiap tahap :

1. Pre-Processing

Pada tahap pre-processing dilakukan pengujian missing value dan normalisasi. Pengujian missing value berguna untuk menghindari error dengan cara menghilangkan data yang kosong atau null. Pada tahap pre-processing data yang digunakan sebelum pre-processing sebanyak 147 data atau baris, kemudian setelah dilakukan pre-processing data masih sama sebanyak 147 data atau rows, hal ini menunjukkan bahwa data tidak ada yang kosong atau bernilai null. Setelah itu melakukan tahap normalisasi yang bertujuan untuk mengubah data agar memiliki distribusi yang serupa. Pada kasus ini menerapkan metode min – max scaler yang bekerja dengan menyesuaikan data dalam rentang tertentu yaitu mulai dari 0 hingga 1. Rumus yang digunakan untuk normalisasi adalah sebagai berikut:

$$x' = \frac{(x - \min_{[i]}(x))}{(\max_{[i]}(x) - \min_{[i]}(x))} \quad (3)$$

x' = Nilai dari masing – masing fitur

$\min_{[i]} x$ = Nilai terkecil pada fitur

$\lceil \max \rceil _x =$ Nilai terbesar pada fitur

Sebagai contoh data nilai 15 dengan nilai fitur 15, 20, 25, 30 yang akan dinormalisasi dengan contoh sebagai berikut :

$$\begin{aligned}x' &= \frac{15 - 15}{30 - 15} \\x' &= \frac{0}{15} \\x' &= 0\end{aligned}$$

Dari hasil tersebut dapat disimpulkan nilai sebelum dinormalisasi 15 dan telah dinormalisasi menjadi 0.

2. Mencari Elbow

Pada tahap ini akan menentukan jumlah kluster terbaik menggunakan metode Elbow. Metode ini bekerja paling baik dalam menentukan nilai k yang optimal. Dalam metode ini, sebenarnya akan memvariasikan jumlah kluster (k) dari 1-10. Kemudian menghitung jumlah sebuah kluster atau Within- Cluster Sum of Square (WCSS) untuk setiap nilai K. Grafik akan bergerak cepat pada suatu titik sehingga menciptakan bentuk siku. Kemudian grafik akan bergerak mendekati atau sejajar sumbu x, jumlah kluster atau nilai k yang optimal akan menyesuaikan titik ini.

3. Silhouette Coefficient

Pada tahap ini akan dilakukan perbandingan metode Silhouette Score menggunakan Principal Component Analysis (PCA) dan tanpa PCA. Silhouette Coefficient digunakan untuk mengevaluasi kluster dan melihat kualitas penempatan data dalam sebuah kluster. Silhouette Coefficient adalah contoh evaluasi yang dilakukan jika kebenaran dasar tidak diketahui, di mana model kluster yang lebih jelas ditentukan oleh skor silhouette coefficient yang lebih tinggi. Silhouette Coefficient terbagi menjadi dua skor sebagai berikut [17] :

- a : Jarak rata – rata antara sampel dan semua titik lain di cluster yang sama.
- b : Jarak rata – rata antara sampel dan semua titik lain di cluster terdekat berikutnya.

Silhouette Coefficient s untuk sampel tunggal diberikan sebagai berikut [17] :

$$s = \frac{b - a}{\max(a, b)} \quad (2)$$

4. Pemodelan / Klasterisasi

Pada penelitian ini akan dilakukan pengukuran jarak Euclidean untuk algoritma k-means. Jenis pengukuran jarak ini biasa digunakan dalam data mining. Dalam jarak euclidean, jarak antara dua titik yang didefinisikan sebagai garis lurus. Jarak Euclidean dihitung dengan [18]:

$$DEuclidean(x_i, x_j) = \sqrt{(x_{im} - x_{jm})^2} = \sqrt{\sum_{m=1}^n (x_{im} - x_{jm})^2} \quad (1)$$

$DEuclidean(x_i, x_j)$: Euclidean Distance

x_i : Data -i

x_j : Data -j

x_{im} : Data -i atribut -m

x_{jm} : Data -j atribut -m

Hasil dari pengujian euclidean distance berupa kluster yang sudah built-in ke dalam algoritma k-means dan dimanifestasikan dalam bentuk data centroid. Algoritma *K-means* bertujuan untuk menjadikan nilai rata-rata dari suatu kluster sebagai centroid dari k tersebut [19]. Grup metode pengelompokan K-Means data berdasarkan kedekatan mereka satu sama lain sesuai dengan jarak Euclidean.

K-means analisis kluster adalah proses iteratif karena merupakan algoritma partisi keras [7]. Pertama, data awalnya dipartisi. Setiap grup dihitung meannya dan kemudian data dipartisi lagi dengan mengalokasikan setiap data ke mean terdekatnya posisi kluster. Dalam bentuknya yang paling sederhana, algoritma ini terdiri dari langkah-langkah berikut [19] :

1. Tentukan jumlah k-cluster (secara acak)
2. Menghasilkan nilai acak untuk cluster centroid sebanyak k-cluster.
3. Hitung menggunakan Euclidean Distance untuk mengetahui jarak data input ke setiap centroid.
4. Klasifikasi data yang memiliki jarak terdekat terhadap centroid.
5. Perbarui nilai pusat atau centroid dari rata – rata cluster.
6. Melakukan iterasi atau pengulangan sampai tidak ada yang berubah.

Setelah itu gunakan nilai akhir rata- rata pusat cluster tersebut sebagai parameter untuk menentukan klasifikasi data.

5. Evaluasi

Setelah data diklasifikasikan kemudian di evaluasi. Pada tahap evaluasi, model akan diukur performanya. Pemodelan akan diukur menggunakan MSE (Main Squared Error), MAE (Mean Absolute Error), Standar Deviasi, Analysis Variance. Receiver Operating Characteristic Curve.

Nilai tengah galat absolut (mean absolute error)

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (4)$$

Nilai tengah galat kuadrat (mean squared error)

$$MSE = \frac{1}{n} \sum_{i=1}^n e_i^2 \quad (5)$$

Deviasi standar galat (standard deviation of error)

$$SDE = \sqrt{\sum_{i=1}^n e_i^2 / (n-1)} \quad (6)$$

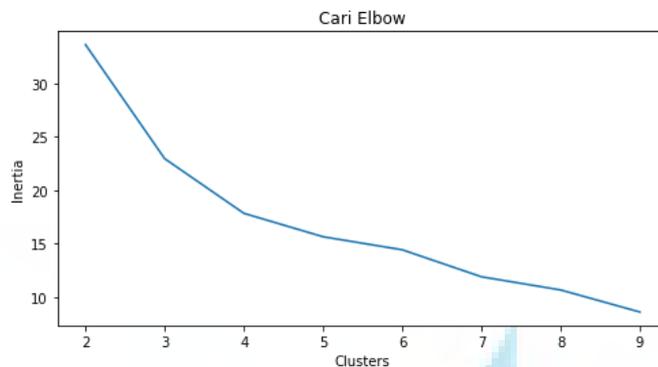
Analysis Variance

$$SDE = \sqrt{\sum_{i=1}^n e_i^2 / (n-1)} \quad (7)$$

IV. HASIL DAN PEMBAHASAN

Pada pengujian kluster terbaik menggunakan silhouette membandingkan metode menggunakan PCA (Principal Component Analysis) dan tanpa PCA yang sudah didekomposisi, dengan menggunakan PCA memperoleh hasil 67 %, sedangkan tanpa PCA memperoleh 70%. Hal ini

membuktikan parameter terbaik yaitu tanpa menggunakan PCA. Sedangkan dalam mencari jumlah cluster terbaik menggunakan metode elbow. Gambar 2 merupakan grafik elbow yang menunjukkan hasil cluster terbaik berjumlah 4.



Gambar 3 Grafik Elbow

Berdasarkan hasil klusterisasi menggunakan K-Means, Klusterisasi terbagi menjadi 4 kluster yaitu kluster 0, 1, 2, dan 3. Pada table menampilkan jumlah kerusakan dari tiap – tiap kluster.

Table 3 Heat Map Klusterisasi

	meninggal	hilang	luka	menderita	mengungsi	rumah	fasdik	faskes	fasib	fasum	kantor	jembatan	pabrik	kios
Cluster 0	1	0	2	0	0	9814	53	1	23	0	9	0	0	3
Cluster 1	12	0	136	1059348	361624	9927	36	14	128	0	35	3	0	4
Cluster 2	12	0	148	6915	2415093	107339	207	162	282	0	201	0	0	1
Cluster 3	5	0	3	2000	109	712	9	0	8	0	0	0	0	0

Berdasarkan table 2 kluster tinggi atau cluster 1 merupakan data yang memiliki dampak bencana paling banyak meliputi meninggal, luka, menderita, mengungsi, kerusakan rumah, fasilitas pendidikan, fasilitas kesehatan, fasilitas ibadah, kantor, jembatan, dan kios. Berdasarkan data yang diperoleh untuk kluster ini menghasilkan korban menderita paling parah sebanyak 1,059,348 orang.

Kluster tengah atas atau cluster 2 merupakan data yang memiliki dampak bencana paling banyak kedua meliputi meninggal, luka, menderita, mengungsi, rumah, fasilitas pendidikan, fasilitas kesehatan, fasilitas ibadah, pada korban menderita, mengungsi, kerusakan rumah dan fasilitas ibadah. Kluster tengah bawah atau cluster 3 merupakan data yang memiliki dampak bencana pada korban luka, menderita, mengungsi, kerusakan rumah, fasilitas pendidikan, fasilitas kesehatan, fasilitas ibadah dan kantor dengan data. Kluster bawah atau cluster 0 merupakan data yang memiliki dampak bencana paling sedikit pada kerusakan rumah, fasilitas pendidikan, fasilitas kesehatan, fasilitas ibadah, kantor, dan kios.

Table 3 menunjukkan centroid setiap cluster sebagai berikut :

Table 4 Centroid

	Cluster 0	Cluster 1	Cluster 2	Cluster 3
meninggal	125661,336	531772,356	1131164,72	5526394,51
hilang	0.00000000e+00	0.00000000e+00	0.00000000e+00	0.00000000e+00
luka	219483,739	2476397,95	2262329,44	17507039,2
menderita	91911793,8	1933230,14	0.00000000e+00	69211,7282
mengungsi	5899031,08	1651931,14	6,93889E-10	19949658,5
rumah	28312992,9	97944647,5	19245914,7	7141463,46
fasdik	176443,515	1556005,28	88249317,8	6573765,26
faskes	-1,73472E-10	317272,139	282114,822	5379640,49
fasib	272679,813	2705308,94	11582940,3	16634798,6
fasum	0.00000000e+00	0.00000000e+00	0.00000000e+00	0.00000000e+00
kantor	-6,93889E-10	837315,923	4388925,74	16653927,6
jembatan	0.00000000e+00	4881,39048	0.00000000e+00	0.00000000e+00
pabrik	0.00000000e+00	0.00000000e+00	0.00000000e+00	0.00000000e+00
kios	1,73472E-10	245979,903	7407310,78	5930,17354

Table 4 merupakan evaluasi pada tiap cluster, berdasarkan hasil evaluasi dapat diperoleh bahwa cluster 1 memiliki nilai MAE dan MSE terendah yaitu 0,017 dan 0,002, hal ini menunjukkan bahwa semakin rendah nilai error yang terjadi maka semakin baik cluster tersebut begitu pula juga sebaliknya. Sedangkan cluster yang memiliki standar deviasi dan variance terbaik yaitu cluster 3 memperoleh standar deviasi 0,212 dan variance 0,045.

Cluster yang memiliki MAE dan MSE yang paling tinggi diperoleh cluster 3 yaitu MAE 0,102 dan MSE 0,039. Sedangkan Standar deviasi dan variance tertinggi diperoleh oleh cluster 1 dengan standar deviasi 0,255 dan variance 0,065.

Table 5 Evaluasi Klaster

	MAE	MSE	Stdev	Variance
Cluster 0	0,021	0,005	0,251	0,063
Cluster 1	0,017	0,002	0,255	0,065
Cluster 2	0,053	0,011	0,249	0,062
Cluster 3	0,102	0,039	0,212	0,045

V. KESIMPULAN

Setelah dilakukan pengujian dan analisis pada data persebaran dampak bencana gempa bumi (2012 – 2021) menggunakan algoritma k-means diperoleh hasil bahwa metode k-means bekerja dengan baik tanpa menggunakan (Principal Component Analysis). Klusterisasi menggunakan elbow menghasilkan 4 cluster yaitu cluster tinggi, cluster menengah atas, cluster menengah bawah dan cluster bawah. Setiap cluster memiliki evaluasi yang berbeda, cluster bawah memiliki nilai MAE dan MSE terbaik yaitu 0,017 dan 0,02 sedangkan cluster yang memiliki standar deviasi dan variance terbaik yaitu cluster menengah bawah memperoleh standar deviasi 0,212 dan variance 0,045.

REFERENSI

- [1] "BENCANA GEOLOGI, (Seri Pengetahuan Bencana)." <https://sumbarprov.go.id/home/news/8753-bencana-geologi-seri-pengetahuan-bencana> (accessed Dec. 31, 2021).
- [2] "RBI RISIKO BENCANA INDONESIA B N P B."
- [3] A. S. Pusat, S. Geologi, B. Geologi, J. Diponegoro, N. 57, and B. Sari, "Seismotektonik dan Potensi Kegempaan Wilayah Jawa."
- [4] 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, doi: 10.14257/IJGDC.2018.11.8.08.
- [5] A. M. Siregar, “PENERAPAN ALGORITMA K-MEANS UNTUK PENGELOMPOKAN DAERAH RAWAN BENCANA DI INDONESIA,” *Intern. (Information Syst. Journal)*, vol. 1, no. 2, pp. 1–10, Feb. 2019, doi: 10.32627/INTERNAL.V1I2.42.
- [6] S. Putriana and D. Andreswari, “Clustering Data Titik Gempa Dengan,” vol. 9, no. 1, 2021.
- [7] P. Novianti, D. Setyorini, and U. Rafflesia, “K-means cluster analysis in earthquake epicenter clustering,” *Int. J. Adv. Intell. Informatics*, vol. 3, no. 2, pp. 81–89, Jul. 2017, doi: 10.26555/IJAIN.V3I2.100.
- [8] H. J. Wattimanela, “Grouping of Tectonic Earthquakes in the Province of Nusa Tenggara Barat Indonesia with K-Means Cluster Method Approach and Determination of Distribution Type,” *Sci. Nat.*, vol. 2, no. 3, pp. 177–191, Sep. 2019, doi: 10.30598/SNVOL2ISS3PP177-191YEAR2019.
- [9] I. H. Rifa, H. Pratiwi, and R. Respatiwan, “CLUSTERING OF EARTHQUAKE RISK IN INDONESIA USING K-MEDOID AND K-MEANS ALGORITHMS,” *MEDIA Stat.*, vol. 13, no. 2, pp. 194–205, Dec. 2020, doi: 10.14710/MEDSTAT.13.2.194-205.
- [10] F. Ramadhani, S. #1, I. #2, and F. Nhita, “Clustering of Earthquake Prone Areas in Indonesia Using K-Medoids Algorithm,” *Indones. J. Comput.*, vol. 4, no. 3, pp. 65–76, 2019, doi: 10.34818/INDOJC.2019.4.3.359.
- [11] “EFFECTS OF EARTHQUAKE DATA CLUSTERING ON THE RESULTS OF STRESS INVERSIONS | Request PDF.”
https://www.researchgate.net/publication/332465791_EFFECTS_OF_EARTHQUAKE_DATA_CLUSTERING_ON_THE_RESULTS_OF_STRESS_INVERSIONS (accessed Jan. 04, 2022).
- [12] J. W. Baker, T. Lin, S. K. Shahi, and N. Jayaram, “New Ground Motion Selection Procedures and Selected Motions for the PEER Transportation Research Program,” 2011.
- [13] F. Mato and T. Toulkeridis, “An unsupervised K-means based clustering method for geophysical post-earthquake diagnosis,” *2017 IEEE Symp. Ser. Comput. Intell. SSCI 2017 - Proc.*, vol. 2018-January, pp. 1–8, Feb. 2018, doi: 10.1109/SSCI.2017.8285216.
- [14] S. Plungsri and K. Puntusavase, “K-means Clustering for Grouping Product Size for Reducing Cost Packaging:,” *J. Ind. Technol.*, vol. 16, no. 2, pp. 30–44, Jul. 2020, doi: 10.14416/J.IND.TECH.2020.07.003.
- [15] G. Weatherill and P. W. Burton, “Delineation of shallow seismic source zones using K-means cluster analysis, with application to the Aegean region,” *Geophys. J. Int.*, vol. 176, no. 2, pp. 565–588, Feb. 2009, doi: 10.1111/J.1365-246X.2008.03997.X/3/176-2-565-FIG016.JPEG.
- [16] “Geoportal Kebencanaan Indonesia.” <https://gis.bnpb.go.id/#tabel> (accessed Feb. 01, 2022).
- [17] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *J. Comput. Appl. Math.*, vol. 20, no. C, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.
- [18] “Learning distance functions.”
- [19] J. Han, M. Kamber, and J. Pei, “Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems),” 2011.

UNIVERSITAS
 MERCU BUANA

KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal dengan judul di atas. Di dalam kertas kerja ini disajikan:

1. Literature Review
2. Analisis dan Perancangan
3. Source Code
4. Dataset yang digunakan
5. Tahapan Eksperimen
6. Hasil Eksperimen

