

**PENERAPAN *PRINCIPAL COMPONENT ANALYSIS* BERBASIS *CLUSTERING*
DATA GEMPA UNTUK MELABELKAN TINGKAT GUNCANGAN**

SKRIPSI

Oleh:
RENATA AMALIA PUTRI
NIM. 220605110074



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026**

**PENERAPAN *PRINCIPAL COMPONENT ANALYSIS* BERBASIS
CLUSTERING DATA GEMPA UNTUK MELABELKAN
TINGKAT GUNCANGAN**

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
RENATA AMALIA PUTRI
NIM. 220605110074

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026**

HALAMAN PERSETUJUAN

PENERAPAN *PRINCIPAL COMPONENT ANALYSIS* BERBASIS *CLUSTERING* DATA GEMPA UNTUK MELABELKAN TINGKAT GUNCANGAN

SKRIPSI

Oleh:
RENATA AMALIA PUTRI
NIM. 220605110074

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 24 April 2026

Pembimbing I,



Dr. Agung Teguh Wibowo Almais, M.T
NIP. 19860301 202321 1 016

Pembimbing II,



Supriyono, M.Kom
NIP. 19841010 201903 1 012

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Supriyono, M.Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

PENERAPAN *PRINCIPAL COMPONENT ANALYSIS* BERBASIS *CLUSTERING* DATA GEMPA UNTUK MELABELKAN TINGKAT GUNCANGAN

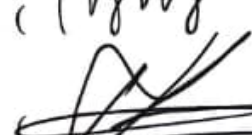
SKRIPSI

Oleh:
RENATA AMALIA PUTRI
NIM. 220605110074

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 25 Mei 2026

Susunan Dewan Penguji

Ketua Penguji	: <u>Dr. Ririen Kusumawati, S.Si, M.Kom</u> NIP. 19720309 200501 2 002
Anggota Penguji I	: <u>Johan Ericka Wahyu Prakasa, M.Kom</u> NIP. 19831213 201903 1 004
Anggota Penguji II	: <u>Dr. Agung Teguh Wibowo Almais, M.T</u> NIP. 19860301 202321 1 016
Anggota Penguji III	: <u>Supriyono, M.Kom</u> NIP. 19841010 201903 1 012

()
()
()
()

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Supriyono, M.Kom
NIP. 19841010 201903 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Renata Amalia Putri
NIM : 220605110074
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Penerapan *Principal Component Analysis* berbasis *Clustering* Data Gempa untuk Melabelkan Tingkat Guncangan

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 25 Mei 2026
Yang membuat pernyataan,



Renata Amalia Putri
NIM. 220605110074

MOTTO

“Mulailah dengan niat, jalani dengan ikhlas, dan akhiri dengan tuntas”

*“By believing passionately in something that does not yet exist, we create it”
(Franz Kafka)*

HALAMAN PERSEMBAHAN

Segala puji bagi Allah Subhanahu Wa Ta'ala atas rahmat, karunia, dan hidayah-Nya, sehingga penulis dapat menuntaskan skripsi ini.

Dengan penuh rasa syukur, penulis persembahkan kepada:

Orang tua saya,

Yang senantiasa memberikan doa, semangat, dan kasih sayang yang tiada henti dan tak terhitung nilainya.

Kakak dan Adik saya,

Yang dengan candaan dan doa yang menghangati perjalanan penulis untuk tidak akan berhenti ditengah jalan.

Seluruh keluarga besar,

Yang selalu menjadi penguat, dan memberikan dukungan serta doa dalam setiap proses yang dilewati.

Dosen pembimbing dan penguji,

Atas ilmu dan bimbingan yang diberikan selama proses penyusunan skripsi.

Teman-teman seperjuangan,

Atas kebersamaan, tenaga, dan waktu selama masa perkuliahan.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dengan menyebut nama Allah SWT. yang Maha Pengasih lagi Maha Penyayang, segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Penerapan *Principal Component Analysis* berbasis *Clustering* Data Gempa untuk Melabelkan Tingkat Guncangan” dengan baik. Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad SAW, yang telah membawa umat manusia dari zaman kegelapan menuju zaman yang terang benderang, yakni agama Islam, semoga kita termasuk golongan yang mendapatkan syafaat beliau di hari kiamat kelak. Aamiin.

Penyusunan skripsi ini tidak terlepas dari bantuan, dukungan, serta doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M.Si., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Agus Mulyono, M.Kes., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Supriyono, M.Kom., selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Agung Teguh Wibowo Almais, M.T., selaku dosen pembimbing I yang telah memberikan bimbingan dengan penuh kesabaran, ketulusan, dan perhatian kepada penulis. Terima kasih atas segala arahan, masukan,

motivasi, serta dukungan yang sangat berarti dalam proses penyusunan hingga penyelesaian skripsi ini.

5. Supriyono, M.Kom., selaku dosen pembimbing II yang telah memberikan arahan, masukan, dan bimbingan kepada penulis dalam proses penyusunan skripsi ini.
6. Dr. Ririen Kusumawati, S.Si., M.Kom., selaku Penguji I dan Johan Ericka Wahyu Prakasa, M.Kom selaku Penguji II yang telah memberikan kritik, saran, dan masukan yang membangun dalam pengujian serta penyempurnaan skripsi ini.
7. Segenap dosen, laboran, dan staf Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah memberikan ilmu, dukungan, serta bantuan kepada penulis selama masa studi hingga selesainya skripsi ini
8. Badan Pengurus Harian (BPH) HIMATIF “Encoder” yang beranggotakan Kak Reyhan, Kak Gianda, Kak Aisha, Kak Lidya, dan Atsila, yang telah berjasa atas pengalaman yang sangat berkesan selama menjabat sebagai Sekretaris II Periode 2023/2024. Khususnya Atsila yang selalu menemani penulis dari awal dan selalu hadir menemani penulis di saat dibutuhkan.
9. Teman-teman bimbingan sekaligus magang, yaitu, Fina, Fadyla, Intan, Faiza, dan Arum, yang telah memberikan dukungan dan kebersamaan selama proses penyusunan skripsi ini.

10. Teman-teman magang di KPKNL Malang, yaitu Dien dan Ita, yang telah menemani penulis selama magang hingga jadi teman yang saling dekat sampai sekarang.
11. Untuk penulis sendiri yang sudah berjuang hingga titik ini dan memilih tidak menyerah hingga akhirnya mampu menyelesaikan skripsi ini.

Penulis sadar bahwa skripsi ini masih jauh dari kesempurnaan, oleh karena itu penulis mengharapkan masukan, kritik, dan saran membangun dari pembaca untuk perbaikan di masa mendatang. Semoga skripsi ini dapat bermanfaat bagi rekan-rekan mahasiswa dan pembaca sekalian serta dapat memberikan sumbangan positif dalam pengembangan ilmu pengetahuan.

Wassalamu'alaikum Wr, Wb.

Malang, 25 Mei 2026

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN.....	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
ABSTRAK	xv
ABSTRACT	xvi
مستخلص البحث.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Pernyataan Masalah	6
1.3 Batasan Masalah.....	6
1.4 Tujuan Penelitian.....	7
1.5 Manfaat Penelitian	7
BAB II STUDI PUSTAKA	8
2.1 Penelitian Terkait.....	8
2.2 Gempa Bumi	12
2.3 <i>Principal Component Analysis</i> (PCA)	13
2.4 <i>PCA Clustering</i>	16
BAB III DESAIN PENELITIAN.....	19
3.1 Desain Sistem.....	19
3.2 Pengumpulan Data	22
3.3 Evaluasi Model.....	23
3.3.1 <i>Calinski-Harabasz Index</i> (CHI).....	23
3.3.2 Validasi Eksternal Data.....	25
3.4 <i>Pseudocode</i> Menggunakan Metode <i>PCA Clustering</i>	25
BAB IV HASIL DAN PEMBAHASAN.....	27
4.1 <i>Pre-processing Data</i>	27
4.1.1 Teknik Pengumpulan Data.....	27
4.2 Normalisasi Data.....	28
4.2.1 <i>StandardScale</i>	28
4.3 <i>Principal Component Analysis</i>	29
4.3.1 Menentukan Nilai Eigen dan Rasio Varians	30
4.3.2 Menentukan Komponen Utama	32
4.4 Visualisasi Grafik	36
4.4.1 Visualisasi Grafik <i>Scatter Plot</i>	37
4.4.2 Visualisasi <i>Bar Plot</i> Nilai Eigen dan Rasio Varians	40
4.4.3 Visualisasi <i>Cluster</i> berdasarkan Label.....	42
4.5 Evaluasi Model.....	43

4.6 Integrasi Islam.....	44
4.7 Hasil	48
BAB V KESIMPULAN DAN SARAN	51
5.1 Kesimpulan	51
5.2 Saran.....	52
DAFTAR PUSTAKA	

DAFTAR GAMBAR

Gambar 3.1 Alur Desain Sistem.....	19
Gambar 4.1 Grafik <i>Scatter Plot</i> Hasil PCA	38
Gambar 4.2 Grafik <i>Bar Plot</i> Nilai Eigen & Rasio Varians	41
Gambar 4.3 Visualisasi <i>Cluster</i> berdasarkan Label	42

DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait.....	9
Tabel 3.1 Data Gempa Bumi	22
Tabel 3.2 <i>Pseudocode Principal Component Analysis</i>	26
Tabel 4.1 Data Gempa Bumi	27
Tabel 4.2 Hasil Normalisasi Data.....	29
Tabel 4.3 Hasil Nilai Eigen dan Rasio Varians	31
Tabel 4.4 Hasil Faktor <i>Loading</i> Variabel	32
Tabel 4.5 Rincian Hasil Komponen Utama (PC)	34
Tabel 4.6 Hasil Pelabelan	49

ABSTRAK

Putri, Renata Amalia. 2026. **Penerapan *Principal Component Analysis* berbasis *Clustering* Data Gempa untuk Melabelkan Tingkat Guncangan**. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Agung Teguh Wibowo Almais, M.T. (II) Supriyono, M.Kom.

Kata kunci: *Clustering*, *Principal Component Analysis*, Data Gempa Bumi

Indonesia memiliki tingkat kerentanan tinggi terhadap bencana alam, khususnya gempa bumi. Perbedaan tingkat guncangan gempa memerlukan upaya mitigasi berbasis teknologi untuk mendukung analisis data gempa. Penelitian ini bertujuan mengelompokkan data gempa menggunakan metode *Principal Component Analysis* (PCA) berbasis *clustering*. Data yang digunakan berupa data gempa harian tahun 2024 dari Badan Meteorologi, Klimatologi, dan Geofisika Stasiun Geofisika Kelas III Malang. PCA digunakan untuk mengekstraksi komponen utama yang merepresentasikan variasi dominan data. Hasil analisis menunjukkan bahwa komponen utama pertama (PC1) mampu menjelaskan 34.1% variasi data dan didominasi oleh variabel *longitude* serta magnitudo. Nilai PC1 digunakan sebagai dasar proses *clustering* untuk mengidentifikasi pola gempa berdasarkan kemiripan karakteristik. Pengujian *Calinski Harabasz Index* menghasilkan nilai sebesar 5626.0269. Berdasarkan hasil tersebut, data gempa dikelompokkan menjadi kategori dirasakan dan tidak dirasakan mengacu pada Skala Intensitas Gempabumi (SIG-BMKG). Hasil penelitian menunjukkan bahwa metode PCA mampu mengurangi kompleksitas data dan membantu pengelompokan tingkat guncangan gempa secara efektif dan efisien.

ABSTRACT

Putri, Renata Amalia. 2026. **Application of Clustering-Based Principal Component Analysis for Labeling Earthquake Ground Shaking Levels**. Thesis. Department of Informatics Engineering Faculty of Science and Technology Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisors: (I) Dr. Agung Teguh Wibowo Almais, M.T. (II) Supriyono, M.Kom.

Keywords: Clustering, Principal Component Analysis, Earthquake Data.

Indonesia is highly vulnerable to natural disasters, particularly earthquakes. This study aims to classify earthquake data using a clustering-based Principal Component Analysis (PCA) method. The *Dataset* consists of daily earthquake records in 2024 from the Badan Meteorologi, Klimatologi, dan Geofisika Geophysics Station Class III Malang. PCA was used to extract the main components representing dominant data variations. The results showed that the first principal component (PC1) explained 34.1% of the variance and was dominated by longitude and magnitude variables. PC1 values were then used for clustering based on earthquake characteristics similarity. The Calinski-Harabasz Index produced a value of 5626.0269. Based on the clustering results, earthquakes were categorized into felt and unfelt groups according to the Earthquake Intensity Scale (SIG-BMKG). The findings demonstrate that PCA effectively reduces data complexity and supports efficient earthquake shaking classification.

مستخلص البحث

بوتري، ريناتا أماليا. ٢٠٢٦. تطبيق تحليل المكونات الرئيسية القائم على التجميع لتصنيف مستويات الاهتزاز الأرضي. قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: الدكتور أغونغ تيغوه ويوو ألاميس، ماجستير تقني المشرف الثاني: سوبريونو، الماجستير.

الكلمات الرئيسية: التجميع، تحليل المكونات الرئيسية، بيانات الزلازل.

تُعدّ إندونيسيا من الدول الأكثر عرضةً للكوارث الطبيعية، ولا سيما الزلازل. تهدف هذه الدراسة إلى تصنيف بيانات القائمة على التجميع. تتكون مجموعة البيانات من سجلات الزلازل (PCA) الزلازل باستخدام منهجية تحليل المكونات الرئيسية اليومية لعام ٢٠٢٤ الصادرة عن محطة الجيوفيزياء من الفئة الثالثة التابعة للهيئة الإندونيسية للأرصاد الجوية وعلم المناخ والجيوفيزياء. في مالانج (BMKG). استُخدم تحليل المكونات الرئيسية لاستخلاص المكونات الرئيسة التي تمثل التباينات السائدة في البيانات. فسّر ما نسبته ٣٤,١٪ من إجمالي التباين، وكان متأثرًا بشكل رئيس بمتغيري خط (PC1) وأظهرت النتائج أن المكون الرئيسي الأول الطول ومقدار الزلازل. ثم استُخدمت قيم المكون الرئيسي الأول في عملية التجميع اعتمادًا على تشابه خصائص الزلازل. وقد أسفر عن قيمة بلغت ٥٦٢٦,٠٢٦٩. وبناءً على نتائج التجميع، صُنِّفت (Calinski-Harabasz Index) مؤشر كالينسكي-هاراباز الزلازل إلى مجموعتين: زلازل محسوسة وزلازل غير محسوسة، وذلك وفقًا لمقياس شدة الزلازل المعتمد من الهيئة الإندونيسية للأرصاد الجوية وعلم المناخ والجيوفيزياء. وتُظهر نتائج الدراسة أن تحليل المكونات الرئيسية يُسهم بفعالية في تقليل تعقيد البيانات، ويدعم عملية تصنيف مستويات الاهتزازات الأرضية الناتجة عن الزلازل بكفاءة.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan negara dengan tingkat kerentanan yang tinggi terhadap bencana alam (Harijoko *et al.*, 2024). Posisi wilayah yang berada di kawasan Cincin Api Pasifik (*Pacific Ring of Fire*) menyebabkan sangat rentan terhadap bencana alam (Wibowo, 2024). Secara geografis, Indonesia terletak pada pertemuan empat lempeng tektonik, yaitu lempeng Benua Asia, lempeng Benua Australia, lempeng Samudera Hindia, dan lempeng Samudera Pasifik (Inayah *et al.*, 2024). Letak geografis tersebut menjadikan Indonesia sangat rawan mengalami berbagai bencana alam, seperti gempa bumi, tsunami, letusan gunung berapi, serta tanah longsor (Dartanto, 2022). Berdasarkan data Badan Nasional Penanggulangan Bencana (BNPB) pada sepanjang tahun 2024 tercatat sebanyak 3.472 bencana alam yang didominasi oleh bencana hidrometeorologi seperti banjir, tanah longsor, kekeringan, dan puting beliung (BNPB, 2024).

Menurut data *World Risk Report* (2017) menegaskan bahwa, kondisi kerentanan Indonesia memiliki persentase paparan bencana sebesar 19.36% dengan nilai kerentanan mencapai 54.19% (Fuady *et al.*, 2021). Artinya, meskipun tingkat keterpaparan bencana tinggi, namun kesiapan Indonesia dalam menghadapi bencana masih tergolong rendah sehingga dampak bencana bisa menjadi lebih parah (Susanto, 2023).

Di antara berbagai bencana tersebut, gempa bumi menjadi salah satu yang paling sering terjadi di Indonesia (Hadi *et al.*, 2019). Gempa bumi dapat terjadi

secara mendadak dengan intensitas yang bervariasi, mulai dari getaran ringan hingga guncangan besar yang menyebabkan kerusakan parah (AlKhaldi, 2024). Setiap peristiwa gempa bumi direkam oleh Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) dalam bentuk data seismik dengan parameter numerik seperti magnitudo, kedalaman, koordinat geografis, dan waktu kejadian (Supendi *et al.*, 2020).

Dalam perspektif Islam, peristiwa bencana alam gempa bumi telah disebutkan dalam Al-qur'an QS. Az-Zalzalah ayat 1 – 5 yang menjelaskan mengenai guncangan besar yang akan terjadi pada hari kiamat, di mana bumi diguncang dengan sangat dahsyat hingga mengeluarkan seluruh isinya. Adapun bunyi ayat tersebut yang berbunyi sebagai berikut.

إِذَا زُلْزِلَتِ الْأَرْضُ زِلْزَالَهَا ۚ ۱ وَأَخْرَجَتِ الْأَرْضُ أَثْقَالَهَا ۚ ۲ وَقَالَ الْإِنْسَانُ مَا هَٰذَا ۚ ۳ يَوْمَئِذٍ تُخْبِرُهَا ۚ ۴ أَنَّ
رَبَّكَ أَوْحَىٰ ۚ ۵

“Apabila bumi diguncangkan dengan guncangan yang dahsyat; dan bumi telah mengeluarkan isi perutnya; dan manusia bertanya, “Apa yang terjadi dengannya (bumi)?”; Pada hari itu (bumi) menyampaikan berita (tentang apa yang diperbuat manusia di atasnya); karena sesungguhnya Tuhanmu telah memerintahkan (yang sedemikian itu) kepadanya.” (QS. Az-Zalzalah: 1–5).

Berdasarkan Tafsir Ibnu Katsir, ayat tersebut menjelaskan guncangan yang menyebabkan bumi menampakkan segala amal perbuatan manusia, baik kebaikan maupun keburukan, sehingga tidak ada yang luput dari perhitungan Allah (Ghoffar, 2004). Dalam konteks yang lebih sederhana, ayat-ayat tersebut dapat dikaitkan dengan fenomena gempa bumi, meskipun skalanya tidak sebanding dengan peristiwa kiamat. Secara ilmiah telah dijelaskan bahwa gempa terjadi akibat pergerakan serta interaksi antar lempeng bumi. Misalnya, di wilayah selatan Pulau

Jawa terdapat pertemuan antara Lempeng Eurasia dan Indo-Australia. Ketika dua lempeng tersebut saling bertemu, tekanan akan terus menumpuk hingga pada titik tertentu lempeng tidak mampu lagi menahannya. Energi yang tersimpan selama ratusan tahun kemudian dilepaskan secara tiba-tiba dalam bentuk gempa bumi. Fenomena inilah yang sejalan dengan gambaran “bumi menyampaikan beritanya.” Proses akumulasi dan pelepasan beban ini berlangsung berulang-ulang, dan dapat direkam baik melalui tanda-tanda alamiah maupun dengan bantuan alat geofisika. Dalam Hadits Nabi SAW juga menegaskan bahwa:

أَتَقُومُ السَّاعَةَ حَتَّى تَكْثُرَ الزَّلَازِلُ

“Tidak akan terjadi kiamat hingga banyak terjadi gempa bumi.” (HR. Bukhori).

Hadits tersebut memperkuat makna yang terkandung dalam Surah Az-Zalzal, bahwa keterkaitan antara fenomena alam berupa guncangan bumi. Dalam konteks keilmuan, fenomena gempa bumi dapat diteliti dan dianalisis secara sistematis agar diperoleh pemahaman yang lebih baik mengenai pola dan intensitasnya. Penelitian ini, yang menggunakan metode *Principal Component Analysis* (PCA) dan *clustering*, diharapkan dapat memberikan kontribusi nyata dalam upaya mitigasi bencana. Dengan demikian, pendekatan ilmiah tidak hanya berfungsi untuk kepentingan akademis, tetapi juga selaras dengan pesan spiritual Islam yang menekankan agar manusia mengambil hikmah dari setiap kejadian alam.

Berdasarkan penelitian yang telah dilakukan sebelumnya, *Principal Component Analysis* (PCA) dikenal sebagai salah satu metode reduksi dimensi data dengan menyederhanakan variabel data tetapi tidak menghilangkan informasi yang

penting. PCA telah digunakan di berbagai bidang, seperti geofisika untuk mengklasifikasikan *facies* seismik (Troccoli *et al.*, 2022), serta dalam pengolahan citra untuk mengurangi kompleksitas data dan mempercepat waktu eksekusi (Rahman, 2020). Selain itu, pemilihan metode PCA dalam penelitian ini merujuk pada studi yang dilakukan oleh (Almais *et al.*, 2023) yang telah berhasil memanfaatkan PCA untuk proses *clustering* sehingga menghasilkan suatu teknik *clustering* baru yaitu PCA-Clustering. Penelitian tersebut menggunakan data kerusakan sektor pasca-bencana, PCA diterapkan dengan menggunakan dua komponen utama, yakni PC1 dengan kontribusi variansi sebesar 85.17% dan PC2 dengan kontribusi variansi 9.36%. Hasil transformasi data kemudian dikelompokkan ke dalam tiga *cluster* berdasarkan nilai koordinator PC1, yaitu *cluster* 1 ($n < 0$) mewakili kerusakan ringan, *cluster* 2 ($0 \leq n < 2$) mewakili kerusakan sedang, dan *cluster* 3 ($n \geq 2$) mewakili kerusakan berat. Proses validasi terhadap data target menunjukkan bahwa pembagian berdasarkan PC1 konsisten dengan kondisi kerusakan yang ditemukan di lapangan. Hal ini membuktikan bahwa PCA mampu melakukan pelabelan data secara efektif, sehingga relevan untuk digunakan pada penelitian ini. Dalam penelitian ini, metode *clustering* juga digunakan untuk mengelompokkan gempa berdasarkan hasil dari tingkat guncangannya.

Principal Component Analysis (PCA) adalah istilah umum untuk sebuah teknik yang menggunakan prinsip-prinsip matematis yang kompleks untuk mentransformasikan sejumlah variabel yang mungkin saling berkorelasi menjadi sejumlah variabel yang lebih sedikit yang disebut komponen utama (*principal*

components) (M. S. Hasan, 2021). Teknik ini bekerja dengan cara mereduksi dimensi data tanpa menghilangkan informasi penting yang terkandung di dalamnya, sehingga analisis dapat dilakukan dengan lebih efisien. Dengan mengadopsi pendekatan yang sama, PCA berbasis *clustering* digunakan untuk membentuk komponen utama dari variabel data gempa setelah melalui proses normalisasi data. Hasil transformasi PCA berupa nilai numerik kemudian dikelompokkan untuk menentukan label tingkat guncangan gempa bumi berdasarkan kategori yang telah ditetapkan yaitu kategori *dirasakan* dan *tidak dirasakan* menggunakan acuan Skala Intensitas Gempabumi (SIG-BMKG) yang disusun oleh BMKG (Muzli *et al.*, 2017).

Pemilihan metode PCA pada penelitian ini didasarkan pada kemampuannya dalam mengidentifikasi pola dan hubungan antarvariabel secara efisien melalui reduksi dimensi. PCA dapat mengekstraksi informasi dominan dari sejumlah besar data sehingga menghasilkan representasi baru yang lebih sederhana dan mudah dianalisis (Uddin *et al.*, 2021). Selain itu, PCA juga mempercepat proses komputasi karena hanya menggunakan komponen utama dengan kontribusi variansi terbesar, yang berarti data yang diolah menjadi lebih ringan namun tetap informatif (Saccenti, 2024). Oleh karena itu, penggunaan PCA-*Clustering* relevan untuk mengolah data gempa bumi dalam jumlah besar, karena mampu mempercepat proses analisis sekaligus menghasilkan pengelompokan tingkat guncangan gempa bumi secara objektif dan efisien.

Adapun kontribusi penulis yaitu penelitian ini dilakukan untuk menganalisis tingkat guncangan gempa bumi menggunakan metode *clustering* dengan teknik

Principal Analysis Component (PCA). Penelitian ini bertujuan untuk mengembangkan metode yang lebih akurat dan efisien dalam menentukan tingkat guncangan gempa berdasarkan data gempa bumi, sehingga dapat memberikan kontribusi yang berarti dalam mendukung upaya mitigasi bencana di Indonesia. Dengan demikian, penelitian ini diharapkan mampu memperluas penerapan metode PCA dalam analisis tingkat guncangan gempa serta menjadi rujukan bagi pengembangan teknik pengelompokan atau klasterisasi data pada penelitian kebencanaan di masa mendatang.

1.2 Pernyataan Masalah

Berdasarkan penjelasan pada latar belakang yang telah diuraikan, permasalahan yang akan dikaji adalah bagaimana hasil pengelompokan yang diperoleh dari penerapan *Principal Component Analysis* (PCA) untuk melabelkan tingkat guncangan gempa.

1.3 Batasan Masalah

Terdapat aspek batasan penelitian untuk menghindari persimpangan data yang dihasilkan. Berikut adalah batasan masalah yang digunakan dalam penelitian ini:

- a. Penelitian ini hanya menganalisis tingkat guncangan gempa bumi.
- b. Data yang digunakan dalam penelitian ini adalah data gempa bumi yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG)

- c. Metode analisis yang digunakan adalah *Principal Component Analysis* (PCA), kemudian hasilnya dikelompokkan dengan teknik *clustering* (*unsupervised learning*).

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah mengevaluasi hasil *clustering* yang dihasilkan oleh metode *Principal Component Analysis* (PCA) dalam mengelompokkan tingkat guncangan gempa bumi.

1.5 Manfaat Penelitian

Dari uraian di atas, maka manfaat penelitian ini adalah sebagai berikut:

- a. Penerapan *PCA Clustering* dalam klasterisasi data gempa bumi berdasarkan tingkat guncangan yang dapat menjadi referensi dalam pengembangan metode analisis data gempa.
- b. Informasi yang dihasilkan melalui analisis PCA berbasis *Clustering* diharapkan mampu memberikan pemahaman yang lebih mendalam dengan merancang program yang tepat sasaran serta memberikan dampak positif. Dengan pemahaman yang lebih baik terhadap penelitian ini, diharapkan dapat berkontribusi secara langsung dalam upaya meningkatkan mitigasi bencana alam gempa bumi.

BAB II

STUDI PUSTAKA

2.1 Penelitian Terkait

Penelitian terdahulu merupakan penelitian yang sudah ada sebelum penelitian ini dilakukan yang relevan dengan topik penelitian. Penelitian tersebut digunakan sebagai acuan ataupun sebagai pedoman, bertujuan untuk memberikan dasar yang konseptual dan teoritis kuat terhadap permasalahan yang dibahas dalam penelitian, serta menunjukkan posisi penelitian yang sedang dilakukan dibandingkan dengan penelitian sebelumnya. Oleh karena itu, kajian studi ini sangat penting untuk dilakukan agar memastikan bahwa penelitian yang dilakukan tidak berdiri sendiri, tetapi terhubung dan berkontribusi terhadap pengetahuan yang sudah ada sebelumnya.

Penelitian terdahulu yang relevan telah dilakukan dengan mengembangkan metode *clustering* data dengan metode *Principal Component Analysis* (PCA) yang berfokus pada penentuan tingkat kerusakan melalui pengurangan dimensi data (Almais *et al.*, 2023). Dalam penelitian tersebut, melakukan pelabelan tingkat kerusakan sektor pasca bencana alam dengan menganalisis 216 data, yang setiap data memiliki lima kriteria. Tingkat kerusakan yang ditentukan adalah rusak ringan, rusak sedang, dan rusak berat. Data direduksi menjadi dua komponen utama yaitu PC1 dan PC2. PC1 memiliki kontribusi terbesar terhadap varians data sebesar 85.17% dengan nilai eigen 4.28%, sedangkan PC2 berkontribusi sebesar 9.36% dengan nilai eigen 0.47%. Hasil dari normalisasi data kemudian dibuat grafik dua dimensi untuk melihat visualisasi data hasil setiap komponen utama (PC).

Komponen tersebut digunakan untuk mengelompokkan data ke dalam tiga klaster yang mewakili kategori kerusakan ringan, sedang, dan berat. Klaster pertama terdiri atas data dengan nilai koordinat kurang dari 0 yang mewakili kerusakan ringan, klaster kedua mencakup data dengan nilai antara 0 hingga 2 yang menunjukkan kerusakan sedang, sedangkan klaster ketiga mencakup data dengan nilai lebih dari 2 yang mengindikasikan kerusakan berat. Hasil PCA kemudian divalidasi dengan cara membandingkan hasil klasterisasi terhadap data target asli, di mana terlihat bahwa PC1 dengan nilai eigen terbesar mampu memberikan distribusi data yang lebih optimal dibandingkan PC2. Proses validasi ini membuktikan bahwa PCA dapat dimanfaatkan secara efektif dalam pelabelan data tanpa pengawasan, khususnya untuk menentukan tingkat kerusakan pasca-bencana. Temuan tersebut menunjukkan bahwa penerapan PCA dalam klasterisasi data kerusakan bangunan pasca-bencana mampu menghasilkan pelabelan yang akurat untuk data *unsupervised* dan bermanfaat dalam mendukung pengambilan keputusan di tahap selanjutnya.

Berikut ini adalah penelitian-penelitian terdahulu lainnya yang disajikan dalam bentuk Tabel 2.1.

Tabel 2.1 Penelitian Terkait

No.	Referensi	Topik	Metode	Subjek
1.	(Uddin <i>et al.</i> , 2021)	Klasifikasi citra	<i>Principal Component Analysis</i> (PCA)	Pengurangan Dimensi dalam Klasifikasi Citra Penginderaan Jauh Hiperspektral (HSI).
2.	(Adiwijaya <i>et al.</i> , 2021)	Clustering data	<i>Principal Component Analysis</i> (PCA) dan K-Means (<i>Clustering</i>)	Mereduksi dimensi Pada Proses <i>Clustering</i> Data Produksi Pertanian Di Kabupaten Bojonegoro.
3.	(Enzellina & Suhaedi, 2022)	Penentuan faktor dominan	<i>Principal Component Analysis</i> (PCA)	Menentukan faktor dominan menggunakan metode

No.	Referensi	Topik	Metode	Subjek
				<i>principal component analysis</i> .
4.	(Sugiarto <i>et al.</i> , 2023)	Peningkatan Fitur Ekstraksi Pada Pengenalan Citra Batik	<i>Discrete Wavelet Transform</i> (DWT), <i>Principal Component Analysis</i> (PCA), KNN, dan SVM	Peningkatan metode ekstraksi fitur untuk pengenalan citra batik menggunakan DWT dan PCA untuk meningkatkan akurasi klasifikasi.
5.	(Ahmad & Irsalinda, 2020)	<i>Clustering</i> gempa berdasarkan kedalaman dan lokasi	Fuzzy K-Means	Data gempa di Indonesia dengan 10 magnitudo ≥ 5 SR.
6.	(Rifa <i>et al.</i> , 2020)	Risiko gempa dan pengelompokan daerah berdasarkan tingkat gempa	K-Means & K-Medoids	Data gempa regional Indonesia (periode tertentu).
7.	(Hu <i>et al.</i> , 2021)	Simulasi gerak tanah (<i>ground motion</i>) untuk gempa sedang	<i>Principal Component Analysis</i> (PCA)	Data gempa magnitudo sedang di wilayah Barat Tiongkok
8.	Penelitian saat ini	<i>Clustering</i> tingkat guncangan gempa	<i>Principal Component Analysis</i> (PCA)	Implementasi PCA <i>Clustering</i> untuk menentukan tingkat guncangan gempa

Penelitian lebih lanjut dengan metode yang sama telah dilakukan oleh Almais *et al.*, (2024) mengimplementasikan metode analisis gabungan GLCM dan PCA untuk meningkatkan keakuratan dalam mendeteksi kerusakan struktural bangunan pasca bencana. Pada proses GLCM bertugas untuk ekstraksi tekstur citra yang berisi informasi lalu dengan teknik PCA untuk mengelompokkan kumpulan data berdasarkan atributnya untuk mengetahui informasi. Hasil penelitian ini membuktikan bahwa kombinasi GLCM dengan menggunakan jarak 0,5 piksel dan PCA mampu mengelompokkan tingkat kerusakan bangunan ke dalam tiga kategori, yaitu kerusakan ringan, sedang, dan berat. Hasil menunjukkan bahwa kombinasi dari analisis PCA menggunakan visualisasi 2D dan 3D hasil GLCM pada jarak 0,5

piksel memberikan akurasi yang lebih tinggi, proses *clustering* lebih dominan mengarah pada kategori kerusakan berat dibandingkan penggunaan jarak 1 piksel.

Oktafiani & Itje Sela (2024) telah mengintegrasikan PCA dan *Synthetic Minority Oversampling Technique* (SMOTE) untuk meningkatkan akurasi klasifikasi kanker payudara. Penelitian ini menggunakan *Dataset Wisconsin Breast Cancer Diagnostic* (WBCD) dengan menerapkan algoritma *Random Forest* (RF) dan *Support Vector Machine* (SVM) sebagai metode klasifikasi. Teknik PCA digunakan untuk mereduksi dimensi data, sedangkan SMOTE diterapkan untuk menyeimbangkan distribusi antara kelas mayoritas dan minoritas. Pendekatan tersebut membantu menyederhanakan data, mengurangi ketidakseimbangan kelas, dan memberikan hasil klasifikasi yang lebih representatif.

Selanjutnya, penelitian yang dilakukan oleh Dinanti & Purwadi (2023) membahas metode *Principal Component Analysis* (PCA) untuk mereduksi dimensi data diabetes sebelum dilakukan klasifikasi menggunakan algoritma *K-Nearest Neighbor* (KNN). Permasalahan yang diangkat adalah jumlah variabel pada *Dataset* yang cukup banyak, sehingga berpotensi memperlambat proses komputasi. Untuk mengatasi hal tersebut, PCA diterapkan dan berhasil mereduksi delapan variabel asli menjadi tiga komponen utama (PC1, PC2, PC3). Data hasil reduksi kemudian diuji menggunakan KNN dengan nilai K yang berbeda, yaitu 3, 5, dan 7. Dari temuan tersebut dapat disimpulkan bahwa penerapan PCA mampu mengurangi jumlah variabel tanpa mengurangi informasi penting sehingga proses klasifikasi lebih optimal, dan penggunaan KNN dengan parameter yang tepat dapat meningkatkan akurasi klasifikasi data diabetes.

Berdasarkan penelitian sebelumnya, menunjukkan efektivitas PCA dalam menyederhanakan dimensi data serta peran pentingnya ketika dikombinasikan dengan metode *clustering* untuk menghasilkan analisis yang lebih tepat dan akurat. Keterkaitan dengan penelitian ini adalah pemanfaatan PCA untuk mereduksi dimensi data yang kompleks pada data gempa serta mengelompokkannya sehingga terbentuk label yang dapat divalidasi oleh pihak BMKG. Dengan pendekatan ini, diharapkan mampu menghasilkan pengelompokan gempa yang lebih akurat dan efisien serta memperluas penerapan metode PCA *Clustering*, khususnya dalam menganalisis data gempa di Indonesia.

2.2 Gempa Bumi

Gempa Bumi adalah pergeseran secara tiba-tiba lapisan tanah yang ada di permukaan bumi (Wahyu & Rushenda, 2022). Pada saat pergeseran ini terjadi maka timbulah getaran yang biasa disebut gelombang seismik. Gelombang ini menjalar keseluruh arah dan menjauhi sumber gempa sehingga tercatat di seismograph, dan beberapa juga naik kepermukaan bumi sehingga dapat dirasakan semua makhluk yang ada dipermukaan bumi (Prathivi, 2020). Gempa bumi dapat terjadi oleh peristiwa letusan gunung api, benturan meteorit, tanah longsor, ledakan bom, dan banyak lagi penyebab lainnya. Namun, umumnya mereka disebabkan oleh gerakan mendadak kerak bumi di sepanjang bidang patahan (Hartatik & Cahya, 2020).

Letak Indonesia secara geografis berada pada pertemuan lempeng kerak bumi yaitu Indo-Australia, lempeng tektonik Pasifik dan Eurasia (Setiawan *et al.*, 2022). Gempa bumi merupakan peristiwa bumi bergetar diakibatkan oleh pelepasan energi di bawah permukaan bumi yang ditandai patahnya lapisan batuan kerak

bumi. Pergerakan lempeng-lempeng tektonik membuat akumulasi energi sehingga menghasilkan gelombang seismik akibat pelepasan energi yang merambat dan membuat guncangan pada permukaan bumi (Philia Christi Latue & Heinrich Rakuasa, 2023).

2.3 *Principal Component Analysis (PCA)*

Principal Component Analysis (PCA) merupakan metode statistik yang umumnya banyak digunakan untuk mengurangi dimensi data dengan tetap mempertahankan informasi yang paling penting serta dapat digunakan untuk analisis data (Saccenti, 2024). Metode ini mampu mengidentifikasi fitur utama, memvisualisasikan hubungan atau relasi, dan menyederhanakan sekumpulan data yang kompleks (Mohammed *et al.*, 2023).

Penerapan reduksi kompleksitas dimensi dengan metode PCA dapat dilakukan dengan mengekstraksi fitur dari perhitungan vektor eigen dan nilai eigen yang diperoleh dari *Principal Component*. Metode PCA dapat menyeleksi sebanyak n variabel menjadi k variabel baru yang disebut *Principal Component*. Hasil k yang diperoleh berjumlah lebih sedikit dari n variabel dan dapat menghasilkan informasi yang sama dengan n variabel. Variabel hasil dari seleksi disebut *Principal Component* atau dapat juga disebut faktor. *Principal Component Analysis (PCA)* berguna untuk menyeleksi fitur dan interpretasi variabel data asli (B. M. S. Hasan, 2021). Berikut ini adalah penjelasan lebih rinci tentang proses PCA beserta langkah-langkah utama dan rumusnya:

1. Normalisasi Data

Dalam implementasi PCA, tahap awal adalah normalisasi data untuk menyeimbangkan skala setiap variabel sehingga tidak ada variabel yang memiliki nilai lebih tinggi dari variabel lain atau mendominasi saat tahap analisis. Dalam penelitian ini, untuk proses normalisasi data menggunakan metode *Standard Scaler* (Almais *et al.*, 2023).

$$X_{Standart} = \frac{X - mean(x)}{\sigma} \quad (2.1)$$

Keterangan:

X = Nilai asli
 $mean(x)$ = Rata-rata variabel tersebut
 σ = Standar deviasi

Untuk menghitung standar deviasi menggunakan rumus:

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}} \quad (2.2)$$

Keterangan:

σ = Standar deviasi
 $\bar{X}$ = Rata-rata (*mean*)
 X_i = Setiap nilai dalam data populasi
 n = Jumlah total nilai dalam *dataset*

Langkah ini sangat penting dilakukan untuk proses tahap selanjutnya, bertujuan agar data yang digunakan memiliki *value* dan rata-rata yang sama setiap variabelnya.

2. Menentukan Matriks Kovarians dan *EigenValue*

Langkah selanjutnya adalah menentukan nilai eigen (*eigenvalue*) dan vektor eigen (*eigenvector*) dari hasil matriks kovarians. Matriks kovarians adalah ukuran yang menggambarkan seberapa besar tingkat variasi atau penyebaran data yang

dimiliki oleh suatu variabel. Nilai eigen merepresentasikan jumlah varians yang dihasilkan oleh masing-masing komponen utama. Vektor eigen untuk memberikan arah dari komponen tersebut. Perhitungan matriks kovarians dapat menggunakan persamaan berikut:

$$C = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.3)$$

Keterangan:

C = Matriks kovarians

n = Jumlah sampel

X_i = Vektor data pada titik ke- i

Nilai varians yang tinggi menunjukkan bahwa data memiliki penyebaran yang luas, sedangkan varians yang rendah mengindikasikan data lebih rapat di sekitar nilai rata-rata. Nilai eigen berada pada vektor eigen dalam bentuk matriks dan digunakan untuk menentukan arah komponen utama pada transformasi data. Semakin besar nilai eigen, semakin besar pula proporsi keragaman data yang dapat dijelaskan oleh komponen utama tersebut.

3. Menentukan Komponen Utama

Setelah melakukan langkah menghitung nilai eigen dan vektor eigen, maka diurutkan berdasarkan nilai eigen terbesar. Komponen utama dibentuk dari kombinasi linier variabel asli berdasarkan vektor eigen. Komponen dengan nilai eigen terbesar dipilih karena mampu memuat variasi informasi paling besar dari data. Jumlah komponen utama yang digunakan ditentukan berdasarkan persentase variasi total yang ingin dipertahankan.

4. Visualisasi 2D dan 3D

Langkah selanjutnya setelah menentukan komponen utama adalah memvisualisasikan ke bentuk grafik. Jika data memiliki banyak variabel, PCA bisa mereduksi dimensinya menjadi 2 komponen utama (PC1 dan PC2). Hasilnya bisa divisualisasikan dalam bentuk grafik *scatter plot* 2D, sehingga dapat menunjukkan pola, klaster, atau pemisahan antar data bisa terlihat lebih jelas.

5. Membuat Label

Langkah terakhir proses PCA adalah pemberian label pada data hasil setelah reduksi dimensi. Label ini memiliki fungsi sebagai penanda kelompok atau *cluster* yang terbentuk. Pada tahap ini, hasil visualisasi data sebelumnya dapat dibedakan berdasarkan kelas atau kategori tertentu sesuai dengan persebaran titik koordinat data. Dengan adanya label, hasil analisis dapat diinterpretasikan dengan lebih mudah karena pola pengelompokan data dapat terlihat secara jelas melalui perbedaan warna atau simbol yang mewakili setiap *cluster*.

Berdasarkan rangkaian tahapan penelitian ini, PCA digunakan untuk untuk membentuk komponen utama dari variabel data gempa setelah normalisasi data. Sehingga PCA dapat digunakan untuk mengelompokkan tingkat guncangan gempa berdasarkan label tertentu.

2.4 PCA Clustering

Clustering merupakan salah satu teknik *machine learning* tanpa pengawasan (*unsupervised*) yang mengungkap pola dan struktur data berdasarkan kemiripan karakteristik antar data dengan mengelompokkan data tidak berlabel (tanpa membutuhkan klasifikasi awal) yang serupa ke dalam beberapa *cluster* atau

kelompok. Objek-objek dalam satu *cluster* memiliki tingkat kemiripan pola atau nilai fitur yang tinggi, sedangkan objek antar *cluster* memiliki tingkat perbedaan yang signifikan (Feng *et al.*, 2022).

Metode pemetaan data berukuran besar ke dalam kelompok-kelompok yang lebih kecil dan homogen secara signifikan mempermudah proses analisis dengan meningkatkan efisiensi dan efektivitas dalam pengelompokan data. Pendekatan ini sangat bermanfaat dalam menangani kumpulan data berukuran besar, karena memungkinkan analisis dilakukan dengan lebih mudah dan memberikan wawasan yang lebih mendalam. Berbagai teknik dan algoritma telah dikembangkan untuk mencapai tujuan ini, masing-masing memiliki keunggulan dan penerapan yang berbeda-beda (Zhang, 2022).

Metode PCA *Clustering* yaitu menggabungkan proses reduksi dimensi dengan *Principal Component Analysis* (PCA) dan pengelompokan data berdasarkan kemiripan pola. PCA memanfaatkan titik koordinat PC dari hasil reduksi dimensi. Pada visualisasi dua dimensi, data titik koordinat merupakan data yang *unsupervised*, belum memiliki label. *Clustering* berfungsi untuk mengelompokkan data tersebut ke dalam rentang nilai tertentu sehingga membentuk satu label berdasarkan kategori yang telah ditentukan *supervised* (Almais *et al.*, 2023).

Dalam penelitian ini, pemilihan metode PCA *Clustering* karena mampu menyederhanakan dimensi data dengan mengekstraksi fitur-fitur utama, sehingga data membentuk komponen utama tanpa kehilangan informasi penting. Hasil PCA dikelompokkan menjadi *cluster-cluster* untuk menggambarkan pola

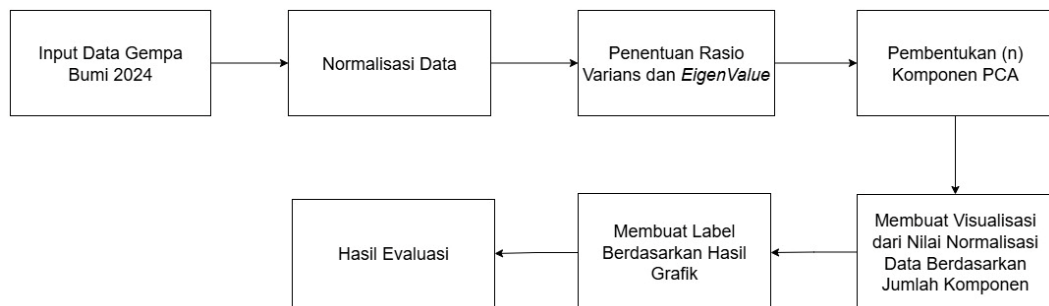
jenis gempa sesuai kategori yang ditetapkan BMKG sekaligus dapat menjadi informasi yang bermanfaat untuk mitigasi bencana dan sistem peringatan dini.

BAB III

DESAIN PENELITIAN

3.1 Desain Sistem

Pada subbab ini, akan menjelaskan mengenai alur dari desain sistem yang menggambarkan tahapan implementasi PCA – *Clustering* dalam mengelompokkan tingkat guncangan gempa bumi. Ilustrasi rancangan sistem pada penelitian ini ditampilkan pada Gambar 3.1. Tahapan dimulai dengan pengumpulan data yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Stasiun Geofisika Kelas III Malang, yang mencakup data gempa bumi dengan berbagai atribut relevan. Data tersebut kemudian diolah dengan menggunakan *Principal Component Analysis* (PCA) untuk melakukan reduksi dimensi data tanpa menghilangkan informasi penting, sehingga menjadi nilai komponen utama. Desain sistem ditunjukkan pada Gambar 3.1.



Gambar 3.1 Alur Desain Sistem
Sumber: (Almais *et al.*, 2023)

Gambar 3.1 menunjukkan alur desain sistem yang digunakan dalam mengelompokkan tingkat guncangan gempa bumi.

1. Input Data Gempa Bumi

Dataset yang digunakan pada penelitian ini adalah *Dataset* Gempa Bumi tahun 2024 yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Stasiun Geofisika Kelas III Malang. *Dataset* ini terdiri dari 4 fitur dan 8.539 baris data meliputi informasi rekapitulasi data gempa selama tahun 2024.

2. Normalisasi Data

Normalisasi bertujuan agar memastikan semua variabel atau fitur berada dalam skala yang setara sehingga tidak ada variabel yang mendominasi ketika melakukan analisis. Normalisasi yang digunakan untuk penelitian ini merupakan *StandardScaler* (Almais *et al.*, 2023). Setiap variabel memiliki skala yang berbeda, maka data perlu dinormalisasi. Normalisasi memastikan bahwa setiap fitur memiliki rata-rata nol ($\mu = 0$) dan standar deviasi ($\sigma = 1$) agar PCA dapat bekerja dengan optimal. Dengan demikian, variabel yang memiliki skala besar tidak akan mendominasi variabel dengan skala kecil dalam perhitungan.

3. Menentukan Matriks Kovarians.

Setelah tahap normalisasi, langkah selanjutnya adalah menentukan rasio varians dan nilai eigen. Rasio varians untuk mengukur korelasi antara variabel-variabel yang ada dalam *Dataset*. Rasio Varians merupakan ukuran untuk persebaran suatu data (Almais *et al.*, 2023). Selanjutnya, perhitungan *eigenvalue* dan *eigenvector* untuk menentukan arah varians terbesar. *Eigenvalue* dari rasio kovarians untuk menggambarkan seberapa besarnya variansi data oleh setiap

komponen utama. Semakin besar nilai eigen, maka semakin banyak variansi dalam data tersebut. *Eigenvector* untuk menunjukkan arah komponen utama dalam ruang variabel asli.

4. Menentukan Komponen Utama PCA

Setelah menghitung hasil *eigenvalue* dan *eigenvector*, maka hasil tersebut digunakan untuk memilih komponen utama PCA. Komponen utama yang memiliki nilai eigen tertinggi artinya yang memiliki banyak informasi akan dipilih untuk melakukan analisis selanjutnya. Proses pemilihan ini bertujuan mereduksi dimensi *Dataset* tanpa kehilangan informasi yang penting.

5. Membuat Visualisasi dalam Bentuk 2D dan 3D

Setelah menentukan komponen utama, kemudian hasil PCA divisualisasikan dalam bentuk grafik 2D dan 3D menggunakan *scatter plot* yang didapat berdasarkan hasil komponen utama. Visualisasi adalah suatu hal yang sangat penting dalam merepresentasikan suatu hasil (Muddana & Vinayakam, 2024). Visualisasi bertujuan agar dapat melihat pola variansi dalam data dengan jelas sehingga memudahkan untuk mengidentifikasi pola dalam data.

6. Membuat Label Berdasarkan Hasil Grafik

Setelah melakukan visualisasi dari komponen utama, maka label dapat dibuat untuk mengelompokkan data. Atau maka melakukan pelabelan berdasarkan data hasil komponen PCA. Melalui hasil visualisasi grafik 2D maupun 3D dari komponen utama, data dapat diberi label sesuai dengan pola atau kelompok (*cluster*) yang terbentuk. Label tersebut ditentukan berdasarkan kesamaan karakteristik yang terlihat pada grafik, sehingga data dengan kemiripan pola akan

tergabung dalam kelompok yang sama. Proses ini berfungsi untuk mempermudah identifikasi pola dan pengelompokan data. Dari pola sebaran data di grafik, dilakukan proses *clustering* untuk memberi label pada data gempa. Pengelompokan ini penting untuk menganalisis pola-pola atau tingkat kerusakan bangunan yang tersegmentasi.

7. Hasil Analisis

Pada tahap akhir, data hasil analisis PCA digunakan untuk pengelompokan data dan sebagai pendukung dalam menentukan keputusan analitis, untuk pengelompokan dan dasar untuk analisis lanjutan.

3.2 Pengumpulan Data

Dalam penelitian ini, data yang digunakan adalah data sekunder. Data sekunder adalah data yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Stasiun Geofisika Kelas III Malang. Data tersebut mencakup informasi berupa laporan terkait gempa pada tahun 2024. Untuk penelitian ini, jumlah data yang digunakan sebanyak 8539 data gempa bumi. Berikut Tabel 3.1 merupakan *Dataset* dari BMKG Kelas III Malang.

Tabel 3.1 Data Gempa Bumi

No.	Latitude	Longitude	Depth	Magnitudo
1.	-8.9	112.93	39	2.8
2.	-9.57	113.86	15	3.3
...				
4270.	-9.2	109.35	12	2.9
4271.	-9.33	113.93	32	3.2
...				
8539.	-9.12	113.55	115	2.7
8540.	-7.96	110.46	16	1.9

Pemilihan data gempa bumi dalam penelitian ini didasarkan pada variabel-variabel utama yang memiliki peran penting dalam menggambarkan karakteristik kejadian gempa, seperti magnitudo, kedalaman, serta koordinat lokasi. Setelah mengetahui isi data yang tercantum pada tabel, maka langkah selanjutnya adalah melakukan pemrosesan serta analisis data, yang mencakup normalisasi dan penerapan metode PCA untuk menghasilkan nilai numerik. Nilai tersebut kemudian digunakan dalam proses clustering sehingga mampu mengelompokkan tingkat guncangan gempa. Dengan demikian, analisis yang perlu dilakukan dapat mengenali pola distribusi data gempa secara lebih sistematis sekaligus memberikan dasar yang kuat untuk mendukung upaya mitigasi bencana.

3.3 Evaluasi Model

Dalam penelitian ini menggunakan metode *Calinski-Harabasz Index* (CHI) untuk melakukan pengujian terhadap hasil *clustering* dari *Principal Component Analysis* (PCA), merupakan matriks evaluasi kualitas hasil dari *clustering*. Hasil dari pengukuran ini selanjutnya digunakan sebagai dasar dalam mengevaluasi performa model *clustering* yang dihasilkan, sehingga dapat diketahui apakah proses reduksi dimensi menggunakan PCA mampu meningkatkan kualitas pengelompokan data secara keseluruhan.

3.3.1 *Calinski-Harabasz Index* (CHI)

Calinski-Harabasz Index disebut juga kriteria rasio varians dalam analisis klaster (Mardhotillah, Bunga; Fadhil, Amril; Elisa & Zurweni, 2023). Menentukan *Calinski-Harabasz Index* (CHI) yaitu dengan menjumlahkan ukuran dispersi *inter-*

cluster dan menjumlahkan ukuran dispersi klaster keseluruhan, dengan mengevaluasi *clustering* berdasarkan kedekatan atau kekompakan jarak antara *centroid* klaster dengan titik-titik data di dalam klaster tersebut, sedangkan pemisahan antar klaster diukur menggunakan jarak antara *centroid* klaster dengan *centroid global* (Ikotun *et al.*, 2025). Metode ini mengacu pada rasio jumlah dispersi (jumlah kuadrat jarak) antara klaster dan dalam klaster untuk semua klaster (Priyatna & Huda, 2026). Indeks CH memiliki kompleksitas komputasi linier $O(n.d)$ sehingga sangat efisien untuk *dataset* besar dan berdimensi tinggi (Ikotun *et al.*, 2025). Semakin besar nilai Indeks CH menunjukkan jumlah *cluster* yang semakin baik (Khairati *et al.*, 2019). Nilai indeks yang tinggi menunjukkan bahwa pemisahan antar *cluster* cukup jelas dan data dalam masing-masing *cluster* memiliki tingkat homogenitas yang baik. Rumus untuk *Calinski-Harabsz Index* ditunjukkan pada persamaan 3.1.

$$\text{Calinski – Harabsz Index (CH)} = \frac{SSB}{SSW} \times \frac{(n - k)}{(k - 1)} \quad (3.1)$$

$$SW = \sum_{l=1}^k \sum_{X_{ij} \in C_l} (X_{ij} - \bar{X}_{jl})^2$$

$$SB = \sum_{l=1}^k n_{jl} (\bar{X}_{jl} - \bar{X}_j)^2$$

$$\bar{X}_{jl} = \frac{1}{n_{jl}} \sum_{X_{ij} \in C_l} X_{ij}$$

$$\bar{X}_j = \frac{1}{n_j} \sum_i^n X_{ij}$$

Keterangan:

CH = Nilai indeks *Calinski–Harabsz*

SSB = *Sum of square between cluster* (tingkat perbedaan objek pada *cluster* yang

	berbeda, jumlah kuadrat antar <i>cluster</i>)
SSW	= <i>Sum of square within cluster</i> (tingkat perbedaan objek dalam satu <i>cluster</i> , Jumlah kuadrat dalam <i>cluster</i>)
n_j	= Banyaknya objek pada variabel ke- j
n_{jl}	= Banyaknya objek pada variabel ke- j <i>cluster</i> ke- l
k	= jumlah <i>cluster</i>
C_l	= <i>cluster</i> ke- l
X_{ij}	= objek ke- i variabel ke- j pada <i>cluster</i> ke- l
$\bar{X}_{jl}$	= nilai rata-rata variabel ke- j pada <i>cluster</i> ke- l

Semakin besar nilai SSB , maka semakin tinggi tingkat dispersi antar *cluster*.

Semakin kecil nilai SSW , maka semakin tinggi tingkat kedekatan atau hubungan antar data di dalam *cluster* tersebut (Wang & Xu, 2019).

3.3.2 Validasi Eksternal Data

Pengujian validasi eksternal adalah melakukan validasi eksternal guna memperoleh hasil yang lebih akurat. Validasi eksternal ini dilakukan dengan cara membandingkan hasil *clustering* berbasis PCA dengan data target asli yang diperoleh dari BMKG seperti, magnitudo dan kedalaman gempa. Tujuannya adalah untuk menilai sejauh mana akurasi hasil pengelompokan sesuai dengan label atau target yang tersedia dari BMKG. Dengan demikian, pengujian yang dilakukan tidak hanya mengevaluasi kualitas internal *cluster* melalui *Calinski–Harabasz Index*, tetapi juga memastikan bahwa hasil *clustering* konsisten dan relevan dengan pola data nyata dari BMKG.

3.4 Pseudocode Menggunakan Metode PCA Clustering

Pseudocode merupakan representasi dari langkah-langkah algoritma yang digunakan dalam proses *clustering* menggunakan metode *Principal Component Analysis* (PCA). *Pseudocode* untuk memberikan gambaran sistematis mengenai alur pemrosesan data, mulai dari input data, normalisasi, perhitungan varians dan

nilai eigen, pembentukan komponen utama PCA, hingga visualisasi hasil dan pemberian label data. Dengan *pseudocode* ini, alur proses PCA dapat dijelaskan secara jelas dan terstruktur, sehingga memudahkan implementasi ke dalam bahasa pemrograman dan memastikan bahwa setiap tahap algoritma berjalan secara efisien.

Tabel 3.2 *Pseudocode Principal Component Analysis*

Algoritma:	Clustering menggunakan PCA untuk pelabelan data
Input:	Membaca <i>Dataset</i> csv menggunakan fungsi <code>pd.read_csv('Gempa.csv')</code>
Proses:	a. Normalisasi data menggunakan fungsi <code>StandarScaler()</code> dan <code>fit_transform()</code> b. Menghitung rasio varians menggunakan fungsi <code>explained_variance_ratio_()</code> c. Menghitung nilai eigen menggunakan fungsi <code>explained_variance_()</code> d. Menentukan komponen utama PCA menggunakan fungsi <code>PCA()</code> e. Membuat data frame dari hasil komponen utama PCA menggunakan fungsi <code>pd.DataFrame()</code> f. Visualisasi hasil normalisasi PCA menggunakan fungsi <code>scatterplot()</code> g. Membuat label data menggunakan percabangan (if - else) dan perulangan (for)
Output:	a. Menggabungkan dataFrame dari komponen PCA dengan label menggunakan <code>concat()</code> b. Hasil visualisasi data setelah pelabelan dengan PCA yang divalidasi dengan data asli dari BMKG menggunakan <code>scatterplot()</code>

BAB IV

HASIL DAN PEMBAHASAN

4.1 *Pre-processing Data*

Pada tahap *pre-processing data* dilakukan untuk mendapatkan data gempa yang berkualitas dan siap untuk diproses ke dalam model menggunakan metode *Principal Component Analysis* (PCA). Tahap *pre-processing* data ini terdapat dua tahap yaitu teknik pengumpulan data dan normalisasi data. Data yang digunakan untuk tahap *pre-processing* berjumlah 8.539 baris. Dalam penelitian ini, tahap *pre-processing* dilakukan dengan beberapa tahap sebagai berikut.

4.1.1 Teknik Pengumpulan Data

Dalam penelitian ini menggunakan data sekunder yang telah diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Stasiun Geofisika Kelas III Malang. Pada data ini berisi rekapan data gempa pada tahun 2024 yang terdiri dari 4 (empat) variabel.

Tabel 4.1 Data Gempa Bumi

No.	Latitude	Longitude	Depth	Magnitudo
1.	-8.9	112.93	39	2.8
2.	-9.57	113.86	15	3.3
...				
4270.	-9.2	109.35	12	2.9
4271.	-9.33	113.93	32	3.2
...				
8539.	-9.12	113.55	115	2.7
8540.	-7.96	110.46	16	1.9

Pemilihan atribut atau variabel ini berdasarkan pada pertimbangan bahwa masing-masing memiliki pengaruh yang signifikan terhadap gempa bumi di suatu

wilayah. Penggunaan data dari BMKG menjamin keakuratan karena diperoleh melalui pengamatan instrumen resmi terstandar.

4.2 Normalisasi Data

Normalisasi data atau standarisasi data bertujuan untuk mengubah data asli menjadi bentuk yang lebih sesuai untuk analisis *Principal Component Analysis* (PCA). Tahap ini sangat penting untuk memastikan semua fitur atau atribut memiliki skala yang serupa. Normalisasi data membantu menghindari bias pada fitur-fitur dengan skala yang berbeda. Variabel latitude dan longitude mewakili koordinat geografi, variabel kedalaman dinyatakan dalam kilometer (km), dan variabel magnitudo dalam skala magnitudo. Perbedaan skala tersebut berpotensi menyebabkan variabel dengan rentang nilai yang lebih besar mendominasi hasil analisis, khususnya pada metode yang berbasis jarak dan variansi seperti *Principal Component Analysis* (PCA).

4.2.1 *StandardScale*

Proses normalisasi pada penelitian ini dilakukan menggunakan metode *StandardScale*, yaitu metode normalisasi yang mentransformasikan data berdasarkan rata-rata (*mean*) dan standar deviasi dari masing-masing variabel. Setiap nilai data dikurangi dengan nilai rata-rata variabel kemudian dibagi dengan standar deviasinya. Proses ini bertujuan untuk memusatkan data pada nilai nol dan menyamakan tingkat penyebaran data.

Setelah dilakukan normalisasi, setiap variabel memiliki nilai rata-rata (*mean*) sebesar 0, yang berarti bahwa data telah dipusatkan pada titik rata-rata

variabel asli. Nilai 0 menunjukkan posisi rata-rata, sedangkan nilai positif dan negatif masing-masing menunjukkan bahwa data berada di atas atau di bawah rata-rata. Dengan demikian, perbedaan posisi data terhadap rata-rata dapat diamati secara langsung tanpa dipengaruhi oleh satuan asli variabel.

Selain itu, hasil normalisasi juga menghasilkan standar deviasi sebesar 1, yang menunjukkan bahwa penyebaran data dinyatakan dalam satuan deviasi standar. Besarnya nilai hasil normalisasi menunjukkan jarak suatu data terhadap rata-rata dalam satuan standar deviasi. Hasil proses normalisasi menunjukkan bahwa seluruh variabel telah berada pada skala yang relatif seimbang.

Tabel 4.2 Hasil Normalisasi Data

No.	Latitude	Longitude	Depth	Magnitudo
1.	-0.472723	1.370183	0.207244	0.155423
2.	-1.039744	2.108433	0.397148	0.990848
3.	0.458207	0.100076	0.497880	0.345832
4.	0.906746	-2.003538	0.523063	0.847087
5.	0.339725	-0.590544	0.271233	2.350853

Normalisasi dengan *StandardScale* menjadi penting karena variabel gempa yang digunakan dalam penelitian ini memiliki satuan dan rentang nilai yang berbeda, seperti latitude dan longitude mewakili koordinat geografi, variabel kedalaman dinyatakan dalam kilometer (km), dan variabel magnitudo dalam skala magnitudo. Tanpa normalisasi, variabel dengan rentang nilai yang lebih besar berpotensi mendominasi proses analisis.

4.3 Principal Component Analysis

Component Analysis dalam penelitian ini dilakukan menggunakan metode *Principal Component Analysis* (PCA) dengan tujuan untuk mereduksi dimensi data gempa yang memiliki banyak variabel sekaligus mengekstraksi informasi utama

yang paling berpengaruh terhadap variasi data. PCA bekerja dengan mentransformasikan variabel-variabel asli yang saling berkorelasi menjadi sejumlah komponen utama baru yang bersifat tidak berkorelasi satu sama lain. Setiap komponen utama merepresentasikan kombinasi linier dari variabel asli dan disusun berdasarkan besarnya variansi data yang mampu dijelaskan.

Penerapan PCA pada penelitian ini dilakukan setelah seluruh data melalui tahap preprocessing dan normalisasi menggunakan *StandardScale*. Langkah ini penting untuk memastikan bahwa setiap variabel memiliki kontribusi yang seimbang dalam pembentukan komponen utama, sehingga hasil PCA benar-benar merefleksikan struktur variasi data gempa yang sesungguhnya dan tidak dipengaruhi oleh perbedaan skala pengukuran antarvariabel.

4.3.1 Menentukan Nilai Eigen dan Rasio Varians

Nilai *eigenvalue* dan *variance ratio* merupakan indikator utama dalam menentukan jumlah serta kontribusi setiap komponen utama pada hasil *Principal Component Analysis* (PCA). Nilai *eigenvalue* diperoleh dari hasil dekomposisi matriks kovarians antarvariabel dan menunjukkan besarnya variasi data yang mampu dijelaskan oleh masing-masing komponen utama. Setiap komponen utama memiliki satu nilai *eigenvalue*, di mana nilai *eigenvalue* yang lebih besar menandakan bahwa komponen tersebut mengandung informasi yang lebih dominan dalam merepresentasikan data asli.

Sementara itu, *variance ratio* atau rasio variansi merupakan nilai yang diperoleh dari perbandingan antara *eigenvalue* suatu komponen utama dengan total *eigenvalue* seluruh komponen. Nilai ini menggambarkan proporsi atau persentase

variasi total data yang dapat dijelaskan oleh masing-masing komponen utama, sehingga dapat digunakan untuk mengevaluasi tingkat kontribusi setiap komponen terhadap keseluruhan informasi data.

Dengan demikian, komponen utama dengan *variance ratio* yang tinggi dipertahankan karena mampu mewakili sebagian besar variasi data, sedangkan komponen dengan *variance ratio* yang rendah tidak dipertahankan karena kontribusinya relatif kecil. Hasil nilai eigen dan rasio varians disajikan dalam Tabel 4.3 berikut.

Tabel 4.3 Hasil Nilai Eigen dan Rasio Varians

Komponen	Nilai Eigen	Rasio Varians
PC1	1.364353	0.341048
PC2	1.064056	0.265983
PC3	0.973796	0.243421
PC4	0.598263	0.149548

Berdasarkan hasil perhitungan nilai eigen dan rasio varians pada tabel di atas, terlihat bahwa beberapa komponen utama memiliki kontribusi yang cukup signifikan terhadap total variasi data. Komponen utama pertama (PC1) memiliki nilai eigen sebesar 1.364353 dengan *variance ratio* 0.341048, yang berarti komponen ini mampu menjelaskan sekitar 34.1% dari total variasi data. Hal ini menunjukkan bahwa PC1 menjadi komponen paling dominan dan berperan penting dalam merepresentasikan pola umum variasi pada *Dataset* gempa bumi.

Komponen utama kedua (PC2) memiliki nilai eigen sebesar 1.064056 dengan *variance ratio* 0.265983, atau sekitar 26.6% dari total variasi. Kontribusi ini masih cukup besar, sehingga PC2 melengkapi informasi tambahan yang tidak tercakup pada PC1. Secara bersama-sama, PC1 dan PC2 mampu menjelaskan sekitar 60.7% total variasi data.

Komponen berikutnya, yaitu PC3, memiliki nilai eigen yang tinggi sebesar 0.973796 dengan *variance ratio* 0.243421 atau sekitar 24.34% dari total variasi data. PC3 tetap memberikan kontribusi penting meskipun tidak sebesar dua komponen pertama, sehingga tiga komponen utama (PC1–PC3) secara kumulatif mampu menjelaskan sekitar 85.04% variasi dalam data. Sementara itu, untuk komponen terakhir PC4 memberikan kontribusi yang lebih kecil dengan *variance ratio* 0.149548 atau 14.95% dan nilai eigen sebesar 0.598263.

Secara keseluruhan, hasil ini menunjukkan bahwa sebagian besar variasi data dapat dijelaskan oleh 2 (dua) komponen utama, yaitu PC1 dan PC2 dengan total kontribusinya lebih dari 50% yaitu 60.7% serta memiliki nilai eigen yang lebih dari 1 (nilai eigen > 1), yang umumnya dianggap sangat memadai dalam analisis PCA. Dengan demikian, reduksi dimensi menjadi dua komponen utama sudah optimal untuk mempertahankan informasi penting dalam *Dataset*, dan komponen-komponen ini dapat digunakan untuk analisis lanjutan seperti visualisasi, *clustering*, atau interpretasi faktor dominan dalam data gempa.

4.3.2 Menentukan Komponen Utama

Berdasarkan hasil analisis sebelumnya, penentuan komponen utama dilakukan dengan mempertimbangkan besarnya nilai *eigenvalue*, *variance ratio*, serta faktor *loading* variable yang telah disajikan dalam Tabel 4.4 berikut.

Tabel 4.4 Hasil Faktor *Loading* Variabel

Variabel	PC1	PC2	PC3	PC4
Latitude	0.055667	0.897257	0.351669	-0.261297
Longitude	0.818659	0.136841	0.21202	0.515981
Depth	-0.222158	-0.381343	0.897281	-0.015092
Magnitudo	0.80106	-0.307958	0.007727	-0.513344

Hasil analisis faktor *loading* menunjukkan bahwa komponen utama pertama (PC1) didominasi secara kuat oleh dua variabel yaitu, Longitude dan Magnitudo, yang masing-masing memiliki *loading* tinggi sebesar 0.818659 dan 0.80106. Nilai ini menegaskan bahwa PC1 hampir sepenuhnya menggambarkan intensitas guncangan dalam data gempa. Variabel lain seperti *Depth* (-0.222158) hanya memberikan kontribusi tambahan yang jauh lebih kecil, sehingga perannya terhadap PC1 dapat dianggap minor atau sangat sedikit.

Pada komponen utama kedua (PC2), variabel dengan pengaruh paling besar atau faktor *loading* terbesar adalah Latitude sebesar 0.897257 dan Longitude sebesar 0.136841. Kedua variabel ini menunjukkan hubungan kuat dalam pembentukan PC2, sehingga komponen utama kedua ini dapat diinterpretasikan sebagai dimensi yang merepresentasikan wilayah kejadian gempa. Variabel lain seperti Magnitudo (-0.307958) dan *Depth* (-0.381343) hanya memberikan kontribusi kecil.

Komponen utama ketiga (PC3) memperlihatkan faktor *loading* tertinggi pada *Depth* sebesar 0.897281, yang menunjukkan bahwa PC3 terutama menggambarkan kedalaman gempa terjadi. Lalu, diikuti oleh variabel Latitude sebesar 0.351669. Variabel lain seperti Longitude (0.21202) dan Magnitudo (0.007727) berkontribusi jauh lebih kecil. Dengan dominasi yang jelas pada variabel *Depth*, PC3 menjadi dimensi yang merepresentasikan variasi kedalaman gempa terjadi.

Pada komponen utama keempat (PC4), variabel dengan kontribusi terbesar adalah Longitude dengan nilai faktor *loading* sebesar 0.515981, disusul oleh *Depth*

(-0.015092), Latitude (-0.261297), dan Magnitudo (-0.513344) yang memberikan kontribusi tambahan kecil. Meskipun *loading*-nya tidak sebesar pada komponen sebelumnya, PC4 dapat diinterpretasikan sebagai dimensi minor yang menggambarkan variasi Longitude.

Secara keseluruhan, pada komponen utama pertama (PC1) menggambarkan intensitas guncangan gempa, pada PC2 menggambarkan wilayah kejadian gempa, PC3 dapat diinterpretasikan sebagai variasi kedalaman gempa, dan komponen utama terakhir (PC4) yang menggambarkan variasi koordinat geografis dan kedalaman. Berikut rincian tabel hasil komponen utama (PC).

Tabel 4.5 Rincian Hasil Komponen Utama (PC)

Komponen	Nilai Eigen	Rasio Varians	Faktor <i>Loading</i> (Terbesar ke Terkecil)
PC1	1.364353	0.341048	1. Longitude 2. Magnitudo 3. Latitude 4. <i>Depth</i>
PC2	1.064056	0.265983	1. Latitude 2. Longitude 3. Magnitudo 4. <i>Depth</i>
PC3	0.973796	0.243421	1. <i>Depth</i> 2. Latitude 3. Longitude 4. Magnitudo
PC4	0.598263	0.149548	1. Longitude 2. <i>Depth</i> 3. Latitude 4. Magnitudo

Berdasarkan hasil analisis *Principal Component Analysis* (PCA), komponen utama pertama hingga ketujuh memiliki nilai eigen, rasio varians, dan faktor *loading* yang berbeda-beda, yang secara bersama-sama menggambarkan struktur variasi data gempa pada penelitian ini.

1. Komponen utama pertama (PC1) memiliki nilai eigen sebesar 1.364353 dengan rasio varians 0.341048 atau sekitar 34.1% dari total variasi data. Berdasarkan faktor *loading*, PC1 didominasi oleh variabel Longitude (0.818659) dan Magnitudo (0.80106) sebagai variabel dengan kontribusi terbesar, diikuti oleh Latitude (0.055667), serta kontribusi yang kecil dari *Depth* (-0.222158). Dominasi variabel Longitude menunjukkan bahwa PC1 dapat diinterpretasikan sebagai komponen yang merepresentasikan lokasi wilayah dan interaksinya dengan faktor besaran getaran gempa.
2. Komponen utama kedua (PC2) memiliki nilai eigen 1.064056 dengan rasio varians 0.265983 atau sekitar 26.6% dari total variasi data. PC2 didominasi oleh Latitude (0.897257) dan Longitude (0.136841). Variabel Magnitudo dan *Depth* memiliki nilai faktor *loading* lebih kecil yaitu masing-masing -0.307958 dan -0.381343. Oleh karena itu, PC2 dapat diinterpretasikan sebagai komponen yang menggambarkan dinamika lokasi wilayah dengan tambahan efek getaran gempa, yang menjadi faktor penting dalam variasi kondisi gempa.
3. Komponen utama ketiga (PC3) memiliki nilai eigen 0.973796 dengan rasio varians sebesar 0.243421 atau sekitar 24.34% dari total variasi data. Variabel dengan faktor *loading* terbesar adalah *Depth* sebesar 0.897281, diikuti oleh Latitude sebesar 0.351669. Komponen Longitude dan Magnitudo memberikan kontribusi relatif lebih kecil dibandingkan variabel lainnya. Dengan demikian, PC3 menggambarkan variasi yang terutama

dipengaruhi oleh variasi intensitas kedalaman dan dipengaruhi oleh letak geografis.

4. Komponen utama keempat (PC4) memiliki nilai eigen 0.968077 dengan rasio varians 0.193593 (19,35%). Nilai faktor *Loading* yang dominan pada PC4 terdapat pada Longitude (0.515981), diikuti oleh *Depth* (-0.015092). Variabel lainnya memberikan kontribusi kecil seperti Latitude memiliki nilai sekitar -0.261297 dan Magnitudo -0.513344. PC4 dapat diinterpretasikan sebagai komponen yang menggambarkan pola area terjadi gempa serta kedalaman, di mana wilayah menjadi faktor paling dominan pada komponen ini.

Secara keseluruhan, hasil PCA menunjukkan bahwa informasi utama dalam *dataset* paling banyak dijelaskan oleh dua komponen pertama (PC1–PC2), dengan PC1 dan PC2 sebagai komponen paling berkontribusi dan menyajikan banyak informasi. Komponen lainnya hanya menyumbang variasi kecil dan lebih bersifat pendukung.

4.4 Visualisasi Grafik

Visualisasi grafik merupakan tahapan penting dalam analisis *Principal Component Analysis* (PCA) karena berfungsi untuk menggambarkan struktur penyebaran data setelah proses reduksi dimensi dilakukan. Melalui visualisasi, pola hubungan antar data dapat diamati secara lebih jelas, termasuk tingkat variasi, kecenderungan pengelompokan, serta keberadaan data yang memiliki karakteristik berbeda atau ekstrem. Visualisasi ini juga membantu dalam mengevaluasi sejauh

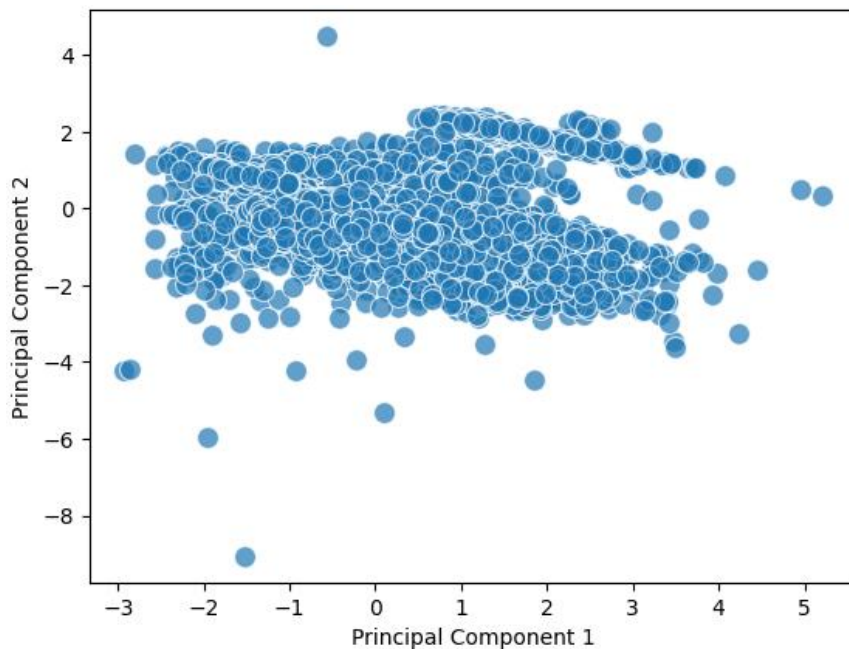
mana komponen utama mampu merepresentasikan variasi data gempa secara efektif.

Pada penelitian ini, visualisasi dilakukan dengan memproyeksikan data gempa ke dalam ruang komponen utama hasil PCA. Representasi grafis ini digunakan sebagai dasar awal untuk mengidentifikasi pola sebaran data dan potensi pembentukan klaster sebelum dilakukan proses *clustering* secara lebih formal.

Visualisasi ini juga membantu dalam mengidentifikasi keberadaan data ekstrem atau *outlier*, yaitu titik-titik yang berada cukup jauh dari pusat sebaran utama. Keberadaan data ekstrem tersebut dapat memengaruhi proses pembentukan klaster dan evaluasi model, sehingga penting untuk diperhatikan dalam tahapan analisis selanjutnya.

4.4.1 Visualisasi Grafik *Scatter Plot*

Visualisasi *scatter plot* menggunakan komponen utama pertama (PC1) dan komponen utama kedua (PC2) dilakukan untuk menggambarkan struktur penyebaran data gempa setelah proses reduksi dimensi dengan PCA. Representasi ini memungkinkan identifikasi pola dan hubungan antar sampel secara lebih jelas dibandingkan visualisasi pada ruang variabel asli yang memiliki dimensi tinggi. Dengan memproyeksikan data ke dalam dua komponen utama yang memiliki kontribusi informasi terbesar, grafik ini memberikan gambaran awal mengenai variasi utama dalam *dataset* serta potensi pembentukan kelompok atau klaster berdasarkan kemiripan karakteristik gempa.



Gambar 4.1 Grafik *Scatter Plot* Hasil PCA

Pada Gambar 4.1 menunjukkan bahwa sebaran titik pada grafik memperlihatkan bagaimana data gempa tersebar setelah direduksi dimensinya dari variabel asli ke dua komponen utama yang memiliki kontribusi informasi terbesar. Terlihat bahwa titik-titik data menumpuk atau membentuk satu kelompok bukan menyebar ke beberapa area dengan pola distribusi yang berbeda. Hal ini mengindikasikan adanya variasi yang cukup kuat antar sampel, terutama terkait karakteristik longitude, latitude, magnitudo, dan parameter gempa lainnya.

Grafik dua dimensi ini menampilkan proyeksi setiap data gempa pada bidang PC1–PC2, di mana sumbu horizontal merepresentasikan PC1 sebagai komponen dengan kontribusi variansi terbesar, sedangkan sumbu vertikal merepresentasikan PC2 sebagai komponen dengan kontribusi variansi kedua terbesar. Penyebaran titik pada grafik menunjukkan bagaimana data terdistribusi

setelah direduksi dari ruang variabel asli yang berdimensi tinggi ke dalam dua komponen utama.

Berdasarkan hasil visualisasi, terlihat bahwa sebaran data cenderung lebih dominan pada arah PC1, yang menunjukkan bahwa variasi utama antar data gempa paling banyak dijelaskan oleh komponen ini. Titik-titik data membentuk pola penyebaran yang hampir sepenuhnya homogen, menunjukkan adanya area dengan kepadatan tinggi dan area dengan kepadatan rendah. Kondisi ini mengindikasikan adanya perbedaan karakteristik gempa antar kelompok data.

Secara keseluruhan, visualisasi dua dimensi ini menunjukkan bahwa data cenderung tersebar dan membentuk satu kelompok besar di sepanjang sumbu PC1, sehingga belum terlihat adanya struktur klaster yang terbentuk secara alami. Meskipun demikian, kondisi ini tetap mendukung penggunaan metode *clustering* pada tahap selanjutnya untuk mengidentifikasi kemungkinan adanya pola pengelompokan yang tidak tampak secara langsung, sehingga analisis pengelompokan dapat dilakukan secara lebih terstruktur dan interpretatif.

Grafik tersebut menampilkan hasil reduksi dimensi data menggunakan metode *Principal Component Analysis* (PCA) ke dalam dua komponen utama, yaitu *Principal Component 1* (PC1) dan *Principal Component 2* (PC2). Setiap titik pada grafik merepresentasikan satu data observasi yang telah ditransformasikan dari variabel asli ke dalam ruang dua dimensi.

Berdasarkan sebarannya, terlihat bahwa variasi data lebih dominan pada sumbu horizontal (PC1) dibandingkan sumbu vertikal (PC2). Hal ini menunjukkan bahwa *Principal Component 1* memiliki kontribusi variasi yang lebih besar

dibandingkan *Principal Component 2*. Dengan kata lain, sebagian besar informasi atau keragaman data berhasil dirangkum dalam komponen pertama.

Sebaran titik-titik pada grafik tampak membentuk satu kelompok besar tanpa pemisahan yang jelas. Hal ini mengindikasikan bahwa struktur data cenderung homogen dalam dua dimensi utama yang ditampilkan. Oleh karena itu, memerlukan metode *clustering* atau pengelompokan data tersebut, untuk mengidentifikasi dan membentuk kelompok data lebih terstruktur.

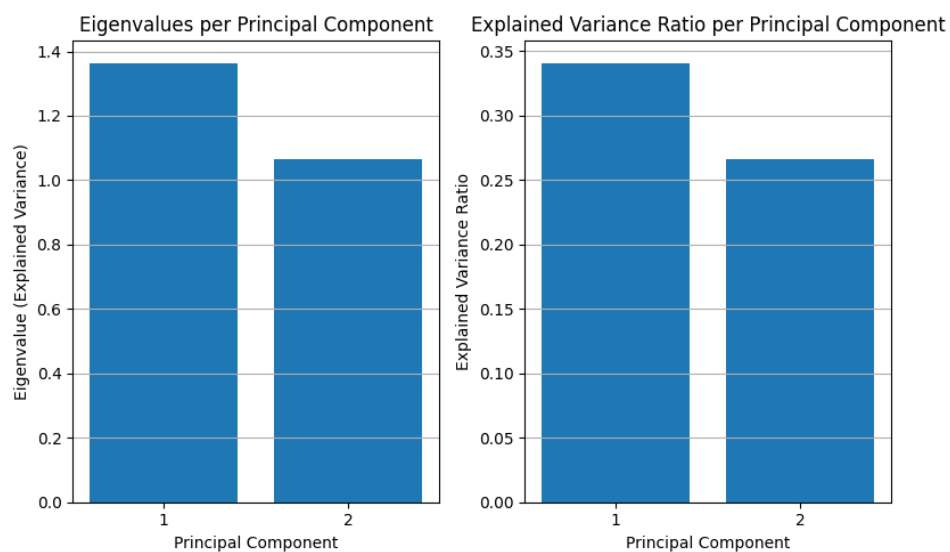
Selain itu, terdapat beberapa titik yang berada cukup jauh dari kumpulan utama data. Titik-titik tersebut dapat diindikasikan sebagai data ekstrem atau *outlier* yang memiliki karakteristik berbeda dibandingkan mayoritas data lainnya. Secara keseluruhan, visualisasi ini menunjukkan bahwa dua komponen utama hasil PCA telah mampu merepresentasikan pola umum data, namun belum memperlihatkan pemisahan kelompok yang signifikan.

4.4.2 Visualisasi *Bar Plot* Nilai Eigen dan Rasio Varians

Pada tahap ini, hasil visualisasi dua dimensi untuk melihat nilai eigen dan rasio varians. Berdasarkan grafik nilai *eigenvalue* dan *explained variance ratio*, terlihat bahwa *Principal Component 1* (PC1) memiliki nilai *eigenvalue* yang lebih besar dibandingkan *Principal Component 2* (PC2). Nilai *eigenvalue* PC1 berada di kisaran sekitar 1.36, sedangkan PC2 sekitar 1.06. Hal ini menunjukkan bahwa komponen pertama mampu menjelaskan variasi data yang lebih besar dibandingkan komponen kedua.

Berdasarkan dari grafik *explained variance ratio*, PC1 mampu menjelaskan sebesar 34.1% dari total variasi data, sementara PC2 menjelaskan sebesar 26.6%.

Explained variance ratio menunjukkan proporsi kontribusi masing-masing komponen terhadap keseluruhan variasi data. Apabila kedua komponen tersebut digabungkan, maka total variasi yang dapat dijelaskan mencapai 60.7%. Hal ini berarti bahwa lebih dari setengah informasi yang terkandung dalam data asli telah berhasil dirangkum hanya dengan menggunakan dua dimensi utama.



Gambar 4.2 Grafik *Bar Plot* Nilai Eigen & Rasio Varians

Pada Gambar 4.2 menunjukkan bahwa *Principal Component* 1 memiliki peran yang lebih signifikan dibandingkan *Principal Component* 2 dalam menjelaskan variasi data. Kedua komponen tersebut telah mampu menyederhanakan kompleksitas data ke dalam bentuk yang lebih ringkas tanpa menghilangkan sebagian besar informasi penting. Oleh karena itu, penggunaan dua komponen utama ini dapat dikatakan cukup representatif untuk menggambarkan pola umum data.

4.4.3 Visualisasi *Cluster* berdasarkan Label

Visualisasi tahap ini berguna untuk melihat pengelompokan atau *cluster* dari hasil PCA. Grafik ini menampilkan hasil visualisasi PCA yang telah diberi label berdasarkan hasil prediksi *clustering*, yaitu kategori “Dirasakan” dan “Tidak Dirasakan”. Setiap titik pada grafik merepresentasikan satu data observasi yang telah direduksi ke dalam dua dimensi utama, yaitu *Principal Component 1* (PC1) dan *Principal Component 2* (PC2). Warna merah menunjukkan data dengan label “Dirasakan”, sedangkan warna oranye menunjukkan data dengan label “Tidak Dirasakan”. Visualisasi grafik tersebut dapat dilihat pada Gambar 4.3.



Gambar 4. 3 Visualisasi *Cluster* berdasarkan Label

Pada Gambar 4.3 terlihat bahwa pemisahan antara kedua label cukup jelas terjadi pada sumbu *Principal Component 1*. Data dengan label “Tidak Dirasakan” cenderung terkonsentrasi pada nilai PC1 yang negatif (sebelah kiri grafik), sedangkan data dengan label “Dirasakan” lebih banyak berada pada nilai PC1 yang

positif (sebelah kanan grafik). Hal ini menunjukkan bahwa komponen utama pertama memiliki peran yang sangat dominan dalam membedakan kedua kelompok tersebut. Dengan kata lain, PC1 menjadi faktor pembeda utama antara data yang termasuk kategori “Dirasakan” dan “Tidak Dirasakan”.

Sementara itu, pada sumbu *Principal Component 2*, kedua kelompok terlihat masih saling tumpang tindih. Penyebaran data secara vertikal tidak menunjukkan pemisahan yang terlalu tegas antar label. Hal ini menunjukkan bahwa komponen kedua (PC2) lebih berperan dalam menginterpretasikan variasi internal dalam masing-masing kelompok, bukan sebagai pemisah utama antar kelompok.

Meskipun secara umum terdapat pemisahan yang cukup baik terutama pada arah horizontal, masih terdapat beberapa titik yang berada di area perbatasan atau bercampur di sekitar nilai PC1 mendekati nol. Titik-titik tersebut dapat diinterpretasikan sebagai data yang memiliki karakteristik yang relatif mirip antara kedua kategori, sehingga berada di wilayah transisi.

Secara keseluruhan, visualisasi ini menunjukkan bahwa proses reduksi dimensi menggunakan PCA berhasil mempertahankan struktur pemisahan data berdasarkan label prediksi. Komponen utama pertama terbukti efektif dalam merepresentasikan perbedaan karakteristik antara kategori “Dirasakan” dan “Tidak Dirasakan”, sehingga mendukung kualitas hasil *clustering* atau pengelompokan yang diperoleh dalam ruang dua dimensi PCA.

4.5 Evaluasi Model

Evaluasi kualitas pengelompokan dalam penelitian ini dilakukan menggunakan metode *Calinski–Harabasz Index* (CHI), yaitu untuk menilai kualitas

keseluruhan hasil *clustering*. Metrik ini mempertimbangkan dua aspek utama, yaitu tingkat kekompakan data dalam satu klaster (*cohesion*) dan tingkat keterpisahan antar klaster (*separation*), yang diukur berdasarkan perbandingan antara variasi antar klaster (*between-cluster dispersion*) dan variasi dalam klaster (*within-cluster dispersion*). Semakin tinggi rasio pada SSB dan SSW, semakin besar nilai indeks CH, yaitu semakin baik *clustering* tersebut. Nilai indeks CH yang lebih tinggi menunjukkan bahwa *cluster* yang terbentuk semakin padat (kompak) dan memiliki pemisahan yang baik antar *cluster*, meskipun tidak ada nilai batas yang secara pasti dianggap sebagai standar penerimaan.

Berdasarkan hasil perhitungan *Calinski–Harabasz Index* diperoleh nilai sebesar 5626.0269, yang merupakan bahwa struktur *cluster* yang terbentuk memiliki tingkat pemisahan yang tergolong baik. Nilai indeks CH yang lebih besar menunjukkan bahwa titik-titik data dalam setiap cluster memiliki kemiripan internal yang kuat dan jarak yang jelas terhadap *cluster* lain. Selain itu, jumlah 2 *cluster* juga sesuai dengan kategori label BMKG, yang hanya membagi kondisi gempa bumi ke dalam 2 kelas, yaitu “Tidak Dirasakan” dan “Dirasakan”.

4.6 Integrasi Islam

Pendekatan *clustering* dalam penelitian memiliki tujuan untuk mengelompokkan data berdasarkan hasil PCA atau pengelompokan data, hal ini serupa dengan bagaimana seluruh manusia dikelompokkan menjadi dua rombongan besar menuju tujuan akhir mereka di akhirat, yaitu rombongan ahli neraka dan rombongan ahli surga. Sebagaimana dijelaskan dalam Al-Qur'an, dalam Surah Az-Zumar ayat 72-73:

قِيلَ ادْخُلُوا أَبْوَابَ جَهَنَّمَ خَالِدِينَ فِيهَا بَعْثٌ مِّنْهُمُ الْمُتَكَبِّرِينَ ۖ وَسَيَقُولُ الَّذِينَ أَتَقَوْا رَبَّهُمْ إِلَىٰ الْجَنَّةِ زُمَرًا حَتَّىٰ إِذَا جَاءُوهَا وَفُتِحَتْ أَبْوَابُهَا وَقَالَ لَهُمْ خَزَنَتُهَا سَلَامٌ عَلَيْكُمْ طِبْتُمْ فَادْخُلُوهَا خَالِدِينَ ۚ ۛۛ

“Dikatakan (kepada mereka), “Masuklah pintu-pintu (neraka) Jahanam (untuk tinggal) di dalamnya selama-lamanya!” Maka, (neraka Jahanam) itu seburuk-buruk tempat tinggal bagi orang-orang yang takabur; orang-orang yang bertakwa kepada Tuhannya diantar ke dalam surga secara berombongan sehingga apabila mereka telah sampai di sana dan pintu-pintunya telah dibuka, para penjaganya berkata kepada mereka, “Salāmun ‘alaikum (semoga keselamatan tercurah kepadamu), berbahagialah kamu. Maka, masuklah ke dalamnya (untuk tinggal) selama-lamanya!” (QS. Az-Zumar: 72–73).

Dalam Tafsir Ibnu Katsir pada ayat tersebut menegaskan bahwa ‘Kelompok-kelompok’ dua kelompok manusia yang berbeda berdasarkan jenis dosa atau ideologi yang sama. Yang berdosa berkumpul dengan sesama pendosa, orang yg beriman dengan sesamanya, dan seterusnya. Surah ini mengajarkan bahwa hidup adalah serangkaian pilihan moral yang akan menentukan kelompok abadi mana seseorang akan bergabung. Pengelompokan ini, seperti yang telah dijelaskan, bukan acak, melainkan berdasarkan amal, keyakinan, dan pergaulan di dunia. Setiap bagian bersama bagian yang lain dan setiap kelompok sesuai dengan kelompok lainnya (Ghoffar, 2004).

Konsep ini selaras dengan metode *clustering*, data dikelompokkan berdasarkan tingkat kemiripan karakteristik tertentu. Dalam penelitian ini, data gempa dikelompokkan berdasarkan atribut seperti wilayah lokasi, kedalaman, dan magnitudo sehingga membentuk *cluster* dengan pola tertentu. Dengan demikian, analisis data gempa menggunakan metode PCA dan *clustering* tidak hanya digunakan dalam analisis ilmiah tetapi dapat dipandang sebagai bagian dari upaya ilmiah untuk mencerminkan keteraturan dan hukum alam (*sunnatullah*) yang telah

Allah SWT. tetapkan. Dalam Hadits Bukhari, Rasulullah ﷺ juga menegaskan bahwa:

أَحَبُّتَ مَنْ مَعَ أَنْتَ

“Engkau bersama dengan orang yang engkau cintai.” (HR. Sahih Bukhari: 6170).

Maksud hadits di atas adalah kita akan dikumpulkan bersama orang yang kita cintai itu, sehingga kita menjadi bagian dari kelompok orang-orang yang dicintai. Dalam *clustering*, objek yang memiliki kemiripan karakteristik akan berada dalam kelompok yang sama. Dalam hadits ini, manusia secara juga dikelompokkan berdasarkan kedekatan kecenderungan, nilai, dan kecintaan. Prinsip dasar *clustering* adalah mengelompokkan objek berdasarkan tingkat kemiripan atribut. Konsep ini memiliki keterkaitan dengan sabda Rasulullah SAW bahwa seseorang akan bersama dengan golongan yang ia cintai, yang menunjukkan adanya pengelompokan berdasarkan kesamaan kecenderungan dan karakteristik.

Pada penelitian ini, proses normalisasi dilakukan dengan tujuan untuk menyeragamkan rentang nilai antar atribut data gempa seperti latitude, longitude, *Depth*, dan magnitudo sehingga data menjadi lebih terstruktur dan optimal untuk dianalisis menggunakan PCA dan *clustering*. Konsep ini dapat dikaitkan dengan firman Allah SWT. dalam surah Al- Ma'idah ayat 6:

يَا أَيُّهَا الَّذِينَ آمَنُوا إِذَا قُمْتُمْ إِلَى الصَّلَاةِ فَاغْسِلُوا وُجُوهَكُمْ وَأَيْدِيَكُمْ إِلَى الْمَرَافِقِ وَامْسَحُوا بِرُءُوسِكُمْ وَأَرْجُلَكُمْ إِلَى الْكَعْبَيْنِ وَإِنْ كُنْتُمْ جُنُبًا فَاطَّهَّرُوا وَإِنْ كُنْتُمْ مَرْضَى أَوْ عَلَى سَفَرٍ أَوْ جَاءَ أَحَدٌ مِنْكُم مِّنَ الْغَائِطِ أَوْ لَمَسْتُمُ النِّسَاءَ فَلَمْ تَجِدُوا مَاءً فَتَيَمَّمُوا صَعِيدًا طَيِّبًا فَامْسَحُوا بِوُجُوْهِكُمْ وَأَيْدِيكُمْ مِنْهُ مَا يُرِيدُ اللَّهُ لِيَجْعَلَ عَلَيْكُمْ مِنْ حَرَجٍ وَلَكِنْ يُرِيدُ لِيُطَهِّرَكُمْ وَلِيُتِمَّ نِعْمَتَهُ عَلَيْكُمْ لَعَلَّكُمْ تَشْكُرُونَ ٦

“Wahai orang-orang yang beriman, apabila kamu berdiri hendak melaksanakan salat, maka basuhlah wajahmu dan tanganmu sampai ke siku serta usaplah

kepalamu dan (basuh) kedua kakimu sampai kedua mata kaki. Jika kamu dalam keadaan junub, mandilah. Jika kamu sakit, 202) dalam perjalanan, kembali dari tempat buang air (kakus), atau menyentuh 203) perempuan, lalu tidak memperoleh air, bertayamumlah dengan debu yang baik (suci); usaplah wajahmu dan tanganmu dengan (debu) itu. Allah tidak ingin menjadikan bagimu sedikit pun kesulitan, tetapi Dia hendak membersihkan kamu dan menyempurnakan nikmat-Nya bagimu agar kamu bersyukur.” (QS. Al-Ma'idah: 6).

Ayat tersebut menjelaskan pentingnya proses pembersihan atau penyucian sebelum melakukan suatu ibadah maupun kegiatan. Dalam Tafsir Ibnu Katsir menjelaskan bahwa ayat ini berkaitan dengan perintah bersuci seperti berwudhu, mandi, dan *tayamum* sebagai bentuk pensucian lahir maupun batin sebelum menjalankan ibadah (Ghoffar, 2004). Sebagaimana manusia diperintahkan untuk bersuci agar berada dalam kondisi yang lebih baik, begitu juga dengan data pada penelitian ini melalui tahap normalisasi agar lebih terstruktur dan siap untuk diproses pada tahap analisis berikutnya. Oleh karena itu, dilakukan normalisasi untuk membersihkan ketidakseimbangan skala data sehingga seluruh atribut memiliki kontribusi yang lebih proporsional.

Selain itu, integrasi ini menunjukkan bahwa ilmu pengetahuan dan teknologi memiliki keterkaitan dengan nilai-nilai Islam. Islam mengajarkan pentingnya kebersihan, keteraturan, dan persiapan sebelum melakukan suatu pekerjaan. Dalam konteks penelitian ini, normalisasi bukan hanya dipahami sebagai proses matematis, tetapi juga mencerminkan prinsip keteraturan dan penyempurnaan sebagaimana diajarkan dalam Al-Qur'an. Dengan demikian, penerapan normalisasi pada penelitian ini tidak hanya bertujuan meningkatkan kualitas analisis data, tetapi juga menggambarkan nilai keilmuan yang selaras

dengan ajaran Islam mengenai pentingnya proses pembersihan dan penyempurnaan sebelum mencapai hasil yang optimal.

4.7 Hasil

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan *Principal Component Analysis* (PCA) berhasil mereduksi dimensi data gempa yang semula terdiri dari empat variabel menjadi sejumlah komponen utama yang lebih ringkas namun tetap mampu mempertahankan sebagian besar informasi penting. Hasil analisis menunjukkan bahwa dua komponen utama pertama (PC1 dan PC2) telah mampu menjelaskan sebesar 60.7% variasi total data. Hal ini menunjukkan bahwa PCA efektif dalam menyederhanakan struktur data tanpa menghilangkan karakteristik utama data gempa.

Hasil reduksi dimensi menggunakan PCA kemudian dimanfaatkan sebagai dasar dalam proses clustering dengan menggunakan skor komponen utama sebagai representasi data. Visualisasi PCA dua dimensi (PC1–PC2) menunjukkan bahwa data memiliki pola penyebaran yang cukup terstruktur, dengan kecenderungan pemisahan antar kelompok terutama pada sumbu PC1. Terlihat bahwa label “Dirasakan” dan “Tidak Dirasakan” mulai membentuk dua kelompok yang relatif terpisah secara horizontal, meskipun masih terdapat beberapa area tumpang tindih di sekitar nilai tengah PC1. Hal ini menunjukkan bahwa komponen utama pertama (PC1) memiliki kontribusi yang lebih dominan dalam membedakan karakteristik data gempa dibandingkan komponen kedua (PC2), sementara PC2 lebih berperan dalam variasi internal masing-masing kelompok.

Tabel 4.6 Hasil Pelabelan

No.	PC1	PC2	Label
1.	1.004969	-0.352439	Dirasakan
2.	2.183259	-0.773692	Dirasakan
3.	-0.050502	0.699144	Tidak Dirasakan
4.	-1.842468	0.969191	Tidak Dirasakan
5.	-1.95835	1.019267	Tidak Dirasakan

Pada Tabel 4.6 menampilkan hasil pelabelan data gempa berdasarkan nilai dua komponen utama hasil reduksi dimensi menggunakan *Principal Component Analysis* (PCA), yaitu *Principal Component 1* (PC1) dan *Principal Component 2* (PC2). Setiap baris pada tabel merepresentasikan satu data observasi yang telah ditransformasikan dari variabel asli ke dalam bentuk skor komponen utama, kemudian diberikan label sesuai hasil pengelompokan atau klasifikasi, yaitu “Dirasakan” dan “Tidak Dirasakan”.

Berdasarkan tabel tersebut, terlihat bahwa data dengan nilai PC1 positif cenderung memperoleh label “Dirasakan”, seperti pada data nomor 1 dan 2 yang memiliki nilai PC1 sebesar 1.004969 dan 2.183259. Sebaliknya, data dengan nilai PC1 negatif cenderung termasuk dalam kategori “Tidak Dirasakan”, seperti pada data nomor 3, 4, dan 5 yang memiliki nilai PC1 kurang dari nol. Hal ini menunjukkan bahwa komponen utama pertama (PC1) memiliki peran dominan dalam membedakan karakteristik gempa yang “Dirasakan” dan “Tidak Dirasakan”.

Sementara itu, nilai PC2 pada tabel menunjukkan variasi tambahan yang membantu menjelaskan perbedaan antar data dalam masing-masing kategori. Hal ini sejalan dengan hasil analisis sebelumnya yang menunjukkan bahwa PC1 memiliki kontribusi variasi yang lebih besar dibandingkan PC2. Dari hasil pelabelan tersebut dapat disimpulkan bahwa proses reduksi dimensi menggunakan

PCA berhasil menyederhanakan struktur data tanpa menghilangkan pola pembeda utama. Nilai skor komponen utama yang dihasilkan mampu merepresentasikan karakteristik gempa secara lebih ringkas, sehingga proses pengelompokan menjadi lebih terarah dan terstruktur. Dengan demikian, tabel ini memperlihatkan bahwa pemisahan kategori “Dirasakan” dan “Tidak Dirasakan” dapat dijelaskan secara matematis melalui distribusi nilai pada komponen utama pertama.

Secara keseluruhan, kombinasi PCA sebagai metode reduksi dimensi dan *clustering* berbasis skor komponen utama terbukti efektif dalam mengidentifikasi pola utama data gempa. PCA berperan penting dalam menyederhanakan data dan menonjolkan struktur variasi utama, sementara *clustering* memanfaatkan struktur tersebut untuk membentuk kelompok data yang bermakna. Berdasarkan hasil evaluasi dan interpretasi gempa, model dengan dua komponen utama ditetapkan sebagai model yang digunakan pada penelitian ini karena memberikan keseimbangan optimal antara kualitas statistik, kejelasan struktur data, dan kemudahan interpretasi.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil pengujian dan analisis pada Bab IV, dapat disimpulkan bahwa proses *pre-processing*, reduksi dimensi menggunakan *Principal Component Analysis* (PCA), serta pengelompokan dengan metode *clustering* berhasil mengidentifikasi pola utama dalam data gempa bumi yang dianalisis. Tahap *pre-processing* yang meliputi normalisasi data dengan *StandardScale* menghasilkan *dataset* yang bersih, konsisten, dan terstandarisasi, sehingga perbedaan skala antar variabel tidak memengaruhi hasil analisis.

Hasil reduksi dimensi menggunakan PCA menunjukkan bahwa sebagian besar variasi data gempa dapat direpresentasikan oleh dua komponen utama pertama (PC1 dan PC2) dengan total kontribusi variasi yang signifikan. Komponen pertama (PC1) menjadi faktor dominan dalam menjelaskan keragaman data, sedangkan komponen kedua (PC2) memberikan kontribusi tambahan dalam menangkap variasi yang tidak terwakili oleh PC1. Hal ini menunjukkan bahwa PCA berhasil menyederhanakan kompleksitas data tanpa menghilangkan informasi penting yang terkandung di dalamnya.

Visualisasi PCA dalam dua dimensi mampu menggambarkan pola persebaran data gempa secara lebih jelas. Terlihat adanya kecenderungan pemisahan data berdasarkan karakteristik tertentu, terutama pada sumbu PC1. Data gempa yang sudah berlabel dirasakan dan tidak dirasakan menunjukkan pola

distribusi yang relatif berbeda, meskipun masih terdapat beberapa titik yang berada pada area perbatasan. Hal ini mengindikasikan bahwa komponen utama pertama memiliki peran penting dalam membedakan karakteristik kedua kategori tersebut.

Selanjutnya, proses *clustering* yang dilakukan terhadap data hasil reduksi dimensi menunjukkan bahwa struktur pengelompokan yang terbentuk mampu merepresentasikan perbedaan karakteristik gempa yang cukup baik. Berdasarkan hasil evaluasi menggunakan *Calinski–Harabasz Index*, diperoleh nilai yang menunjukkan tingkat kekompakan dan keterpisahan *cluster* yang optimal pada jumlah *cluster* tertentu. Nilai indeks CH yang lebih tinggi menunjukkan bahwa anggota dalam satu *cluster* memiliki kemiripan yang tinggi dan memiliki perbedaan yang cukup jelas dengan *cluster* lainnya.

Secara keseluruhan, hasil analisis menunjukkan bahwa kombinasi metode PCA dan *clustering* mampu memberikan pemahaman yang komprehensif terhadap struktur data gempa bumi. Metode ini tidak hanya menyederhanakan dimensi data, tetapi juga membantu mengidentifikasi pola utama yang relevan dalam membedakan karakteristik gempa yang dirasakan dan tidak dirasakan. Informasi yang diperoleh dari penelitian ini diharapkan dapat mendukung analisis kebencanaan, peningkatan kewaspadaan terhadap risiko gempa, serta pengambilan keputusan berbasis data dalam konteks mitigasi bencana.

5.2 Saran

1. Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk penelitian selanjutnya agar pengembangan analisis data gempa dapat dilakukan secara lebih mendalam

dan komprehensif. Penelitian selanjutnya dapat menggunakan *dataset* dengan rentang waktu yang lebih panjang atau cakupan wilayah yang lebih luas agar pola yang dihasilkan menjadi lebih representatif dan generalisasi model dapat ditingkatkan. Penambahan variabel lain seperti kedalaman gempa, jarak episentrum terhadap permukiman, atau kondisi geologi setempat juga dapat dipertimbangkan untuk memperoleh hasil pengelompokan yang lebih akurat.

2. Kedua, metode reduksi dimensi dan *clustering* dapat dikembangkan dengan membandingkan PCA dengan teknik lain seperti t-SNE atau UMAP untuk visualisasi, serta membandingkan algoritma clustering seperti K-Means, DBSCAN, atau *Hierarchical Clustering* guna melihat perbedaan performa dalam membentuk struktur *cluster*. Perbandingan ini dapat memberikan gambaran metode mana yang paling sesuai untuk karakteristik data gempa.

Dengan adanya pengembangan tersebut, diharapkan penelitian mengenai analisis dan pengelompokan data gempa dapat memberikan kontribusi yang lebih signifikan dalam mendukung sistem mitigasi bencana dan pengambilan keputusan berbasis data.

DAFTAR PUSTAKA

- Adiwijaya, Wisesty, U. N., Lisnawati, E., Aditsania, A., Kusumo, D. S., Di, P., Bojonegoro, K., Hedyati, D., & Suartana, I. M. (2021). Penerapan Principal Component Analysis (PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi. *Journal of Computer Science*, 14(11), 49–54.
- Ahmad, A., & Irsalinda, N. (2020). Cluster Analysis of Earthquake's Data Clustering in Indonesia using Fuzzy K-means Clustering. *Proceeding International Conference on Science and Engineering*, 3(April), 3–7. <https://doi.org/10.14421/icse.v3.457>
- AlKhaldi, K. H. (2024). 96 - *Earthquakes* (G. B. T.-C. D. M. (Third E. Ciottone, Ed.; pp. 601–604). Elsevier. <https://doi.org/https://doi.org/10.1016/B978-0-323-80932-0.00096-3>
- Almais, A. T. W., Susilo, A., Naba, A., Sarosa, M., Crysdian, C., Tazi, I., Hariyadi, M. A., Muslim, M. A., Basid, P. M. N. S. A., Arif, Y. M., Purwanto, M. S., Parwatiningtyas, D., Supriyono, & Wicaksono, H. (2023). Principal Component Analysis-Based Data Clustering for Labeling of Level Damage Sector in Post-Natural Disasters. *IEEE Access*, 11(May), 74590–74601. <https://doi.org/10.1109/ACCESS.2023.3275852>
- Almais, A. T. W., Susilo, A., Naba, A., Sarosa, M., Juwono, A. M., Crysdian, C., Muslim, M. A., & Wicaksono, H. (2024). Characterization of Structural Building Damage in Post-Disaster using GLCM-PCA Analysis Integration. *IEEE Access*, 12(September), 146190–146201. <https://doi.org/10.1109/ACCESS.2024.3469637>
- BNPB. (2024). Indeks Risiko Bencana Indonesia Tahun 2024. *Badan Nasional Penanggulangan Bencana*, 03, 1–184.
- Dartanto, T. (2022). Natural disasters, mitigation and household welfare in Indonesia: Evidence from a large-scale longitudinal survey. *Cogent Economics and Finance*, 10(1). <https://doi.org/10.1080/23322039.2022.2037250>
- Dinanti, A., & Purwadi, J. (2023). Analisis Performa Algoritma K-Nearest Neighbor dan Reduksi Dimensi Menggunakan Principal Component Analysis. *Jambura Journal of Mathematics*, 5(1), 155–165. <https://doi.org/10.34312/jjom.v5i1.17098>

- Enzellina, G., & Suhaedi, D. (2022). Penggunaan Metode Principal Component Analysis dalam Menentukan Faktor Dominan. *Jurnal Riset Matematika*, 101–110. <https://doi.org/10.29313/jrm.v2i2.1192>
- Feng, G., Lin, J., & Wang, K. (2022). Researches Advanced in Clustering Algorithms. *Highlights in Science, Engineering and Technology*, 16, 168–177. <https://doi.org/10.54097/hset.v16i.2498>
- Fuady, M., Munadi, R., & Fuady, M. A. K. (2021). Disaster mitigation in Indonesia: between plans and reality. *IOP Conference Series: Materials Science and Engineering*, 1087(1), 012011. <https://doi.org/10.1088/1757-899x/1087/1/012011>
- Ghoffar, M. A. (2004). Lubaabut Tafsir Min Ibni Katsiir. In *Pustaka Imam asy-Syafi'i* (p. Jilid (1-8)).
- Hadi, H., Agustina, S., & Subhani, A. (2019). Penguatan Kesiapsiagaan Stakeholder dalam Pengurangan Risiko Bencana Alam Gempabumi. *Geodika: Jurnal Kajian Ilmu Dan Pendidikan Geografi*, 3(1), 30. <https://doi.org/10.29408/geodika.v3i1.1476>
- Harijoko, A., Anggraini, A., Setianto, A., Hendratno, A., & ... (2024). *Geologi Gempa Bumi Indonesia*. books.google.com. <https://books.google.com/books?hl=en&lr=&id=ojPuEAAQBAJ&oi=fnd&pg=PA1&dq=gempa&ots=qf95JtN03j&sig=39GC7PdnKn0EHZrtQO9bkw411Gg>
- Hartatik, H., & Cahya, A. S. D. (2020). Clusterisasi Kerusakan Gempa Bumi di Pulau Jawa Menggunakan SOM. *Jurnal Ilmiah Intech: Information Technology Journal of UMUS*, 2(02), 25–34. <https://doi.org/10.46772/intech.v2i02.286>
- Hasan, B. M. S. (2021). A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *Journal of Soft Computing and Data Mining*, 2(1), 20–30. <https://doi.org/10.30880/jscdm.2021.02.01.003>
- Hasan, M. S. (2021). Triboinformatic modeling of dry friction and wear of aluminum base alloys using machine learning algorithms. *Tribology International*, 161. <https://doi.org/10.1016/j.triboint.2021.107065>
- Hu, J. J., Zhang, H., Jin, C. Y., Wang, Z. W., & Hu, L. (2021). A method to simulate ground motion based on PCA and PSO intelligent algorithms—a case study of moderate magnitude earthquakes in western China. *Gongcheng*

Lixue/Engineering Mechanics, 38(3), 159–168.
<https://doi.org/10.6052/j.issn.1000-4750.2020.05.0293>

Ikotun, A. M., Habyarimana, F., & Ezugwu, A. E. (2025). Cluster validity indices for automatic clustering : A comprehensive review. *Heliyon*, 11(2), e41953.
<https://doi.org/10.1016/j.heliyon.2025.e41953>

Inayah, J., Fanani, A., & Utami, W. D. (2024). Klasterisasi Data Kejadian Gempa Bumi di Indonesia Menggunakan Metode K-Medoids. *JUSTIN (Jurnal Sistem Dan* <https://jurnal.untan.ac.id/index.php/justin/article/view/73594>

Khairati, A. F., Adlina, A. A., Hertono, G. F., & Handari, B. D. (2019). *Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA*. 2, 161–170.

Mardhotillah, Bunga; Fadhil, Amril; Elisa, E., & Zurweni. (2023). *Indeks Calinski – Harabasz Analisis Fuzzy C – Means dan K – Means Cluster Kabupaten/Kota di Provinsi Jambi Menurut Potensi Pertambangan, Penggalian, Pengadaan Listrik, dan Gas*. 2(1), 16–22.

Mohammed, M. A., Akawee, M. M., Saleh, Z. H., Hasan, R. A., Ali, A. H., & Sutikno, T. (2023). The effectiveness of big data classification control based on principal component analysis. *Bulletin of Electrical Engineering and Informatics*, 12(1), 427–434. <https://doi.org/10.11591/eei.v12i1.4405>

Muddana, A. L., & Vinayakam, S. (2024). Data Visualizations. In *Python for Data Science* (pp. 229–254). Springer Nature Switzerland.
https://doi.org/10.1007/978-3-031-52473-8_12

Muzli, M., Masturyono, M., Murjaya, J., & Riyadi, M. (2017). Studi Awal Penyusunan Skala Intensitas Gempabumi Badan Meteorologi Klimatologi Dan Geofisika. In *Jurnal Meteorologi dan Geofisika* (Vol. 17, Number 2).
<https://doi.org/10.31172/jmg.v17i2.440>

Oktafiani, R., & Itje Sela, E. (2024). Breast Cancer Classification with Principal Component Analysis and Smote using Random Forest Method and Support Vector Machine. *International Journal of Computer Applications*, 186(16), 975–8887. <https://doi.org/10.5120/ijca2024923537>

Philia Christi Latue, & Heinrich Rakuasa. (2023). Pemanfaatan Data Penginderaan Jauh dan Sistim Informasi Geografis Untuk Identifikasi Perkembangan Lahan Terbangun pada Wilayah Rawan Gempa Bumi di Kota Ambon. *INSOLOGI: Jurnal Sains Dan Teknologi*, 2(3), 476–485.
<https://doi.org/10.55123/insologi.v2i3.1899>

- Prathivi, R. (2020). Optimasi Algoritme Naive Bayes Untuk Klasifikasi Data Gempa Bumi di Indonesia Berdasarkan Hiposentrum. *Telematika*, 13(1), 36–43. <https://doi.org/10.35671/telematika.v13i1.928>
- Priyatna, T. R., & Huda, M. (2026). *Cluster Analysis of Cities / Districts in West Kalimantan based on Stunting Response Indicators using the Calinski Harabasz Index*. 11(1), 1–15.
- Rahman, M. A. (2020). Employing PCA and t-statistical approach for feature extraction and classification of emotion from multichannel EEG signal. *Egyptian Informatics Journal*, 21(1), 23–35. <https://doi.org/10.1016/j.eij.2019.10.002>
- Rifa, I. H., Pratiwi, H., & Respatiwan, R. (2020). Clustering of Earthquake Risk in Indonesia Using K-Medoids and K-Means Algorithms. *Media Statistika*, 13(2), 194–205. <https://doi.org/10.14710/medstat.13.2.194-205>
- Saccenti, E. (2024). A gentle introduction to principal component analysis using tea-pots, dinosaurs, and pizza. *Teaching Statistics*, 46(1), 38–52. <https://doi.org/10.1111/test.12363>
- Setiawan, I. N., Krismawati, D., Pramana, S., & Tanur, E. (2022). Klasterisasi Wilayah Rentan Bencana Alam Berupa Gerakan Tanah Dan Gempa Bumi Di Indonesia. *Seminar Nasional Official Statistics*, 2022(1), 669–676. <https://doi.org/10.34123/semnasoffstat.v2022i1.1538>
- Sugiarto, E., Budiman, F., Muslih, M., Arifin, Z., Fahmi, A., & Hendriyanto, N. (2023). Peningkatan Fitur Ekstraksi Berbasis Discrete Wavelet Transform dan Principal Component Analysis Pada Pengenalan Citra Batik. *Jurnal Transformatika*, 20(2), 13. <https://doi.org/10.26623/transformatika.v20i2.5613>
- Supendi, P., Rohadi, S., Karnawati, D., & Nugraha, A. D. (2020). *Analysis of the destructive earthquakes end of West Java , Indonesia*. 02003, 1–6.
- Susanto, T. Y. (2023). On the earthquake distribution modeling in sumatra by cauchy cluster process: Comparing log-linear and log-Additive intensity models. *Sains Malaysiana*, 52(2), 655–667. <https://doi.org/10.17576/jsm-2023-5202-25>
- Troccoli, E. B., Cerqueira, A. G., Lemos, J. B., & Holz, M. (2022). K-means clustering using principal component analysis to automate label organization in multi-attribute seismic facies analysis. *Journal of Applied Geophysics*, 198, 104555. <https://doi.org/https://doi.org/10.1016/j.jappgeo.2022.104555>

- Uddin, Md. P., Mamun, Md. Al, & Hossain, Md. A. (2021). PCA-based Feature Reduction for Hyperspectral Remote Sensing Image Classification. *IETE Technical Review*, 38(4), 377–396. <https://doi.org/10.1080/02564602.2020.1740615>
- Wahyu, A., & Rushenda. (2022). Klasterisasi Dampak Bencana Gempa Bumi. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, 8(1), 175–179. <https://doi.org/10.26418/jp.v8i1>
- Wang, X., & Xu, Y. (2019). *An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index*. <https://doi.org/10.1088/1757-899X/569/5/052024>
- Wibowo, A. (2024). Clustering Indonesian Provinces by Disaster Intensity using K-Means Algorithm: a Data Mining Approach. *Disaster Advances*, 17(12), 1–8. <https://doi.org/10.25303/1712da0108>
- Zhang, M. (2022). *Unsupervised Learning Algorithms in Big Data: An Overview*. 910–931. https://doi.org/10.2991/978-2-494069-89-3_107