

**PREDIKSI KERUSAKAN BANGUNAN PASCAGEMPA BUMI  
BERBASIS *CLUSTERING***

**TESIS**

**Oleh:  
AHMAD ZARKONI  
NIM. 240605210011**



**PROGRAM STUDI MAGISTER INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2025**

**HALAMAN PENGAJUAN**  
**PREDIKSI KERUSAKAN BANGUNAN PASCAGEMPA BUMI**  
**BERBASIS *CLUSTERING***

**TESIS**

**Diajukan kepada:**

**Universitas Islam Negeri Maulana Malik Ibrahim Malang**  
**Untuk memenuhi Salah Satu Persyaratan dalam**  
**Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh:**  
**AHMAD ZARKONI**  
**NIM. 240605210011**

**PROGRAM STUDI MAGISTER INFORMATIKA**  
**FAKULTAS SAINS DAN TEKNOLOGI**  
**UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM**  
**MALANG**  
**2025**

**PREDIKSI KERUSAKAN BANGUNAN PASCAGEMPA BUMI  
BERBASIS *CLUSTERING***

**TESIS**

**Diajukan kepada:**

**Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk Memenuhi Salah Satu Persyaratan Dalam  
Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh:**

**AHMAD ZARKONI  
NIM. 240605210011**

**PROGRAM STUDI MAGISTER INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2025**

**HALAMAN PERSETUJUAN**  
**PREDIKSI KERUSAKAN BANGUNAN PASCAGEMPA BUMI**  
**BERBASIS *CLUSTERING***

**TESIS**

**Oleh:**  
**AHMAD ZARKONI**  
**NIM. 240605210011**

Telah diperiksa dan disetujui untuk diuji:  
Tanggal: 12 November 2025

Pembimbing I,

Pembimbing II,



Dr. Agung Teguh Wibowo Almais, M.T  
NIP. 19860301202321 1 016

Dr. Cahyo Crys dian, M.Cs  
NIP. 19740424 200901 1 008

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



Prof. Dr. H. Muhammad Faisal, S.Kom., M.T.  
NIP. 19740510 200501 1 007

**HALAMAN PENGESAHAN**  
**PREDIKSI KERUSAKAN BANGUNAN PASCAGEMPA BUMI**  
**BERBASIS *CLUSTERING***

**TESIS**

**Oleh:**  
**AHMAD ZARKONI**  
**NIM. 240605210011**

Telah Dipertahankan di Depan Dewan Penguji Tesis  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan\  
Untuk Memperoleh Gelar Magister Komputer (M.Kom)  
Tanggal: 12 November 2025

**Susunan Dewan Penguji**

Penguji I	: <u>Dr. Mokhamad Amin Hariyadi, M.T</u> NIP. 19670018 200501 1 001	
Penguji II	: <u>Dr. Usman Pagalay, M.Si</u> NIP. 19650414 200312 1 001	
Pembimbing I	: <u>Dr. Agung Teguh Wibowo Almais, M.T</u> NIP. 19860301 202332 11 016	
Pembimbing II	: <u>Dr. Cahyo Crysdian, M.Cs</u> NIP. 19740424 200901 1 008	

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



Prof. Dr. Muhammad Faisal, S.Kom., M.T.  
NIP. 19740510 200501 1 007

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Name : Ahmad Zarkoni  
NIM : 240605210011  
Program Studi : Magister Informatika  
Fakultas : Sains dan Teknologi

Menyatakan dengan sebenarnya bahwa Tesis yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan 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 Tesis ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 12 November 2025

Yang Membuat Pernyataan,



Ahmad Zarkoni  
240605210011

**MOTTO**

رَضَا الشَّيْخِ أَفْضَلُ مِنْ عِلْمِهِ

**"Ridho Seorang Guru Jauh Lebih Mahal Dari Pada Ilmunya"**

## HALAMAN PERSEMBAHAN

Assalamu'alaikum Wr. Wb.

Kupersembahkan karya Tesis terbaikku ini kepada:

1. Istri tercinta UM, Zahro NA Lirboyo, Zahro TA Ploso serta seluruh keluarga besar, yang telah memberikan dukungan moral, materiel, dan doa yang tiada henti sehingga menjadi kekuatan terbesar bagi penulis.
2. Sahabat-sahabat dan budies friend Tomy Ivan Sugiharto, serta rekan-rekan seperjuangan Syahreni, Holidiyatul Amani, Fanny Brawijaya, Ilham, Riska Yulinda, Samsul H Wiyono, yang telah menjadi teman diskusi dan pemberi semangat dalam suka maupun duka selama menempuh studi.

Wassalamu'alaikum Wr. Wb.

Malang, 12 November 2025



Ahmad Zarkoni  
NIM. 240605210011

## KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Segala puji dan syukur penulis panjatkan kehadirat Allah SWT, Tuhan semesta alam, yang atas rahmat, taufik, dan hidayah-Nya, penulis dapat menyelesaikan penyusunan Tesis. Tesis ini disusun sebagai salah satu syarat untuk memperoleh gelar Magister (S2) pada Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Penyusunan Tesis ini tidak akan terwujud tanpa bimbingan, arahan, motivasi, dan bantuan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis ingin menyampaikan ucapan terima kasih yang tak terhingga kepada:

1. Bapak Dr. H. Agus Mulyono, M.Si, selaku Dekan Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang.
2. Bapak Prof. Dr. Muhammad Faisal, M.T, selaku Ketua Program Studi Informatika UIN Maulana Malik Ibrahim Malang.
3. Bapak Dr. Agung Teguh Wibowo Almais, M.T., selaku Dosen Pembimbing I, yang telah meluangkan waktu, tenaga, dan pikiran dengan penuh kesabaran dalam memberikan arahan, bimbingan, dan motivasi yang tak ternilai harganya selama proses penelitian ini.
4. Bapak Dr. Cahyo Crysdiyan, M.Cs., selaku Dosen Pembimbing II, yang telah memberikan masukan, koreksi, dan wawasan berharga demi kesempurnaan penelitian ini.
5. Para penguji Bapak Dr. Mokhamad Amin Hariyadi, M.T., Dr. Usman Pagalay, M.Si.
6. Bapak Dr. Totok Chamidy, M.Kom., selaku dosen wali Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang, atas ilmu dan pelayanan yang telah diberikan.
7. Seluruh pihak yang tidak dapat penulis sebutkan satu per satu, yang telah memberikan bantuan baik secara langsung maupun tidak langsung.

Penulis menyadari sepenuhnya bahwa Tesis ini masih jauh dari kesempurnaan, baik dari segi substansi maupun metodologi. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan untuk perbaikan di masa depan.

Akhir kata, semoga Tesis ini dapat memberikan manfaat dan kontribusi, baik secara teoretis bagi pengembangan ilmu pengetahuan di bidang Informatika dan Geofisika, maupun secara praktis bagi masyarakat dan lembaga terkait.

Wassalamu'alaikum Wr. Wb.

Malang, 12 November 2025



Ahmad Zarkoni  
NIM. 240605210011

## DAFTAR ISI

HALAMAN PENGANTAR.....	i
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
PERNYATAAN KEASLIAN TULISAN .....	v
MOTTO .....	vi
HALAMAN PERSEMBAHAN .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xii
DAFTAR TABEL.....	xiii
ABSTRAK.....	xiv
ABSTRACT.....	xv
مستخلص البحث.....	xvi
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang.....	1
1.2 Pernyataan Masalah .....	5
1.3 Tujuan Penelitian .....	6
1.4 Hipotesis .....	6
1.5 Manfaat Penelitian .....	6
1.6 Batasan Masalah .....	7
1.7 Sistematika Penulisan .....	7
BAB II TINJAUAN PUSTAKA.....	9
2.1 <i>Prediction-based Clustering</i> .....	9
2.2 Kerangka Teori .....	19
2.3 Integrasi Al-Qur'an, Tafsir dan Hadis .....	25
2.4 Gempa Bumi.....	27
2.5 Gempa Merusak Wilayah Bawean .....	28
BAB III DESAIN DAN INTEGRASI SISTEM PENELITIAN .....	31
3.1 Kerangka Konsep.....	31
3.2 <i>Long Short Term Memory (LSTM)</i> .....	31
3.2.1 Formulasi Matematis Model LSTM .....	36
3.3 <i>Principal Component Analysis (PCA)</i> .....	38
3.4 Persiapan Data .....	40

3.5	Desain Sistem .....	41
3.5.1	Step 1: Inisialisasi dan Pemuatan Data .....	43
3.5.2	Step 2: Pra-Pemrosesan Data (Persiapan LSTM).....	43
3.5.3	Step 3: Proses Model Prediksi (LSTM).....	44
3.5.4	Step 4: <i>Input</i> Fitur dan Normalisasi (Persiapan PCA).....	45
3.5.5	Step 5: Proses PCA- <i>Clustering</i> .....	45
3.6	Skenario Evaluasi .....	46
3.7	Validasi Internal ( <i>Prediction</i> ).....	47
3.8	Instrument Penelitian .....	50
BAB IV UJI COBA .....		51
4.1	Skenario Uji Coba.....	51
4.2	Hasil Pelatihan Model Uji Coba .....	53
4.3	Model Matematis LSTM .....	65
4.4	Pembahasan Performa Model LSTM .....	67
4.5	Penentuan Model <i>Prediction</i> Terbaik.....	68
4.6	Implikasi untuk Tahap <i>Clustering</i> .....	69
4.7	Hasil <i>Fase Clustering (Prediction-based Clustering)</i> .....	69
4.8	Persiapan Data <i>Input</i> (Hasil Prediksi LSTM).....	69
4.9.1	Implementasi <i>Principal Component Analysis</i> (PCA).....	70
4.9.2	Penentuan Kluster ( <i>Clustering Logic</i> ) .....	71
4.9.3	Hasil dan Visualisasi Kluster.....	72
4.9.4	Perhitungan Matematis <i>Silhouette Score</i> .....	73
4.9	Validasi Kluster PCA dengan Skala MMI .....	75
4.10.1	Hasil Pemetaan Kluster PC1 dan Label MMI .....	75
4.10.2	Analisis dan Interpretasi Validasi MMI .....	76
4.10	Validasi Eksternal (Data Gempa Bumi Dirasakan) .....	77
BAB V PENUTUP.....		80
5.1	Kesimpulan .....	80
5.1.1	Integrasi Metode Prediksi LSTM dengan PCA- <i>Clustering</i> .....	80
5.1.2	Tingkat Efisiensi Integrasi LSTM dengan PCA- <i>Clustering</i> .....	81
5.1.3	Identifikasi Faktor-Faktor yang Memengaruhi Kerusakan Bangunan Pascagempa Bumi.....	81
5.2	Saran .....	82
5.2.1	Saran Praktis (Implikasi) .....	82
5.2.2	Saran untuk Penelitian Masa Depan .....	83

DAFTAR PUSTAKA .....	84
----------------------	----

## DAFTAR GAMBAR

Gambar 1.1 Peta Seismisitas PGR VII 2024.....	2
Gambar 2.1 Kerangka teori .....	20
Gambar 2.2 Perubahan dan Pemulihan Elastisitas .....	27
Gambar 2.3 Peta Pulau Bawean .....	29
Gambar 2.4 Peta Struktur Seismik Pulau Bawean .....	29
Gambar 3.1 Kerangka Konsep .....	31
Gambar 3.2 Arsitektur <i>Long Short-Term Memory</i> (LSTM) .....	32
Gambar 3.3 Alir Perhitungan <i>Forget Gate</i> .....	33
Gambar 3.3 Alir <i>Input Gate</i> Perkalian Bagian A dan Bagian B .....	33
Gambar 3.4 Alir Penambahan <i>Forget Gate</i> dan <i>Input Gate</i> .....	33
Gambar 3.5 Alir <i>Output Gate</i> ( <i>Sigmoid</i> ) dan <i>Tanh</i> .....	34
Gambar 3.9 Alir <i>Principal Component Analisis</i> (PCA)- <i>Clustering</i> .....	38
Gambar 3.6 Desain Sistem .....	42
Gambar 4.1 Sebaran Gempa Bawean (sumber: BMKG) .....	52
Gambar 4.2 Visualisasi Magnitudo, Kedalaman, dan Jarak .....	53
Gambar 4.3 <i>Eigen Values</i> dan <i>Rasio Varians</i> .....	71
Gambar 4.5 Visualisasi Kluster PCA (PC1 vs PC2).....	72
Gambar 4.6 Label MMI dengan PC1 Berdasarkan Magnitudo .....	76
Gambar 4.7 Peta Seismisitas Gempa Bumi Dirasakan Wilayah PGR VII 2024 ...	77

## DAFTAR TABEL

Tabel 2.1 Kerangka Teori LSTM PCA- <i>clustering</i> .....	22
Tabel 3.1 Data Gempa Bumi.....	41
Tabel 3.2 Keterangan Variabel data.....	41
Tabel 4.1 Parameter Model Skenario A.....	54
Tabel 4.2 Hasil Evaluasi Model Skenario A.....	54
Tabel 4.3 Parameter Model Skenario B.....	57
Tabel 4.4 Hasil Evaluasi Model Skenario B.....	58
Tabel 4.5 Parameter Model Skenario C.....	61
Tabel 4.6 Hasil Evaluasi Model Skenario C.....	61
Tabel 4.7 Perbandingan Model Terbaik Antar Skenario.....	67
Tabel 4.8 Hasil Normalisasi Menggunakan <i>StandardScaler</i> .....	70
Tabel 4.9 Jumlah Hasil Kluster Berdasarkan PC1.....	72
Tabel 4.10 Data Sampel.....	73
Tabel 4.11 Hasil <i>Silhouette Score</i> .....	74
Tabel 4.10 Pemetaan Hasil Kluster PC1 terhadap Label MMI.....	76
Tabel 4.12 Daftar 12 Titik Gempa Bumi Dirasakan Wilayah PGR VII 2024.....	77

## ABSTRAK

Zarkoni, Ahmad. 2025. **Prediksi Kerusakan Bangunan Pascagempa Bumi Berbasis *Clustering***. Tesis. Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing (1) Dr. Agung Teguh Wibowo Almais, M.T (II) Dr. Cahyo Crysdiyan, M.Cs.

Kata Kunci: Gempa Bumi, Penilaian Kerusakan, *Prediction-based Clustering*, LSTM, *Principal Component Analysis* (PCA), Skala MMI, Pembelajaran Mesin.

Kerusakan bangunan pascagempa bumi merupakan tantangan mitigasi bencana yang memerlukan identifikasi dini yang cepat dan akurat. Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi efisiensi model hibrida *Prediction-based Clustering* dengan mengintegrasikan *Long Short-Term Memory* (LSTM) untuk prediksi time-series dan *Principal Component Analysis* (PCA)-*Clustering* untuk pengelompokan tanpa pengawasan, guna menentukan tingkat kerusakan bangunan berdasarkan parameter seismologis. Data gempa yang meliputi Magnitudo (M), Kedalaman (D), dan Jarak *epicenter* (R) digunakan sebagai input utama. Metode LSTM (Model C3) berhasil memodelkan pola sekuensial variabel dengan kinerja prediksi yang solid (Test  $R^2$ : 0.5380). Hasil prediksi tersebut kemudian menjadi input untuk PCA-*Clustering*, yang berhasil mengelompokkan data menjadi tiga kluster tingkat kerusakan (ringan, sedang, dan Berat) berdasarkan nilai *Principal Component* 1 (PC1). Hasil paling signifikan adalah tingkat validitas eksternal model yang baik. Ditemukan kecocokan 1-ke-1 yang sempurna (presisi 100%) antara 12 kejadian yang diklusterisasi oleh model sebagai kluster "Dampak Rusak Berat" dengan 12 kejadian yang secara independen dicatat oleh BMKG sebagai "Gempa Bumi Dirasakan" (kejadian gempa kritis). Temuan ini membuktikan bahwa integrasi LSTM PCA-*Clustering* sangat efisien dan memiliki utilitas praktis yang substansial dalam memfilter dan mengisolasi kejadian gempa yang berpotensi menyebabkan kerusakan signifikan, menjadikannya alat yang dapat diandalkan untuk sistem peringatan dini dan penilaian dampak bencana.

## ABSTRACT

Zarkoni, Ahmad. 2025. **Prediction Of Building Damage After An Earthquake Based On Clustering**. Thesis. Master of Informatics study Program Faculty of Science and Technology State Islamic University Maulana Malik Ibrahim Malang. Advisor (I) Dr. Agung Teguh Wibowo Almais, M.T (II) Dr. Cahyo Crysdiyan, M.Cs.

Keywords: Earthquake, Damage Assessment, Prediction-based Clustering, LSTM, Principal Component Analysis (PCA), MMI Scale, Machine Learning.

Post-earthquake building damage constitutes a critical challenge for disaster mitigation, requiring prompt and accurate early identification. This study aims to develop and evaluate the efficiency of a hybrid Prediction-based Clustering model that integrates Long Short-Term Memory (LSTM) for time-series prediction and Principal Component Analysis (PCA)-Clustering for unsupervised grouping, to determine building damage levels based on seismological parameters. Earthquake data including Magnitude (M), Depth (D), and Epicentral Distance (R) were utilized as the primary inputs. The LSTM method (Model C3) successfully modeled the sequential patterns of the variables with solid predictive performance (Test  $R^2$ : 0.5380). The resulting predictions were then fed into PCA-Clustering, which successfully grouped the data into three damage level clusters (Light, Moderate, and Severe) based on the Principal Component 1 (PC1) values. The most significant finding is the exceptional level of external validity. A perfect 1-to-1 match (100% precision) was found between the 12 occurrences classified by the model as the "Severe Damage Impact" cluster and the 12 occurrences independently recorded by BMKG (Indonesia's Meteorology, Climatology, and Geophysical Agency) as "Felt Earthquakes" (critical earthquake events). This finding proves that the LSTM-PCA-Clustering integration is highly efficient and possesses substantial practical utility in filtering and isolating earthquake events that potentially cause significant damage, establishing it as a reliable tool for early warning systems and disaster impact assessment.

## مستخلص البحث

زاركوني، أحمد. 2025. التنبؤ بأضرار المباني بعد الزلزال بناءً على التجميع. أطروحة. برنامج ماجستير المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية في مالانج المشرفان: (I) د. أغونغ يوغو ويوووو المئيس، م.ت (II) د. كاهيو كريسديان، م.ت.

الكلمات المفتاحية: الزلازل، تقييم الأضرار، التجميع القائم على التنبؤ، LSTM، تحليل المكونات الرئيسية (PCA)، مقياس MMI، التعلم الآلي.

تعد أضرار المباني بعد الزلزال تحديًا حاسمًا للتخفيف من حدة الكوارث، وتتطلب تحديدًا مبكرًا وسريعًا ودقيقًا. تهدف هذه الدراسة إلى تطوير وتقييم كفاءة نموذج هجين للتصنيف العنقودي القائم على التنبؤ (*Prediction-based Clustering*) من خلال دمج تقنية الذاكرة طويلة المدى القصيرة (*LSTM*) للتنبؤ بالسلاسل الزمنية، وتقنية التصنيف العنقودي بتحليل المكونات الرئيسية (*PCA-Clustering*) للتجميع غير المراقب، لتحديد مستوى أضرار المباني بناءً على المعايير الزلزالية. تم استخدام بيانات الزلازل التي تشمل الحجم (*M*)، والعمق (*D*)، والمسافة المركزية (*R*) كمدخلات رئيسية. نجحت طريقة (*LSTM*) (النموذج *C3*) في نمذجة الأنماط المتسلسلة للمتغيرات بأداء تنبؤي قوي (مربع *R* للاختبار: 0.5380). وتم بعد ذلك استخدام نتائج التنبؤ كمدخلات لـ *PCA-Clustering*، التي نجحت في تجميع البيانات إلى ثلاثة مستويات من الأضرار (خفيفة، ومتوسطة، وخطيرة) بناءً على قيم المكون الرئيسي الأول (*PC1*). كانت النتيجة الأكثر أهمية هي المستوى الاستثنائي للصلاحية الخارجية للنموذج. حيث تم العثور على تطابق مثالي (دقة 100%) بين 12 حالة صنفها النموذج على أنها عنقود "تأثير الضرر الخطير" و12 حالة سجلتها *BMKG* (وكالة الأرصاد الجوية والمناخ والجيوفيزياء الإندونيسية) بشكل مستقل على أنها "زلزال محسوس" (أحداث زلزال حرجة). تثبت هذه النتيجة أن دمج (*LSTM-PCA-Clustering*) يتميز بكفاءة عالية ويمتلك فائدة عملية جوهرية في تصفية وعزل أحداث الزلازل التي من المحتمل أن تسبب أضرارًا كبيرة، مما يجعله أداة موثوقة لأنظمة الإنذار المبكر وتقييم آثار الكوارث.

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Bencana alam termasuk fenomena alam yang menimbulkan berbagai dampak negatif seperti kerusakan infrastruktur. Persiapan rehabilitasi dan rekonstruksi pascabencana dilakukan untuk menilai tingkat kerusakan dan kerugian yang harus ditangani pemerintah, sehingga penanganan dapat disesuaikan dengan kondisi nyata di lapangan dan mencegah dampak yang lebih besar (Almais et al., 2020).

Dalam paradigma integrasi ilmu (Pohon Ilmu), fenomena alam dipahami sebagai ayat *kauniyah* (tanda kebesaran Allah) sekaligus *sunnatullah* (hukum alam) yang dapat dipelajari. Al-Quran secara eksplisit menggambarkan fenomena gempa bumi dalam surat Al-Hadid ayat 22 dan Al-Zalzalah ayat 1-2:

مَا أَصَابَ مِنْ مُصِيبَةٍ فِي الْأَرْضِ وَلَا فِي أَنْفُسِكُمْ إِلَّا فِي كِتَابٍ مِنْ قَبْلِ أَنْ نَبْرَأَهَا إِنَّ ذَلِكَ عَلَى اللَّهِ يَسِيرٌ ٢٢

Artinya: “Tidak ada bencana (apa pun) yang menimpa di Bumi dan tidak (juga yang menimpa) dirimu, kecuali telah tertulis dalam Kitab (Lauh Mahfuz) sebelum Kami mewujudkannya. Sesungguhnya hal itu mudah bagi Allah.” (Q.S. Al-Hadid: 22)

إِذَا زُلْزِلَتِ الْأَرْضُ زِلْزَالَهَا وَأَخْرَجَتِ الْأَرْضُ أَثْقَالَهَا ٢

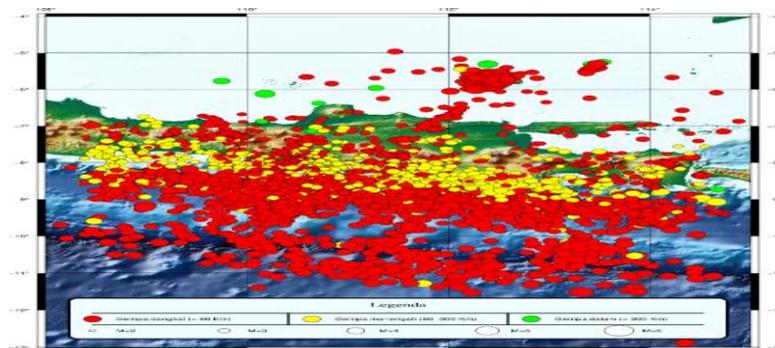
Artinya “Apabila bumi diguncangkan dengan guncangan yang dahsyat, bumi mengeluarkan isi perutnya, (Q.S. Al-Zalzalah: 1-2)

Menurut kitab *Al-Tafsir al-Wajiz* karya al-Wahidi, segala peristiwa di alam, termasuk gempa, banjir, dan berbagai bencana, serta musibah yang menimpa diri manusia, semuanya telah ditetapkan dan tertulis di *Lauh Mahfuz* sebelum terjadi. Semua itu berada dalam ketentuan Allah dan sangatlah mudah bagi-Nya untuk mewujudkannya. Dan menurut Wahbah bin Musthafa Az-Zuhaili dalam kitab *At-Tafsirul Munir*, termasuk tanda-tanda Kiamat adalah guncangan bumi yang sangat

dahsyat dan dengan keluarnya isi perut bumi berupa benda-benda terpendam dan orang-orang mati.

Secara ilmiah, manifestasi dari guncangan dahsyat tersebut salah satunya terjadi karena Indonesia terletak di pertemuan tiga lempeng tektonik utama dunia: Eurasia, Indo-Australia, dan Pasifik, serta lempeng mikro Filipina di utara Sulawesi. Kondisi ini menjadikan Indonesia rentan terhadap gempa tektonik yang terkadang disertai tsunami. Pencatatan gempa signifikan dan merusak menjadi krusial untuk pemetaan sejarah kegempaan wilayah. Berdasarkan hasil monitoring BMKG, sejak tahun 2013 tercatat tren peningkatan aktivitas gempa bumi di Indonesia, dari kisaran 4.000–5.000 kejadian per tahun menjadi lebih dari 10.000 kejadian per tahun (Ekaptiningrum, 2024). Lonjakan signifikan pada tahun 2018 dan 2019 berkorelasi dengan beberapa gempa kuat dan merusak (Senkaya et al., 2024).

Berdasarkan data Stasiun Geofisika Malang mencatat 8.576 kejadian gempa bumi di Jawa Tengah dan Jawa Timur sepanjang tahun 2024 (Buletin Gempa Bumi Tahunan, BMKG Malang). Gambar 1.1 menyajikan peta seismisitas sebaran gempa bumi berdasarkan magnitudo dan kedalaman.



Gambar 1.1 Peta Seismisitas PGR VII 2024

Berdasarkan Gambar 1, terlihat bahwa seismisitas gempa bumi di Jawa Timur

dan Jawa Tengah tahun 2024 diklasifikasikan berdasarkan magnitudo dan kedalaman. Titik merah menunjukkan gempa dangkal (kedalaman  $<60$  km), kuning menunjukkan gempa menengah (60–300 km), dan hijau menunjukkan gempa dalam ( $>300$  km).

Beberapa penelitian terdahulu telah mengkaji pengelompokan tingkat kerusakan bangunan pascagempa. Pertama, Bachriwindi et al. (2019) mengidentifikasi lima parameter kerusakan bangunan pascabencana: kondisi bangunan, struktur, kerusakan fisik, fungsi, dan status bangunan meskipun tidak spesifik pada jenis bencana tertentu. Selanjutnya, Almais et al. (2024) mengintegrasikan metode *Gray-Level Co-occurrence Matrix* (GLCM) dan *Principal Component Analysis* (PCA) untuk mengelompokkan kerusakan bangunan dalam tiga kategori: ringan, sedang, dan berat. Validasi dengan membandingkan jarak piksel 0,5 dan 1 mengungkapkan akurasi lebih tinggi pada jarak 0,5 piksel, dengan klaster kerusakan berat dominan pada koordinat tertentu.

Teknik ini terbukti mampu merepresentasikan kondisi kerusakan bangunan pascabencana secara akurat melalui parameter kedalaman. Penelitian Almais et al. (2023) lebih lanjut menunjukkan bahwa PCA berhasil mengelompokkan data kerusakan ke dalam tiga klaster: rusak ringan ( $n < 0$ ), rusak sedang ( $0 \leq n < 2$ ), dan rusak berat ( $n \geq 2$ ). Validasi terhadap data target asli mengonfirmasi bahwa *Principal Component 1* (PC1) menghasilkan distribusi data lebih baik dibanding PC2, membuktikan efektivitas PCA sebagai metode pelabelan data tidak terawasi untuk klasifikasi kerusakan pascabencana. Namun, penelitian ini masih terbatas karena tidak spesifik pada jenis bencana tertentu. Berbeda dengan pendekatan sebelumnya, Senkaya et al. (2024) secara khusus memfokuskan analisis pada

bencana gempa bumi dengan parameter tanggal, *origin time*, magnitudo, kedalaman, dan jarak. Integrasi metode *prediction* dan *clustering* dalam penelitian ini menjadi acuan untuk memprediksi tingkat kerusakan bangunan pascagempa (Rifaldi & Ahdika, 2024; Senkaya et al., 2024).

Penelitian ini mengadopsi pendekatan Almais et al. (2024) dalam pemodelan dampak kerusakan bangunan pascabencana menggunakan *PCA-Clustering*, serta memperluas aplikasinya untuk wilayah Jawa Timur dan Jawa Tengah. Hasilnya diharapkan dapat mendukung perencanaan pembangunan di daerah rawan gempa sebagai bentuk mitigasi risiko (Liu et al., 2024; Scala et al., 2024). Sementara itu penelitian Manur et al. (2018) membahas tektonik tersier terkini di wilayah Bawean, menunjukkan bahwa regangan bersifat tidak homogen secara regional dan wilayah tersebut telah terbagi menjadi beberapa domain yang dipisahkan oleh patahan geser utama dan zona ekstensional (Manur, 2018). Elemen struktural utama yang terjadi dari barat ke timur adalah Palung Muriah, Lengkungan Bawean. Kontribusi penelitian terletak pada penerapan *prediction-based clustering* untuk memprediksi tingkat kerusakan bangunan pascagempa bumi berbasis data historis. Metode ini diharapkan mampu meningkatkan efektivitas perencanaan tanggap darurat dan pengambilan keputusan mitigasi bencana, bagi BMKG, pendekatan ini dapat menjadi alat prediksi dan pengelompokan kerusakan yang optimal dalam mendukung strategi penanganan bencana.

Dalam konteks ilmiah, masa pascabencana dapat diwujudkan melalui metode *prediction*. Oleh karena itu, *prediction-based clustering* tingkat kerusakan bangunan pascagempa bumi diupayakan sebagai solusi ikhtiar (usaha) ilmiah dan mitigasi berbasis data. Hasil penelitian ini diharapkan dapat dimanfaatkan untuk

perencanaan pembangunan di Jawa Timur dan Jawa Tengah, wilayah dengan aktivitas seismik tinggi sebagai bentuk kesiapsiagaan terhadap potensi gempa merusak.

## 1.2 Pernyataan Masalah

Keterbatasan metode konvensional dalam penilaian kerusakan bangunan pascagempa: Metode penilaian kerusakan bangunan pascagempa seperti di Pulau Bawean sering lambat, subjektif, dan membutuhkan banyak sumber daya manusia, sehingga sulit menentukan prioritas respons darurat dan rekonstruksi secara cepat, yang memerlukan integrasi metode prediksi seperti LSTM dengan *PCA-Clustering* untuk penentuan tingkat kerusakan yang lebih efisien.

Kesenjangan dalam integrasi teknik prediksi dan klasterisasi untuk penilaian otomatis: Meskipun metode *deep learning* seperti LSTM digunakan untuk prediksi dan *clustering* untuk pengelompokan, belum ada integrasi yang baik antara keduanya untuk penilaian kerusakan bangunan secara otomatis, sehingga perlu pengukuran efisiensi integrasi LSTM dengan *PCA-Clustering* dalam mengukur tingkat kerusakan pascagempa.

Kurangnya identifikasi faktor-faktor penyebab kerusakan bangunan: Faktor-faktor yang memengaruhi kerusakan bangunan pascagempa belum diidentifikasi secara menyeluruh, terutama dalam konteks variasi geografis dan seismik seperti di Indonesia, sehingga penelitian ini bertujuan mengidentifikasi faktor-faktor tersebut untuk mendukung penilaian yang lebih akurat.

### 1.3 Tujuan Penelitian

1. Mengintegrasikan metode *Prediction* yaitu LSTM dengan *PCA-Clustering* untuk menentukan tingkat kerusakan bangunan pascagempa bumi.
2. Mengetahui tingkat efisiensi integrasi LSTM dengan *PCA-Clustering* untuk mengukur tingkat kerusakan bangunan pascagempa bumi.
3. Mengidentifikasi faktor-faktor yang memengaruhi kerusakan bangunan pascagempa bumi.

### 1.4 Hipotesis

Berdasarkan perumusan masalah, hipotesis penelitian ini adalah sebagai berikut:

1. H1 (Hipotesis Alternatif)

Model *hybrid* yang mengintegrasikan metode *Long Short-Term Memory* (LSTM) untuk *prediction* dan *PCA-Clustering* untuk klasterisasi memiliki akurasi yang lebih tinggi dan signifikan secara statistik dalam memprediksi serta mengkluster tingkat kerusakan bangunan pascagempa, dibandingkan dengan penggunaan metode tunggal (LSTM atau *PCA-Clustering*).

2. H0 (Hipotesis Nol)

Model *hybrid* yang mengintegrasikan LSTM dan *PCA-Clustering* tidak memberikan perbedaan akurasi yang signifikan secara statistik dalam memprediksi dan mengkluster tingkat kerusakan bangunan pascagempa dibandingkan dengan metode tunggal.

### 1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

### 1. Secara Teoretis (Ilmiah)

Penelitian ini berkontribusi pada pengembangan model hybrid yang mengintegrasikan metode *prediction* (LSTM) dengan teknik *clustering* (PCA-Clustering). Kontribusi ini menawarkan pendekatan baru, yang disebut *prediction-based clustering*, untuk menganalisis dan memprediksi tingkat kerusakan bangunan pascagempa.

### 2. Secara Praktis (Industri/Lembaga)

Hasil penelitian ini dapat menjadi acuan bagi instansi terkait (seperti PUPR) dan pengembang swasta di wilayah Jawa Timur dan Jawa Tengah. Model ini dapat digunakan sebagai alat bantu dalam pengambilan keputusan untuk perencanaan pembangunan di wilayah berpotensi gempa tinggi, terutama dalam menentukan standar mutu dan konstruksi bangunan yang lebih tahan bencana.

## 1.6 Batasan Masalah

Untuk menjaga agar penelitian ini tetap fokus dan terarah, penelitian ini dibatasi pada hal-hal berikut:

1. Data yang digunakan dalam penelitian ini adalah data pusat gempa tahun 2024 yang berlokasi studi di Pulau Bawean.
2. Metodologi yang digunakan berfokus pada metode *prediction* menggunakan *Long Short-Term Memory* (LSTM) dan metode klasterisasi menggunakan *PCA-Clustering* untuk memproses dan mengidentifikasi tingkat kerusakan bangunan pascagempa.

## 1.7 Sistematika Penulisan

Penulisan Tesis ini disusun dalam lima bab yang saling berkaitan untuk

memberikan gambaran yang utuh mengenai penelitian yang dilakukan.

Sistematika penulisannya adalah sebagai berikut:

#### 1. BAB I: PENDAHULUAN

Bab ini berisi pendahuluan yang mencakup latar belakang masalah, perumusan masalah, tujuan penelitian, hipotesis, manfaat penelitian, batasan masalah, dan diakhiri dengan sistematika penulisan.

#### 2. BAB II: TINJAUAN PUSTAKA

Bab ini menguraikan landasan teoretis yang relevan dengan penelitian, seperti teori mengenai gempa bumi, penilaian kerusakan bangunan, metode prediction, *Long Short-Term Memory (LSTM)*, *Principal Component Analysis (PCA)*, dan *Clustering*. Bab ini juga mengulas penelitian-penelitian terdahulu yang relevan (*state-of-the-art*).

#### 3. BAB III: METODOLOGI PENELITIAN

Bab ini menjelaskan secara rinci langkah-langkah penelitian, mulai dari kerangka penelitian, metode pengumpulan data, arsitektur model *prediction-based clustering* yang diusulkan, serta metode pengujian dan evaluasi model.

#### 4. BAB IV: HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi dan pengujian model. Data hasil eksperimen akan dianalisis dan dibahas secara mendalam untuk menjawab hipotesis dan tujuan penelitian.

#### 5. BAB V: KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari keseluruhan hasil penelitian dan pembahasan, serta memberikan saran untuk pengembangan penelitian selanjutnya berdasarkan temuan yang diperoleh.

## BAB II TINJAUAN PUSTAKA

### 2.1 *Prediction-based Clustering*

Penelitian terkait *prediction* yang mana telah banyak dilakukan oleh beberapa peneliti terdahulu dengan berbagai metode, beberapa penelitian tersebut diantaranya:

Menurut penelitian yang dilakukan Tao Bai, et al. (2022) berfokus pada permasalahan deteksi dan lokalisasi gempa bumi yang merupakan tugas mendasar dalam seismologi, tetapi sangat menantang karena sinyal seismik seringkali bercampur dengan *noise*, terutama untuk gempa mikro (Bai & Tahmasebi, 2022). Penelitian ini mengusulkan model *hybrid* bernama *attention-based LSTM-FCN* yang menggabungkan keunggulan *Convolutional Neural Network* (FCN) dan *Long Short-Term Memory* (LSTM) untuk meningkatkan akurasi deteksi dan resolusi lokalisasi gempa. Meskipun model *deep learning* seperti *ConvNetQuake* telah menunjukkan keberhasilan dalam deteksi gempa, akurasi klasifikasi dan resolusi geografisnya masih terbatas. Selain itu, model tersebut hanya mampu mengklasifikasikan gempa ke dalam enam wilayah geografis, yang kurang memadai untuk analisis yang lebih rinci. Model mencapai akurasi klasifikasi sebesar 89.1%, meningkat 14.5% dibandingkan *ConvNetQuake*. Presisi: 96.2%, *Recall*: 99.9%, dan *F1-Score*: 98.0%, Model mampu mengklasifikasikan gempa ke dalam 15 wilayah geografis dengan akurasi tetap tinggi (78.5%). Penelitian ini berhasil menunjukkan bahwa kombinasi FCN, LSTM, dan mekanisme *attention* dapat meningkatkan kinerja deteksi dan lokalisasi gempa secara signifikan. Model ini juga lebih fleksibel dalam hal resolusi geografis dibandingkan pendekatan

sebelumnya.

Menurut penelitian yang dilakukan Pei-Ru Jian, et al. (2022) berfokus rangkaian gempa 2018–2022 di Hualien menunjukkan distribusi temporal dan spasial yang kompleks, dengan mekanisme fokal yang tidak selalu sesuai dengan struktur sesar yang diketahui di permukaan (Jian & Wang, 2022). Hal ini menyulitkan identifikasi geometri struktur seismogenik yang bertanggung jawab atas gempa-gempa tersebut. Penelitian ini berkontribusi dalam mengkarakterisasi struktur seismogenik di Hualien menggunakan algoritma *machine learning* tanpa pengawasan (*unsupervised*) seperti DBSCAN dan PCA, serta metode *back-projection MUSIC*. Hasilnya memberikan pemahaman baru tentang interaksi struktur gempa dan transfer tegangan statis selama rangkaian gempa 2018–2022. Metode yang diusulkan pada penelitian ini menggunakan dua pendekatan utama yaitu *MUSIC Back-Projection*: Untuk memetakan proses ruptur gempa kuat dalam rangkaian gempa. Algoritma DBSCAN dan PCA: Untuk mengelompokkan gempa dan mengidentifikasi orientasi planar dari struktur seismogenik. Penelitian ini berhasil mengungkap kompleksitas struktur seismogenik di Hualien, termasuk rotasi bidang sesar sepanjang pantai dan hubungannya dengan subduksi Lempeng Laut Filipina. Hasil ini mendukung pemahaman tentang dinamika gempa di zona tumbukan-subduksi.

Menurut penelitian yang dilakukan oleh Aarón Cofré, et al. (2022) berfokus pada ketergantungan metode tradisional pada jaringan stasiun seismik yang padat, dan kesulitan dalam mengekstrak fitur gempa secara *real-time* dari data *waveform* mentah menggunakan pendekatan berbasis fisika (Cofre et al., 2022). Penelitian ini mengusulkan model *Long Short-Term Memory* (LSTM) berbasis *single-station*

yang bekerja secara *end-to-end* untuk mengestimasi magnitudo gempa langsung dari data seismik mentah. Kontribusi utamanya meliputi: (1) eliminasi kebutuhan *feature engineering* manual, (2) kemampuan prediksi hanya dengan satu stasiun seismik, dan (3) reduksi latensi komputasi hingga 70% dibanding metode berbasis jaringan. Model LSTM dirancang untuk memproses *waveform* seismik tiga-komponen (*vertical, north-south, east-west*) secara langsung. Arsitektur ini menggabungkan lapisan *convolutional* untuk ekstraksi fitur lokal dan LSTM untuk memodelkan ketergantungan temporal. Pelatihan menggunakan dataset gempa *Chile* dengan augmentasi data untuk generalisasi. Pengujian pada dataset independen menunjukkan akurasi estimasi magnitudo dengan *mean absolute error* (MAE) 0.21 dan korelasi 0.93 terhadap nilai referensi. Model ini unggul dibanding pendekatan *random forest* (MAE 0.35) dan CNN tradisional (MAE 0.28), terutama untuk gempa bermagnitudo 4.0–6.5.

Penelitian Peng Zhang, et al. (2022) berjudul berfokus pada tantangan memprediksi gempa di tambang yang dapat membahayakan keselamatan pekerja dan merusak infrastruktur (Zhang et al., 2022). Untuk itu, penelitian ini menggunakan metode *Spatial Temporal-DBSCAN (Density-Based Spatial Clustering of Applications with Noise)*, yang menggabungkan analisis *spasial* dan *temporal* dari data seismik. Hasil penelitian menunjukkan bahwa metode ini berhasil meningkatkan akurasi prediksi gempa dengan mengelompokkan data seismik ke dalam klaster-klaster yang menunjukkan pola temporal dan spasial tertentu. Pendekatan ini memungkinkan prediksi waktu dan lokasi gempa yang lebih tepat, memberikan manfaat besar untuk mitigasi risiko dan keselamatan di tambang yang rawan gempa.

Menurut penelitian yang dilakukan Agung Teguh Wibowo Almais, et al. (2023) mengidentifikasi dua masalah utama yaitu kesulitan dalam mengolah data multidimensi yang mencakup berbagai parameter kerusakan, dan ketiadaan sistem otomatis yang dapat mengelompokkan wilayah berdasarkan tingkat kerusakan dengan akurasi tinggi (Almais et al., 2023). Keterbatasan ini sering menyebabkan ketidakmerataan dalam alokasi sumber daya pascabencana. mengusulkan *framework* berbasis *Principal Component Analysis* (PCA) dan *clustering* untuk mengkluster tingkat kerusakan wilayah pascabencana. Tahapan penelitian yang pertama adalah pengumpulan data multidimensi (fisik bangunan, infrastruktur, korban jiwa), transformasi PCA untuk ekstraksi fitur dominan, *clustering* untuk mengelompokkan wilayah menjadi beberapa level kerusakan (ringan, sedang, berat), dan validasi menggunakan data historis bencana. Pengujian dilakukan pada data bencana gempa bumi di Indonesia dengan 15 parameter awal. PCA berhasil mereduksi dimensi menjadi 3 komponen utama (mencakup 92% variansi data). Hasil *clustering* menunjukkan akurasi 89% dibandingkan dengan label ahli, dengan waktu pemrosesan 40% lebih cepat dibanding metode konvensional. Temuan utama menunjukkan bahwa pendekatan ini mampu mengidentifikasi pola kerusakan yang konsisten dengan penilaian ahli, mengurangi kompleksitas data tanpa kehilangan informasi kritis, dan menyediakan visualisasi yang intuitif untuk pengambilan keputusan. Sistem ini juga terbukti adaptif untuk berbagai skenario bencana berbeda.

Menurut penelitian yang dilakukan Li Chenchao, et al. (2024) penelitian pada analisis data sekuensial dan pengurangan dimensi data, yang memungkinkan deteksi kerusakan lebih cepat dan lebih efisien (Chenchao et al., 2024). Metode

yang digunakan pada penelitian ini yaitu LSTM-FCN (*Long Short-Term Memory Fully Convolutional Network*). Hasil dari penelitian menggunakan metode *Long-Short Term Memory (LSTM) auto-encoder* yang digabungkan dengan *Impulse Response Function (IRF)* berhasil memberikan metode yang efektif untuk kuantifikasi kerusakan struktural. Model *LSTM auto-encoder* menunjukkan kemampuan luar biasa dalam mendeteksi dan mengidentifikasi perubahan dalam respons dinamis struktur seiring dengan waktu, yang merupakan indikasi awal adanya kerusakan. Dengan memanfaatkan data vibrasi dan respons struktural yang tercatat, model ini dapat mengidentifikasi kerusakan pada bangunan bahkan dengan perubahan kecil yang tidak selalu terdeteksi dengan metode konvensional.

Menurut penelitian yang dilakukan Yuchuan Yang, et al. (2024) penelitian ini mengatasi dua keterbatasan utama yaitu ketidakmampuan model *existing* dalam memproses data deret waktu multivariat dari berbagai sumber (InSAR, curah hujan, data geologi) secara terintegrasi, dan kesulitan dalam mengekstrak pola temporal jangka panjang dan pendek secara simultan dari data deformasi tanah (Yang et al., 2024). Penelitian ini mengusulkan *framework hybrid* yang menggabungkan *Temporal Fusion Transformer (TFT)* dengan *clustering* deret waktu multivariat. Kontribusi utamanya yaitu arsitektur *deep learning* yang secara khusus dirancang untuk prediksi deformasi tanah berbasis InSAR, sistem *clustering* untuk mengidentifikasi pola deformasi *spatio-temporal*, serta *prototype* yang terbukti efektif di wilayah kompleks seperti Tibet Timur. *Framework* ini mencapai RMSE 2.3 mm dalam prediksi deformasi (35% lebih akurat dibanding LSTM dan 28% lebih baik dari Transformer standar). Analisis *clustering* berhasil mengidentifikasi 4 pola deformasi dominan dengan *silhouette score* 0.71. Model

juga menunjukkan robust terhadap missing data hingga 15%. Temuan kunci menunjukkan bahwa integrasi TFT dan *clustering multivariat* secara signifikan meningkatkan akurasi prediksi jangka menengah (3-6 bulan), *attention mechanism* dalam TFT berhasil mengkuantifikasi kontribusi relatif berbagai faktor pemicu (curah hujan menjadi dominan dengan *weight* 0.62), dan pendekatan ini memungkinkan identifikasi *early warning sign* melalui analisis pola cluster.

Menurut penelitian yang dilakukan oleh Stefania Gentili, et al. (2024) berfokus pada dua masalah spesifik yaitu karakterisasi proses difusi fluida yang memicu rangkaian gempa Molise 2018 di Italia Selatan, dan identifikasi hubungan temporal-spasial antara migrasi fluida dengan distribusi hiposenter gempa (Gentili et al., 2024). Wilayah ini menjadi kasus menarik karena kompleksitas struktur geologinya yang melibatkan interaksi sesar normal dan *strike-slip*. Peneliti menggunakan metodologi *hybrid* yang menggabungkan analisis relokasi hiposenter dengan teknik *double-difference*, pemodelan difusi fluida berbasis hukum Darcy, inversi mekanisme fokal untuk menentukan parameter sumber, analisis statistik pola *spasial-temporal* seismisitas, data diperoleh dari jaringan *seismograf* INGV dan stasiun temporer yang dipasang khusus pascagempa utama. Hasil utama menunjukkan relokasi hiposenter mengungkap migrasi kluster seismik dengan kecepatan  $\sim 0.5$  km/hari, model difusi fluida menunjukkan kecocokan baik ( $R^2=0.89$ ) dengan pola observasi, parameter difusi ( $D \sim 0.3$  m<sup>2</sup>/s) konsisten dengan sistem sesar permeable, mekanisme fokal dominan menunjukkan kombinasi sesar normal dan *strike-slip*. Temuan kunci penelitian ini mengungkap bahwa sekues gempa Molise 2018 dipicu oleh difusi fluida tekanan tinggi sepanjang zona sesar yang sudah ada, migrasi hiposenter menunjukkan pola

anisotropik yang terkontrol struktur geologi lokal, terdapat korelasi kuat antara zona permeabilitas tinggi dengan densitas gempa, model ini berhasil menjelaskan ~75% aktivitas *aftershock* dalam 30 hari pertama.

Menurut penelitian yang dilakukan oleh Huiwei Wang, et al. (2024) berfokus pada risiko gempa bumi di wilayah rawan seismik memerlukan sistem mitigasi yang efektif untuk meningkatkan ketahanan bangunan (Wang et al., 2024). Sistem *Earthquake Early Warning* (EEW) saat ini masih memiliki keterbatasan dalam memprediksi dampak gempa secara akurat, terutama pada skala regional. Ketidakpastian dalam estimasi parameter seismik dan respons bangunan sering menyebabkan peringatan yang tidak tepat, berpotensi meningkatkan risiko kerusakan infrastruktur dan korban jiwa.

Menurut penelitian yang dilakukan oleh Agung Teguh Wibowo Almais, et al. (2024) kerusakan bangunan pascabencana masih menghadapi tantangan signifikan dalam hal akurasi dan kecepatan *asesmen* (Almais et al., 2024). Metode konvensional yang mengandalkan survei lapangan manual tidak hanya memakan waktu tetapi juga rentan terhadap variasi subjektivitas penilai. Keterlambatan dalam penilaian ini dapat berdampak serius pada proses rehabilitasi dan rekonstruksi. Penelitian ini mengatasi dua masalah spesifik yaitu kesulitan ekstraksi fitur tekstural yang konsisten dari citra bangunan rusak, dan kebutuhan akan sistem klasifikasi kerusakan yang *robust* terhadap variasi kondisi pencahayaan dan sudut pengambilan gambar. Tantangan ini krusial terutama di daerah bencana dengan infrastruktur monitoring terbatas. Pendekatan penelitian mencakup tiga tahap kunci yaitu ekstraksi fitur tekstural menggunakan GLCM dengan 5 parameter (*contrast, correlation, energy, homogeneity, entropy*), reduksi

dimensi dengan PCA untuk mendapatkan komponen utama paling informatif, klasifikasi menggunakan *Support Vector Machine* (SVM) dengan kernel RBF, dataset terdiri dari 1.200 citra bangunan dari 4 bencana berbeda dengan 4 level kerusakan (ringan, sedang, berat, runtuh). Hasil pengujian menunjukkan akurasi klasifikasi mencapai 92.4% pada data uji, kombinasi fitur GLCM *correlation* + PCA *component* 1&2 memberikan performa terbaik, waktu pemrosesan per citra hanya 0.8 detik (50x lebih cepat dari survei manual), model menunjukkan ketahanan baik terhadap *noise* (akurasi tetap >85% pada SNR 15dB). Temuan utama penelitian mengungkap bahwa integrasi GLCM-PCA berhasil menangkap karakteristik tekstural kunci yang berkorelasi dengan tingkat kerusakan struktural, pendekatan ini efektif untuk berbagai material bangunan (beton, kayu, bata), sistem mampu membedakan kerusakan struktural dari kerusakan non-struktural dengan akurasi 89%, visualisasi PCA memungkinkan identifikasi pola kerusakan yang tidak terlihat secara manual.

Menurut penelitian yang dilakukan oleh Yuxuan Tao, et al. (2025) penilaian kerusakan regional pascagempa, terutama setelah serangkaian gempa utama (*mainshock*) dan gempa susulan (*aftershock*) yang merupakan tantangan kritis dalam rekayasa gempa (Tao et al., 2025). Penelitian ini menyoroti keterbatasan kerangka kerja berbasis energi (*energy framework*) yang ada dalam menilai kerapuhan (*fragility*) struktur secara cepat di wilayah luas. Selain itu, integritas antara analisis dinamik nonlinier dengan prediksi kerusakan berbasis data masih belum optimal, terutama untuk skenario gempa berurutan yang kompleks. Kontribusi utama penelitian ini adalah penggabungan *deep learning* (DL) ke dalam kerangka kerja berbasis energi untuk mempercepat penilaian kerusakan

regional. Metode ini memungkinkan prediksi kerapuhan struktur dengan akurasi tinggi dan waktu komputasi yang jauh lebih singkat, sehingga cocok untuk aplikasi *real-time* dalam manajemen bencana. Penulis mengusulkan model *hybrid* yang menggabungkan simulasi dinamik berbasis energi dengan jaringan saraf tiruan (*neural networks*). Data energi disipasi struktur dari simulasi *mainshock-aftershock* dikombinasikan dengan parameter gempa seperti *Peak Ground Acceleration* (PGA) sebagai input pelatihan DL. PGA digunakan untuk mengkarakterisasi intensitas guncangan pada setiap seri gempa, baik pada tahap *mainshock* maupun *aftershock*. Model DL kemudian memprediksi tingkat kerusakan dan kerapuhan struktur dengan mempertimbangkan akumulasi dampak dari variasi PGA pada serangkaian gempa tersebut. Eksperimen dilakukan dengan data gempa sintetik dan historis yang mencakup variasi PGA (0.1g–0.8g) untuk merepresentasikan skenario gempa rendah hingga tinggi. Model DL (CNN-LSTM) menunjukkan sensitivitas terhadap perubahan PGA, dengan akurasi prediksi >90% pada rentang PGA 0.2g–0.6g, yang merupakan wilayah kritis untuk kerusakan struktural. Hasil juga mengungkap bahwa akumulasi kerusakan akibat *aftershock* dengan PGA  $\geq 30\%$  dari *mainshock* meningkatkan probabilitas keruntuhan hingga 40%.

Menurut penelitian yang dilakukan oleh Yhon Fuentes, et al. (2025) sistem peringatan dini gempa bumi yang memerlukan prediksi parameter seismik yang akurat dan cepat, seperti *Peak Ground Acceleration* (PGA), untuk meminimalkan risiko bencana (Fuentes et al., 2025). Penelitian ini berfokus pada keterbatasan model prediksi PGA berbasis regresi statistik atau jaringan saraf tunggal dalam mengintegrasikan data multivariabel (misalnya: waktu, magnitudo, kedalaman)

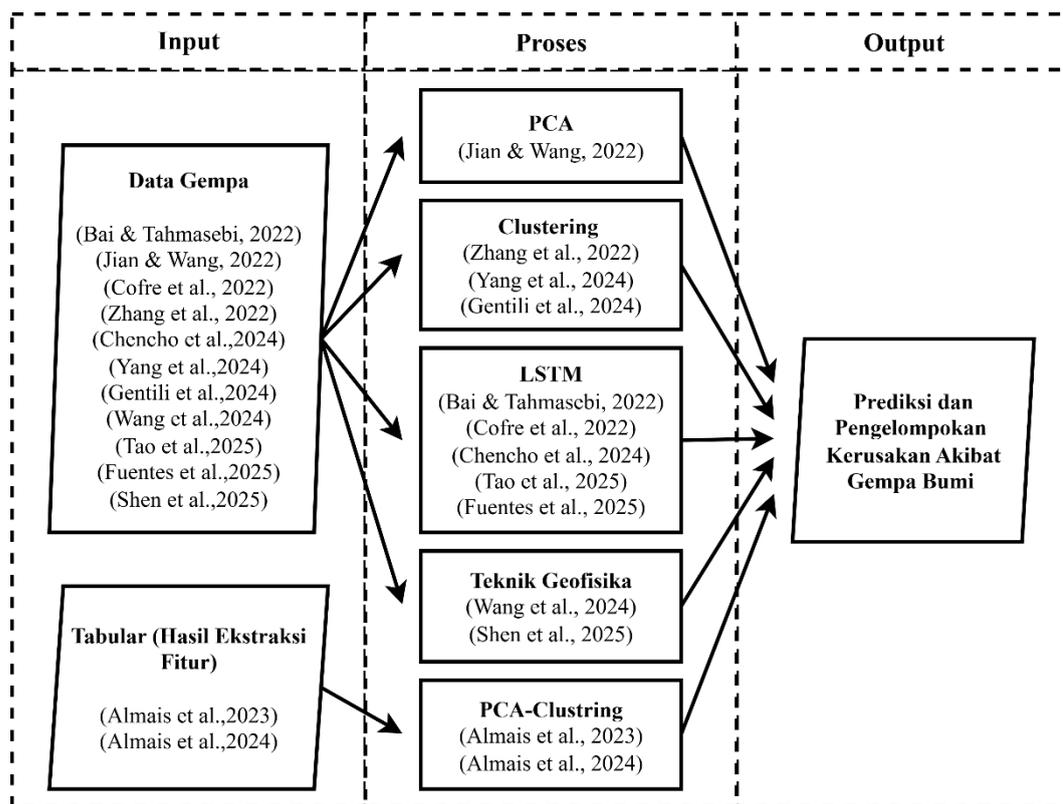
dan pola temporal jangka panjang. Akurasi prediksi yang tidak stabil pada kondisi gempa kompleks menjadi tantangan utama. Penelitian mengembangkan arsitektur *hybrid* yang menggabungkan *Long Short-Term Memory* (LSTM) dan *Temporal Transformers* untuk meningkatkan akurasi prediksi PGA. Model dirancang untuk menangkap pola temporal jangka pendek (via LSTM) dan ketergantungan jangka panjang (via *Transformers*), serta mengintegrasikan data multivariabel secara efektif. Manfaatnya mencakup peningkatan respons sistem dan mitigasi risiko gempa yang lebih tepat. Eksperimen dilakukan pada dataset gempa bumi historis dengan variasi magnitudo dan kedalaman. Model *hybrid* menunjukkan kinerja superior dibandingkan model tunggal (LSTM atau *Transformers*), dengan reduksi kesalahan RMSE hingga 18% dan peningkatan korelasi prediksi ( $R^2 > 0.92$ ). Hasil ini konsisten pada berbagai skenario gempa, termasuk kejadian dengan magnitudo tinggi ( $>6.0$  Mw). Penelitian membuktikan bahwa kombinasi LSTM dan *Transformers* mampu meningkatkan akurasi prediksi PGA dengan memanfaatkan informasi temporal lokal dan global. Model ini juga menunjukkan ketahanan terhadap *noise* data, menjadikannya kandidat potensial untuk integrasi dalam sistem secara *real-time*.

Menurut penelitian yang dilakukan oleh Xuanye Shen, et al. (2025) sistem peringatan dini gempa (*Earthquake Early Warning/EEW*) memerlukan estimasi magnitudo yang cepat dan akurat untuk meminimalkan dampak bencana (Shen et al., 2025). Namun, sifat gempa yang kompleks dan bervariasi membuat prediksi magnitudo dalam waktu singkat menjadi tantangan besar. Metode tradisional berbasis pemodelan matematis dan *shallow machine learning* memiliki keterbatasan dalam akurasi dan generalisasi, terutama ketika hanya informasi awal

gelombang P yang tersedia. Penelitian ini menggunakan dataset gempa dari jaringan K-NET dan Kik-net Jepang yang mencakup periode Januari 2008 hingga Juni 2021, dengan magnitudo berkisar antara 3.0 hingga 7.7. Data diproses melalui beberapa tahapan, yaitu: picking gelombang P menggunakan metode rata-rata jangka panjang dan pendek, koreksi *baseline*, serta filtering frekuensi rendah dengan *Butterworth high-pass*. Empat model *deep learning* dibandingkan dalam penelitian ini. Pertama, *MagNet* yang menggabungkan CNN 1D dan Bi-LSTM untuk ekstraksi fitur temporal. Kedua, DCRNN yang mengembangkan *MagNet* dengan menambahkan lapisan CNN dan mengurangi ukuran kernel secara bertahap untuk memperluas bidang persepsi. Hasil eksperimen menunjukkan bahwa DCRNNamp memberikan performa terbaik dengan MAE 0.287 dan  $R^2$  0.737 dalam 3 detik pertama setelah kedatangan gelombang P. Model ini juga lebih stabil dalam menangani gempa besar ( $>6.5$  M) dibandingkan model lainnya. Sementara itu, *Exams* menunjukkan kinerja yang kurang optimal pada tahap awal prediksi, dengan  $R^2$  hanya 0.354 dalam 1 detik pertama. Namun, akurasinya meningkat seiring bertambahnya waktu.

## 2.2 Kerangka Teori

Pada tahapan ini membahas tentang metode yang telah dibahas pada subBab sebelumnya yang membahas dengan membandingkan berdasarkan studi literatur pada penelitian ini disertai dengan alur desain penelitian yang terdiri dari input, process dan output seperti pada gambar 2.2.



Gambar 2.1 Kerangka teori

Pada penelitian ini penulis menggunakan data gempa bumi dengan variabel Magnitudo dan Kedalaman sebagai parameter input. Kemudian diolah menggunakan metode LSTM PCA-Clustering untuk melakukan *clustering* gempa bumi. Pada metode PCA-Clustering dilakukan pengukuran nilai Tingkat kerusakan bangunan akibat gempa bumi untuk mendapatkan *cluster* terbaik. sedangkan pada metode LSTM dilakukan pengukuran MAE (*Mean Absolute Error*) untuk mendapatkan evaluasi terbaik. Sehingga hasil yang diperoleh adalah *cluster* tingkat kerusakan bangunan akibat gempa bumi di wilayah Jawa Timur, Jawa Tengah.

Gempa bumi adalah peristiwa bergetarnya bumi akibat pelepasan energi di dalam bumi secara tiba-tiba yang ditandai dengan patahnya lapisan batuan pada

kerak bumi. Akumulasi energi dihasilkan dari pergerakan lempeng-lempeng tektonik yang dinamis. Energi yang dihasilkan dipancarkan ke segala arah berupa gelombang gempa bumi sehingga efeknya dapat dirasakan sampai ke permukaan bumi. Menurut teori lempeng tektonik, permukaan bumi terpecah menjadi beberapa lempeng tektonik besar. Lempeng tektonik adalah segmen keras kerak bumi yang mengapung di atas *astenosfer* yang cair dan panas. Oleh karena itu, lempeng tektonik saling berinteraksi satu sama lain. Daerah perbatasan lempeng-lempeng tektonik merupakan tempat-tempat yang memiliki kondisi tektonik aktif yang menyebabkan gempa bumi, gunung berapi, dan pembentukan dataran tinggi (Mia et al., 2024). Secara umum, struktur geologi aktif di Jawa didominasi oleh sesar geser dan sesar naik dengan sesar turun sebagai struktur minor. Daerah Jawa Timur didominasi oleh sesar patahan turun (Pusat Studi Gempa Nasional, 2022)

Gempa bumi merupakan fenomena alam yang kompleks dan sulit diprediksi secara akurat. Oleh karena itu, pendekatan berbasis analisis deret waktu (*time series*) mulai digunakan dalam penelitian untuk mendeteksi pola-pola tersembunyi dari data gempa bumi yang tersedia. Dalam konteks ini, eksplorasi data awal (EDA) dan penerapan metode *machine learning* memberikan kontribusi penting dalam memetakan tren kejadian gempa secara statistik maupun prediktif. Penelitian yang dilakukan oleh (Azis et al., 2020) menggabungkan teknik visualisasi dan model prediktif untuk menganalisis frekuensi dan magnitudo gempa, dengan memanfaatkan data historis dari katalog gempa global. Pendekatan ini bertujuan untuk menghasilkan wawasan yang lebih baik terhadap perilaku gempa bumi serta mendukung sistem peringatan dini yang lebih responsif dan berbasis data.

Tabel 2.1 Kerangka Teori LSTM PCA-clustering

Penulis	Judul Artikel	Metode	Hasil
Tao Bai, et al. (2022)	Attention-based LSTM-FCN for earthquake detection and location	Attention-based LSTM-FCN	Model mampu meningkatkan akurasi deteksi gempa kecil dan mengurangi kesalahan lokalisasi
Aziz, et al. (2020)	Time series analysis on earthquakes using EDA and machine learning	LSTM auto-encoder	LSTM auto-encoder efektif mendeteksi kerusakan struktur secara akurat meski dengan noise dan sensor terbatas.
Pei-Ru Jian, et al. (2022)	Applying unsupervised machine-learning algorithms and MUSIC back-projection to characterize 2018–2022 Hualien earthquake sequence	Algoritma unsupervised dan MUSIC back-projection	Kluster spasial-temporal gempa berkorelasi dengan struktur geologi aktif
Aarón Cofré, et al. (2022)	End-to-End LSTM-Based Earthquake Magnitude Estimation With a Single Station	Jaringan LSTM end-to-end	Estimasi magnitudo akurat ( $\pm 0.3$ ) dengan data single stasiun
Peng Zhang, et al. (2022)	Prediction Method for Mine Earthquake in Time Sequence Based on Clustering Analysis	Analisis Pengelompokan	Klaster temporal membantu identifikasi periode risiko tinggi
Agung Teguh Wibowo Almais, et al. (2023)	Principal Component Analysis-Based Data Clustering for Labeling of Level Damage Sector in Post-Natural Disasters	Principal Component Analysis (PCA) dan Klasterisasi Data	Klaster kerusakan sesuai tingkat (ringan, sedang, berat)
Chencho, et al. (2024)	Structural damage quantification using long short-term memory (LSTM) auto-encoder and impulse response functions	Auto-Encoder LSTM dan Analisis Impulse Response	Akurasi deteksi kerusakan mencapai 92% pada data uji
Yuchuan Yang, et al. (2024)	Advanced Prediction of Landslide Deformation Through Temporal Fusion Transformer and ultivariate Time-Series	Temporal Fusion Transformer (TFT) dan Klasterisasi Deret Waktu	Prediksi deformasi 3 hari sebelumnya dengan RMSE 1.2 cm

Penulis	Judul Artikel	Metode	Hasil
	Clustering of InSAR: Insights From the Badui Region, Eastern Tibet		
Stefania Gentili, et al. (2024)	Seismic clusters and fluids diffusion: a lesson from the 2018 Molise (Southern Italy) earthquake sequence	Analisis statistik dan Klasterisasi Spasial-Temporal	Klaster gempa terkait migrasi fluida pada zona sesar
Agung Teguh Wibowo Almais, et al. (2024)	Characterization of Structural Building Damage in Post-Disaster using GLCM-PCA Analysis Integration	Integrasi GLCM (Gray-Level Co-Occurrence Matrix) dan PCA	Integrasi GLCM-PCA efektif klasifikasi kerusakan (akurasi 89%)
Xuanye Shen, et al. (2025)	Comparative Analysis of Deep Learning Methods for Real-Time Estimation of Earthquake Magnitude	CNN, RNN, dan Transformer	LSTM unggul dalam latensi rendah ( $\pm 0.1$ detik)
Yhon Fuentes, et al. (2025)	Peak Ground Acceleration Prediction for Earthquake Early Warning with Multivariable Long Short-Term Memory Networks and Temporal Transformers	Multivariabel-LSTM dan Temporal Transformer	Model mengurangi kesalahan PGA hingga 15% dibanding metode konvensional
Yuxuan Tao, et al. (2025)	Integrating Deep Learning Into an Energy Framework for Rapid Regional Damage Assessment and Fragility Analysis Under Mainshock-Aftershock Sequences	CNN-LSTM	Model memperkirakan kerusakan 30% lebih cepat dengan akurasi 85%
Penelitian ini	Prediction-Based Clustering Tingkat Kerusakan Bangunan Pascagempa Bumi	Long-Short Time Memory (LSTM) dan PCA-Clustering	Tingkat Kerusakan Bangunan Pascagempa Bumi

Berdasarkan tabel 2.1 diperoleh beberapa metode *prediction* dan *clustering* yang dapat digunakan untuk melakukan prediksi dan pengelompokan data gempa

bumi yang bersumber dari berbagai artikel ilmiah pada beberapa penelitian terdahulu. Prediksi dan pengelompokan tingkat kerusakan bangunan pascagempa bumi merupakan masalah kompleks yang memerlukan analisis data temporal, faktor struktural, dan kondisi lingkungan. Studi sebelumnya Almais et al, (2023; 2024) dan Chencho et al, (2024) telah mengadopsi metode statis seperti *Principal Component Analysis* (PCA) dan *Long Short-Term Memory* (LSTM) untuk identifikasi kerusakan berbasis citra satelit atau data respons dinamis. Namun, pendekatan tersebut cenderung terpisah antara analisis temporal (*prediction*) dan klusterisasi, sehingga kurang mampu memodelkan evolusi kerusakan secara holistik. Di sisi lain, metode klusterisasi berbasis reduksi dimensi yang menawarkan efisiensi komputasi belum diintegrasikan dengan prediksi dinamik untuk skenario pascabencana.

Meskipun studi Almais et al. (2024) telah mengembangkan integrasi GLCM-PCA untuk mengekstraksi fitur tekstur kerusakan, analisis tersebut mengabaikan faktor temporal seperti getaran berulang atau riwayat beban struktural. Sementara itu, model prediktif seperti Yuxuan Tao et al. (2025) telah mengembangkan kerangka energi berbasis *deep learning* untuk menilai kerusakan infrastruktur akibat gempa utama dan *aftershock*, namun penelitian tersebut terbatas pada evaluasi kerusakan *post-hoc* tanpa memprediksi bagaimana kluster kerusakan berevolusi setelah serangkaian gempa. Penelitian tersebut mengisi celah dengan memvalidasi model *LSTM-based PCA-Clustering* pada simulasi *mainshock-aftershock*, di mana kluster kerusakan diproyeksikan tidak hanya berdasarkan gempa utama, tetapi juga respons struktural terhadap aktivitas gempa susulan. Pendekatan ini memberikan wawasan baru tentang kapasitas model dalam

memprediksi evolusi kerusakan pada kondisi seismik yang dinamis. Penelitian ini menjawab tantangan tersebut dengan mengintegrasikan LSTM (untuk memodelkan ketergantungan temporal) dan PCA (untuk klasterisasi berdasarkan kerusakan bangunan pascagempa bumi), sehingga memungkinkan identifikasi kluster kerusakan yang dipengaruhi oleh interaksi multidimensi.

### 2.3 Integrasi Al-Qur'an, Tafsir dan Hadis

Ayat ayat ini mengingatkan bahwa goncangan dan kerusakan yang terjadi di bumi, termasuk gempa dan bencana alam lainnya, berada dalam kekuasaan Allah sebagai bagian dari ketetapan-Nya. Pada saat yang sama, sebagian kerusakan di darat dan laut juga muncul akibat ulah manusia yang merusak keseimbangan alam. Melalui peristiwa tersebut, Allah memberikan pelajaran agar manusia kembali kepada kebenaran dan memperbaiki perilaku, yang terdapat pada surat Al-Baqarah ayat 4 dan surat Ar-Rum ayat 41:

إِذَا رُجَّتِ الْأَرْضُ رَجًا ۝٤

Artinya: “*Apabila bumi digoncangkan sedahsyat-dahsyatnya*”. (Q.S. Al-Baqarah: 4)

ظَهَرَ الْفَسَادُ فِي الْبَرِّ وَالْبَحْرِ بِمَا كَسَبَتْ أَيْدِي النَّاسِ لِيُذِيقَهُمْ بَعْضَ الَّذِي عَمِلُوا لَعَلَّهُمْ يَرْجِعُونَ ۝٤١

Artinya: “*Telah tampak kerusakan di darat dan di laut disebabkan perbuatan tangan manusia. (Melalui hal itu) Allah membuat mereka merasakan sebagian dari (akibat) perbuatan mereka agar mereka kembali (ke jalan yang benar)*” (Q.S. Ar-Rum: 41).

Menurut kitab tafsir *Almadinah Al-Munawwarah* dari Syaikh Prof. Dr. Imad Zuhair Hafidz menjelaskan Jika bumi mulai bergetar dan berguncang dengan keras, dan gunung-gunung hancur berkeping-keping sehingga menimbulkan debu yang bertebaran. Dan dari kitab tafsir *Wajiz* menjelaskan bahwa kerusakan di

darat maupun di laut, baik di kota maupun desa, muncul akibat perbuatan manusia yang mengikuti hawa nafsu dan menyimpang dari fitrahnya. Allah membiarkan mereka merasakan sebagian akibat dari tindakan tersebut agar mereka sadar dan kembali kepada perilaku yang sesuai dengan fitrah dan tuntunan yang benar.

Selanjutnya ayat Al-Qur'an dan hadis yang menjelaskan pengelompokan manusia ke dalam suku dan bangsa (*lita'ārafū*) menegaskan bahwa perbedaan tersebut dimaksudkan agar manusia saling mengenal dan membangun hubungan harmonis. Namun, ukuran kemuliaan yang hakiki di sisi Allah bukan ditentukan oleh ras atau keturunan, melainkan oleh tingkat ketakwaan. Prinsip ini ditegaskan dalam Surah Al-Hujurat ayat 13 serta hadis riwayat Ahmad dan At-Tirmidzi.

يَا أَيُّهَا النَّاسُ إِنَّا خَلَقْنَاكُمْ مِنْ ذَكَرٍ وَأُنْثَىٰ وَجَعَلْنَاكُمْ شُعُوبًا وَقَبَائِلَ لِتَعَارَفُوا ۗ إِنَّ أَكْرَمَكُمْ عِنْدَ اللَّهِ أَتَقْوَىٰ ۗ إِنَّ اللَّهَ عَلِيمٌ خَبِيرٌ

Artinya: *"Wahai manusia! Sungguh, Kami telah menciptakan kamu dari seorang laki-laki dan seorang perempuan, kemudian Kami jadikan kamu berbangsa-bangsa dan bersuku-suku agar kamu saling mengenal. Sesungguhnya yang paling mulia di antara kamu di sisi Allah ialah orang yang paling bertakwa. Sungguh, Allah Maha Mengetahui, Mahateliti".* (Q.S. Al-Hujurat Ayat 13)

يَا أَيُّهَا النَّاسُ، أَلَا إِنَّ رَبَّكُمْ وَاحِدٌ، وَإِنَّ أَبَانَكُمْ وَاحِدٌ، أَلَا لَافْضَلٌ لِعَرَبِيٍّ عَلَىٰ أُعْجَمِيٍّ، وَلَا لِعَجَمِيٍّ عَلَىٰ عَرَبِيٍّ، وَلَا لَأَحْمَرَ عَلَىٰ أَسْوَدَ، وَلَا لَأَسْوَدَ عَلَىٰ أَحْمَرَ، إِلَّا بِالتَّقْوَىٰ

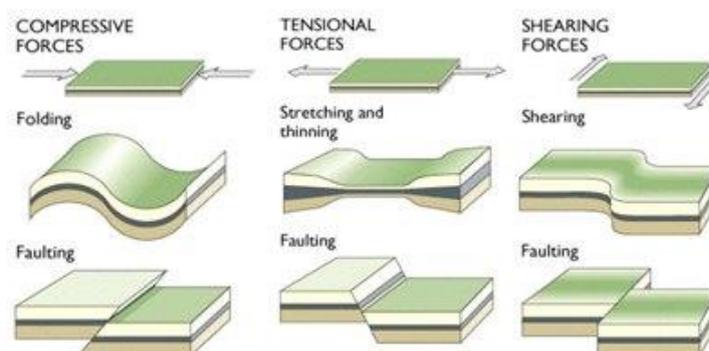
Artinya *"Wahai manusia! Ketahuilah, sesungguhnya Tuhan kalian itu satu, dan sesungguhnya bapak kalian itu satu (yakni Adam). Ketahuilah, tidak ada kelebihan bagi orang Arab atas orang non-Arab ('Ajam), dan tidak ada kelebihan bagi orang non-Arab atas orang Arab; tidak ada kelebihan bagi orang berkulit merah atas orang berkulit hitam, dan tidak ada kelebihan bagi orang berkulit hitam atas orang berkulit merah, melainkan dengan takwa."* (HR. Ahmad dan At-Tirmidzi).

Ayat dan hadis tersebut menegaskan bahwa seluruh manusia berasal dari satu asal dan diciptakan dengan berbagai perbedaan bukan untuk menjadi dasar superioritas, tetapi untuk membangun relasi, saling memahami, dan bekerja sama. Keragaman suku, bangsa, atau warna kulit merupakan bentuk kategorisasi alamiah

yang tidak memiliki kaitan dengan nilai kemuliaan. Satu-satunya ukuran keutamaan di sisi Allah adalah ketakwaan. Dengan demikian, pengelompokan manusia semata-mata berfungsi sebagai sarana mengenali keberagaman, bukan sebagai penentu derajat atau nilai seseorang.

## 2.4 Gempa Bumi

Gempa bumi adalah peristiwa getaran yang terjadi di permukaan bumi yang ditandai dengan retakan atau patahnya lapisan batuan kerak bumi. Energi yang memicu gempa bumi terkumpul dari pergerakan lempeng tektonik dan menyebar ke segala arah dalam bentuk gelombang seismik.



Gambar 2.2 Perubahan dan Pemulihan Elastisitas

Gambar 2.2 menunjukkan bahwa jika dua gaya bekerja secara berlawanan pada batuan di kerak bumi, batuan tersebut akan terdeformasi karena elastisitasnya. Jika gaya ini bertahan lama, daya dukung batuan akan mencapai puncaknya, menyebabkan pergeseran dan patahan cepat sepanjang bidang patahan. Setelah itu, batuan akan stabil dan kembali ke bentuk awalnya melalui proses yang disebut pemulihan elastis, tetapi dengan posisi atau bentuk yang berbeda. Gempa bumi terjadi ketika batuan bergerak dengan cepat sebagai akibat dari pergeseran ini. Dengan itu menunjukkan bahwa asperitas, atau permukaan

kasar pada patahan yang sudah ada, mencegah batu di kedua sisi patahan untuk meluncur. Tekanan mendistorsi batu sampai asperitas pecah, yang melepaskan tekanan dan membuat batu kembali ke bentuk awalnya.

Energi yang dilepaskan oleh gempa bumi dilepaskan dalam bentuk gelombang P (*Primer*) dan gelombang S (*Sekunder*) yang merambat melalui kerak bumi. Gelombang P adalah gelombang kompresional yang dapat melalui baik *medium* padat maupun fluida, sedangkan gelombang S adalah gelombang geser yang hanya dapat melalui *medium* padat dan bergerak lebih lambat daripada gelombang P. Kedua gelombang ini dapat dihitung secara matematis dengan mengikuti rumus berikut:

$$v_p = \sqrt{\frac{k + \frac{4}{3}G}{\rho}} \quad (2.1)$$

$$v_s = \sqrt{\frac{G}{\rho}} \quad (2.2)$$

Keterangan:

$v_p$  : *Primer wave*

$v_s$  : *Sekunder wave*

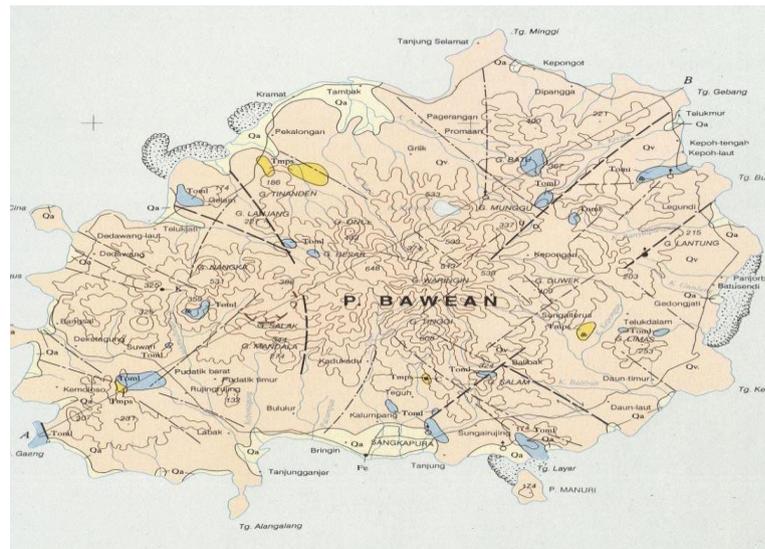
$\rho$  : Densitas ( $kg/m^3$ )

$k$  : Modulus bulk ( $P_a$ )

$G$  : Modulus geser

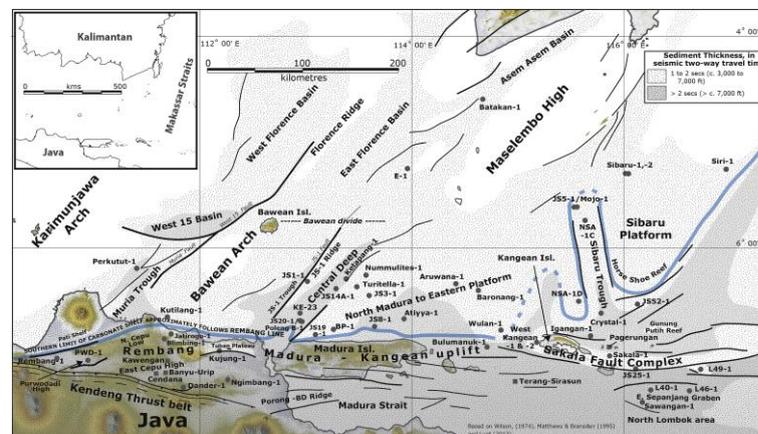
## 2.5 Gempa Merusak Wilayah Bawean

Batuan vulkanik Balibak (Qv) yang terdiri dari perselingan lava, breksi gunung api, dan tuf berumur *Quarter* mendominasi pulau Bawean. Puncak tertinggi terletak di Gunung besar.



Gambar 2.3 Peta Pulau Bawean

Di sisi lain, *basement* Bawean terdiri dari Batugamping Gelam (*Toml*) berumur *Miosen* yang terdiri dari Batugamping terumbu dan klastik. Batupasir kepongan (*Tmps*) yang mengandung kuarsa juga ada.



Gambar 2.4 Peta Struktur Seismik Pulau Bawean

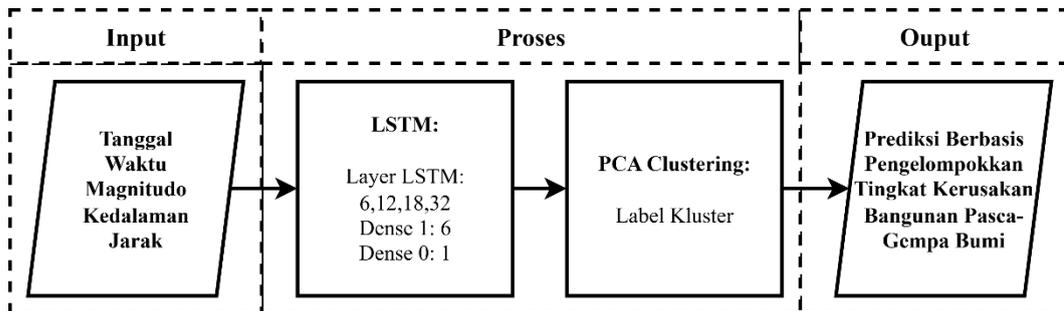
Studi terbatas telah dilakukan tentang sesar yang masih hidup di wilayah Bawean, Indonesia. Namun, tektonik Tersier di area Bawean, seperti yang ditunjukkan oleh Manur dan Barraclough (1994) dan diilustrasikan oleh Lunt (2019), menunjukkan bahwa regangan regional tidak merata, area tersebut dibagi

menjadi beberapa domain yang terpisah oleh sesar geser mendatar besar dan zona ekstensional. Rotasi tektonik dari domain-domain ini, serta deformasi internalnya, disebabkan oleh geser sinistral regional. Ini termasuk *Muriah Trough*, *Bawean Arch*, *Tuban-Camar Trough*, *JS-1 Ridge*, *Central Previous Hit Deep Next Hit Depression*, *Masalembo Basin*, dan *North-Madura Shelf*. Dua fase tektonik utama telah ditemukan: *fase rift Paleogen* dikendalikan oleh satu dan *fase reaktivasi Neogen* dari wilayah Bawean.

## BAB III DESAIN DAN INTEGRASI SISTEM PENELITIAN

### 3.1 Kerangka Konsep

Kerangka konsep yang terdiri dari input data gempa, bangunan, dan lingkungan, yang diproses melalui analisis faktor dan pemodelan prediksi yang menghasilkan *output* tingkat kerusakan bangunan pascagempa pada Gambar 3.1.



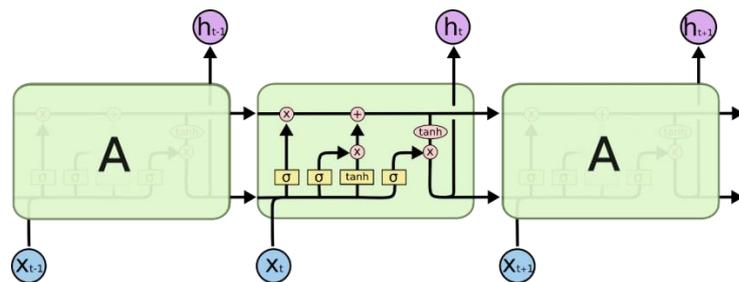
Gambar 3.1 Kerangka Konsep

Pada Gambar 3.1 menunjukkan alur pemrosesan data seismik untuk memprediksi tingkat kerusakan bangunan pascagempa bumi. Terdapat dua jalur pemrosesan data. Jalur pertama menggunakan data sederhana seperti tanggal, epicentrum, kedalaman, dan magnitudo yang langsung diproses dengan model LSTM untuk prediksi. Jalur kedua memanfaatkan data yang lebih lengkap, seperti waktu origin, lokasi, dan dilakukan standarisasi dengan *Standard Scaler*, lalu diproses dengan LSTM dan PCA untuk menghasilkan prediksi yang lebih akurat terhadap tingkat kerusakan bangunan.

### 3.2 *Long Short Term Memory (LSTM)*

*Long Short-Term Memory (LSTM)* adalah jenis jaringan saraf tiruan (*neural network*) yang dikembangkan untuk mengatasi masalah yang sering muncul pada jaringan saraf berulang (*Recurrent Neural Networks, RNN*), khususnya masalah

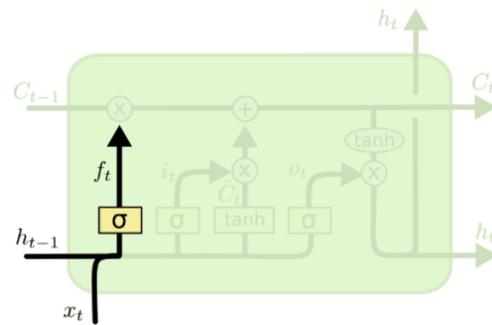
*vanishing gradient*. Diperkenalkan oleh Sepp Hochreiter dan Jürgen Schmidhuber pada tahun 1997, LSTM dirancang untuk memungkinkan jaringan tersebut menyimpan dan mengingat informasi dalam jangka panjang, yang sangat berguna untuk menangani urutan data yang memiliki ketergantungan jangka panjang. LSTM memiliki struktur yang lebih kompleks dibandingkan dengan RNN standar, karena menggabungkan *gates* atau pintu yang mengatur aliran informasi. Ada tiga *gates* utama dalam LSTM: *forget gate*, yang menentukan informasi apa yang akan dibuang dari *state sel*; *input gate*, yang mengontrol informasi baru apa yang akan ditambahkan ke *state sel*; dan *output gate*, yang memutuskan *output* yang akan dihasilkan pada setiap langkah waktu.



Gambar 3.2 Arsitektur *Long Short-Term Memory* (LSTM)

Memutuskan informasi apa yang akan dibuang dari status sel adalah langkah pertama dalam LSTM. "*Forget Get*" adalah lapisan *sigmoid* yang membuat keputusan ini. Lapisan ini memantau  $h_{t-1}$  dan  $x_t$  dan menghasilkan nilai yang berkisar antara 0 dan 1 untuk setiap nilai dalam status sel  $C_{t-1}$ . Angka 1 menunjukkan "*completely keep this*", sementara angka 0 menunjukkan "*completely get rid of this*". Komponen LSTM:

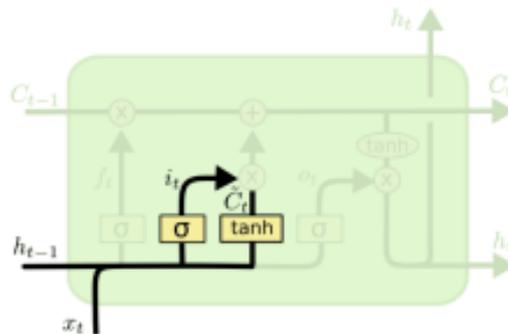
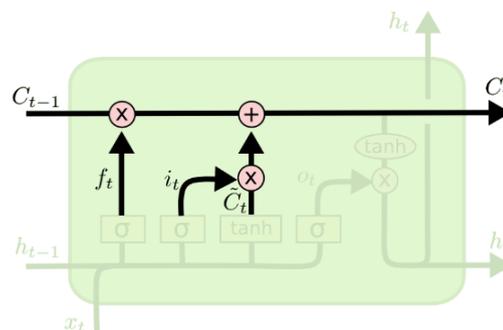
1. *Forget Gate*

Gambar 3.3 Alir Perhitungan *Forget Gate*

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3.1)$$

Menerima informasi yang terdiri dari *state* tersembunyi dari *cell* sebelumnya dan informasi baru dari input saat ini.

## 2. *Input Gate*

Gambar 3.3 Alir *Input Gate* Perkalian Bagian A dan Bagian BGambar 3.4 Alir Penambahan *Forget Gate* dan *Input Gate*

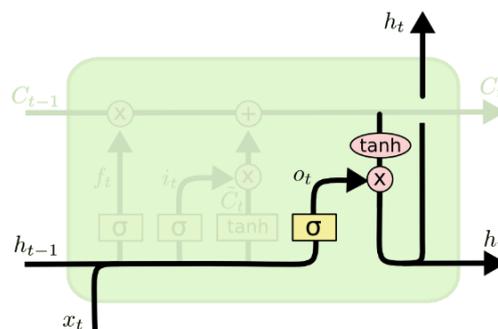
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.2)$$

$$\hat{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3.3)$$

Langkah berikutnya adalah menentukan informasi baru yang akan disimpan dalam status *hidden state* terdapat dua bagian dari proses ini. Nilai mana yang akan diperbarui pertama-tama oleh lapisan *sigmoid* yang dikenal sebagai “*Input Gate*”. Kemudian, fungsi *sigmoid* dan fungsi *tanh* dikalikan untuk membuat vektor nilai kandidat ( $\hat{C}_t$ ) dilanjutkan dengan menggabungkan dan memproses informasi ini. Terakhir, memilih produk apa yang akan dihasilkan *hidden state*.

Meskipun versi yang telah difilter, keluaran ini akan didasarkan pada status *hidden state* pada sel. Pertama-tama menjalankan lapisan *sigmoid* untuk menentukan bagian mana dari status sel yang akan dihasilkan. Kemudian, memasukkan status *hidden state* ke dalam *tanh* (untuk memastikan nilai berada di antara -1 dan 1) dan mengalikannya dengan keluaran gerbang *sigmoid*. Dengan cara ini, akan mengeluarkan bagian *hidden state* yang telah ditentukan.

### 3. *Output Gate*



Gambar 3.5 Alir *Output Gate* (*Sigmoid*) dan *Tanh*

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (3.4)$$

$$h_t = o_t * \tanh(C_t) \quad (3.5)$$

Setelah menerima informasi tentang *hidden state* dari sel sebelumnya dan informasi tentang *hidden state* dari sel baru, *gate output* akan menggabungkan

keduanya dan memproses informasi tersebut dengan fungsi *sigmoid*. Selanjutnya, fungsi tanh akan memproses *hidden state* yang baru, dan hasil dari fungsi tanh akan dikalikan dengan hasil dari fungsi *sigmoid* untuk mendapatkan informasi yang akan disimpan pada *hidden state* yang baru. Setelah itu, kedua *hidden state* dan *cell state* yang baru akan diproses oleh fungsi *sigmoid*.

Keterangan:

$f_t$  : *Output* dari *forget gate* pada waktu  $t$ , yaitu nilai antara 0 dan 1 yang menunjukkan seberapa banyak informasi dari memori sebelumnya yang harus dilupakan.

$\sigma$  : Fungsi aktivasi *sigmoid*, menghasilkan nilai antara 0 dan 1.

$x_t$  : *Input* saat ini (pada waktu  $t$ ), bisa berupa data input (misalnya nilai sensor, kata, angka).

$w_f$  : Bobot (*weight*) yang dikalikan dengan input  $x_t$  untuk *forget gate*.

$h_{t-1}$  : *Hidden state* dari waktu sebelumnya, yaitu informasi tersembunyi yang dibawa dari langkah sebelumnya.

$w_{h-f}$  : Bobot (*weight*) untuk *hidden state* pada *forget gate*.

Keunggulan utama LSTM dibandingkan dengan RNN tradisional adalah kemampuannya untuk menjaga informasi relevan dalam jangka waktu yang lebih lama dan mengurangi masalah *vanishing gradient* yang sering menghambat pelatihan jaringan saraf dengan urutan data panjang. LSTM telah terbukti sangat efektif dalam berbagai aplikasi seperti pemrosesan bahasa alami (NLP), pengenalan suara, dan prediksi deret waktu (*prediction time series*), di mana informasi dari waktu sebelumnya sangat penting untuk membuat keputusan pada

waktu berikutnya. Berkat struktur dan mekanismenya yang unik, LSTM telah menjadi model yang sangat populer dan sering digunakan dalam berbagai bidang seperti terjemahan mesin, analisis sentimen, dan pengenalan pola.

### 3.2.1 Formulasi Matematis Model LSTM

Model *Long Short-Term Memory* (LSTM) pada penelitian ini digunakan untuk mempelajari pola sekuensial dari parameter gempa bumi. Variabel yang digunakan sebagai *input* meliputi:

1. Magnitudo (M)
2. Kedalaman (D)
3. Jarak *epicenter* ke stasiun (R)

Sehingga vektor *input* pada setiap *timestep* ke- $t$ , pada persamaan 3.6:

$$x_t = \begin{bmatrix} M_t \\ D_t \\ R_t \end{bmatrix} \quad (3.6)$$

Semua variabel ini dinormalisasi sebelum digunakan untuk memastikan stabilitas perhitungan.

LSTM terdiri dari tiga gerbang utama: *forget gate*, *input gate*, dan *output gate*.

Secara umum hubungan matematis antar variabel dirumuskan sebagai berikut:

1. *Forget Gate*

Gerbang ini menentukan berapa banyak memori lama yang dipertahankan pada persamaan 3.7:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (3.7)$$

Dengan memasukkan variabel gempa pada persamaan 3.8:

$$W_f x_t = \begin{bmatrix} w_{f,M}^{(1)} M_t + w_{f,D}^{(1)} D_t + w_{f,R}^{(1)} R_t \\ w_{f,M}^{(2)} M_t + w_{f,D}^{(2)} D_t + w_{f,R}^{(2)} R_t \\ w_{f,M}^{(3)} M_t + w_{f,D}^{(3)} D_t + w_{f,R}^{(3)} R_t \end{bmatrix} \quad (3.8)$$

## 2. *Input Gate*

Mengatur jumlah informasi baru yang masuk pada persamaan 3.9:

$$i_t = \sigma(W_{i^x t} + U_{i^h t-1} + b_i) \quad (3.9)$$

Dengan integrasi variabel pada persamaan 3.10:

$$W_{i^x t} = \begin{bmatrix} W_{i,M^M t}^{(1)} + W_{i,D^D t}^{(1)} + W_{i,R^R t}^{(1)} \\ W_{i,M^M t}^{(2)} + W_{i,D^D t}^{(2)} + W_{i,R^R t}^{(2)} \\ W_{i,M^M t}^{(3)} + W_{i,D^D t}^{(3)} + W_{i,R^R t}^{(3)} \end{bmatrix} \quad (3.10)$$

## 3. *Candidate Cell*

Menentukan konten baru yang diusulkan masuk ke memori pada persamaan

3.11:

$$\tilde{c}_t = \tanh(W_{g^x t} + U_{g^h t-1} + b_g) \quad (3.11)$$

Dengan integrasi variabel pada persamaan 3.12:

$$W_{g^x t} = \begin{bmatrix} W_{g,M^M t}^{(1)} + W_{g,D^D t}^{(1)} + W_{g,R^R t}^{(1)} \\ W_{g,M^M t}^{(2)} + W_{g,D^D t}^{(2)} + W_{g,R^R t}^{(2)} \\ W_{g,M^M t}^{(3)} + W_{g,D^D t}^{(3)} + W_{g,R^R t}^{(3)} \end{bmatrix} \quad (3.12)$$

## 4. *Pembaruan Cell State*

$$C_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \quad (3.13)$$

Pada persamaan 3.13 ini, tiga variabel gempu berperan langsung pada pembentukan memori model melalui pengaruhnya terhadap  $i_t$  dan  $\tilde{c}_t$ .

## 5. *Output Gate dan Hidden State*

$$o_t = \sigma(W_{o^x t} + U_{o^h t-1} + b_o) \quad (3.14)$$

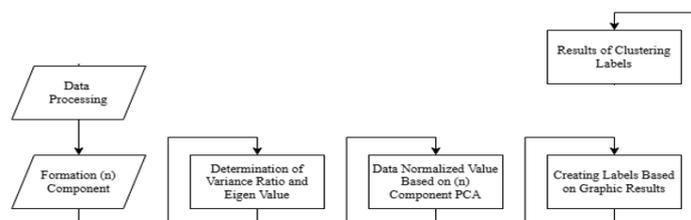
$$h_t = o_t \times \tanh(c_t) \quad (3.15)$$

Ketiga variabel input Magnitudo, Kedalaman, dan Jarak diintegrasikan langsung ke dalam setiap komponen LSTM melalui *matriks* bobot  $W_i$ ,  $W_f$ ,  $W_g$ , dan  $W_o$  dengan persamaan 3.14 dan 3.15. Dengan demikian, setiap perubahan

pada nilai ketiga variabel tersebut akan memengaruhi nilai *gate* dan *hidden state* pada setiap *timestep*. Hal ini membuat LSTM mampu mempelajari pola temporal gempa bumi secara adaptif.

### 3.3 *Principal Component Analysis (PCA)*

*Principal Component Analysis (PCA)* adalah teknik statistik yang digunakan untuk mereduksi dimensi data dengan mengidentifikasi komponen utama yang paling signifikan, yang dapat menjelaskan sebagian besar variabilitas dalam data. PCA dimulai dengan melakukan standarisasi data, jika diperlukan, untuk memastikan setiap variabel memiliki skala yang sama. Selanjutnya, dihitung *matriks kovarians* untuk menggambarkan hubungan antar variabel, kemudian dilakukan dekomposisi *eigen* untuk menemukan *eigenvector* dan *eigenvalue*. *Eigen vector* yang memiliki *eigenvalue* terbesar dipilih sebagai komponen utama, yang mencerminkan arah dengan variansi terbesar dalam data. Proses ini memungkinkan kita untuk memproyeksikan data ke ruang komponen utama yang lebih rendah dimensinya, tetapi tetap mempertahankan informasi yang relevan. PCA sangat berguna dalam analisis data dengan dimensi tinggi, memungkinkan pengurangan kompleksitas dan meningkatkan efisiensi dalam pemrosesan data untuk aplikasi seperti *machine learning* dan visualisasi data.



Gambar 3.9 Alir *Principal Component Analysis (PCA)-Clustering*

Langkah-langkah perhitungan algoritma PCA *Clustering* sebagai berikut:

1. Mempersiapkan Data

Mengumpulkan data Gempa Bumi yang dilakukan untuk menilai kerusakan bangunan pascagempa bumi pada sektor wilayah Jawa Timur dan Jawa Tengah.

2. Menormalkan Data

Selanjutnya, melakukan proses normalisasi data untuk memastikan bahwa data tersebut sesuai dengan standar PCA (*Principal Component Analysis*). Normalisasi adalah langkah penting untuk membuat data menjadi sebanding dan memiliki derajat yang sama. Rumus untuk menghitung nilai standardisasi pada persamaan 3.16 sebagai berikut:

$$X_{standart} = \frac{X - \text{mean}(X)}{\text{standartdeviation}(X)} \quad (3.16)$$

Di mana nilai rata-rata (*mean*) dihitung dengan persamaan 3.17:

$$\underline{x} = \sum_{i=1}^n Xi/n \quad (3.17)$$

3. Menentukan Rasio Varian dan Nilai *Eigen*

PCA bertujuan mengurangi dimensi data sambil mempertahankan sebanyak mungkin informasi dalam persamaan 3.18:

$$s^2 = \sum_{i=1}^n (Xi - \underline{x})^2 / (n - 1) \quad (3.18)$$

4. Menentukan Jumlah Komponen Utama (PC)

Tentukan jumlah komponen utama (*n*) berdasarkan rasio varian kumulatif yang signifikan (biasanya 85-95% dari total varian). Pilih *n* nilai *eigen* terbesar untuk membentuk komponen utama.

5. Membuat Visualisasi

#### 6. Membuat Rentang Nilai

Rentang nilai dapat dilakukan dengan memaksimalkan nilai tertinggi dalam suatu set data dan meminimalkan hilangnya informasi yang disebabkan oleh reduksi data. Dalam penelitian kami, untuk membuat rentang nilai menggunakan hasil dari PC.

#### 7. Melakukan *Clustering*

*Clustering* merupakan suatu teknik pembelajaran yang membagi data menjadi beberapa bagian data tanpa pengawasan menjadi beberapa kelompok data yang homogen. Dalam penelitian kami untuk melakukan *clustering* data menggunakan data titik koordinat PC berdasarkan rentang nilai yang telah ditentukan. Hasil data *clustering* memerlukan validasi untuk mengetahui tingkat kebenarannya.

#### 8. Menganalisis Hasil

Interpretasi setiap *Cluster* berdasarkan karakteristiknya. Hubungkan hasil analisis dengan dampaknya pada sektor-sektor yang dipengaruhi gempa bumi di wilayah tersebut.

#### 9. Membuat Rentang Nilai

Untuk mendukung interpretasi, rentang nilai berdasarkan karakteristik data:

- Kerusakan ringan:  $< 0$
- Kerusakan sedang:  $0 \leq n \leq 2$
- Kerusakan berat  $i: > 2$

### 3.4 Persiapan Data

Data yang digunakan pada penelitian ini adalah data gempa bumi yang merupakan data *time series* yang berupa signal yang diolah pada *Seiscom* jaringan

*InaTEWS Multi Stasiun* untuk mendapatkan parameter-parameter gempa seperti *origin time*, koordinat, magnitudo, kedalaman dan jarak pada [repogempa.bmkg.go.id/repo\\_new/](http://repogempa.bmkg.go.id/repo_new/) yang diakses pada tanggal 12 Maret 2025 sebanyak 483 data parameter gempa bumi periode tahun 2024 yang terjadi di Wilayah Pusat Gempa khususnya pada kepulauan Bawean yang terdiri 5 variabel dalam penelitian data gempa bumi seperti pada Tabel 3.1.

Tabel 3.1 Data Gempa Bumi

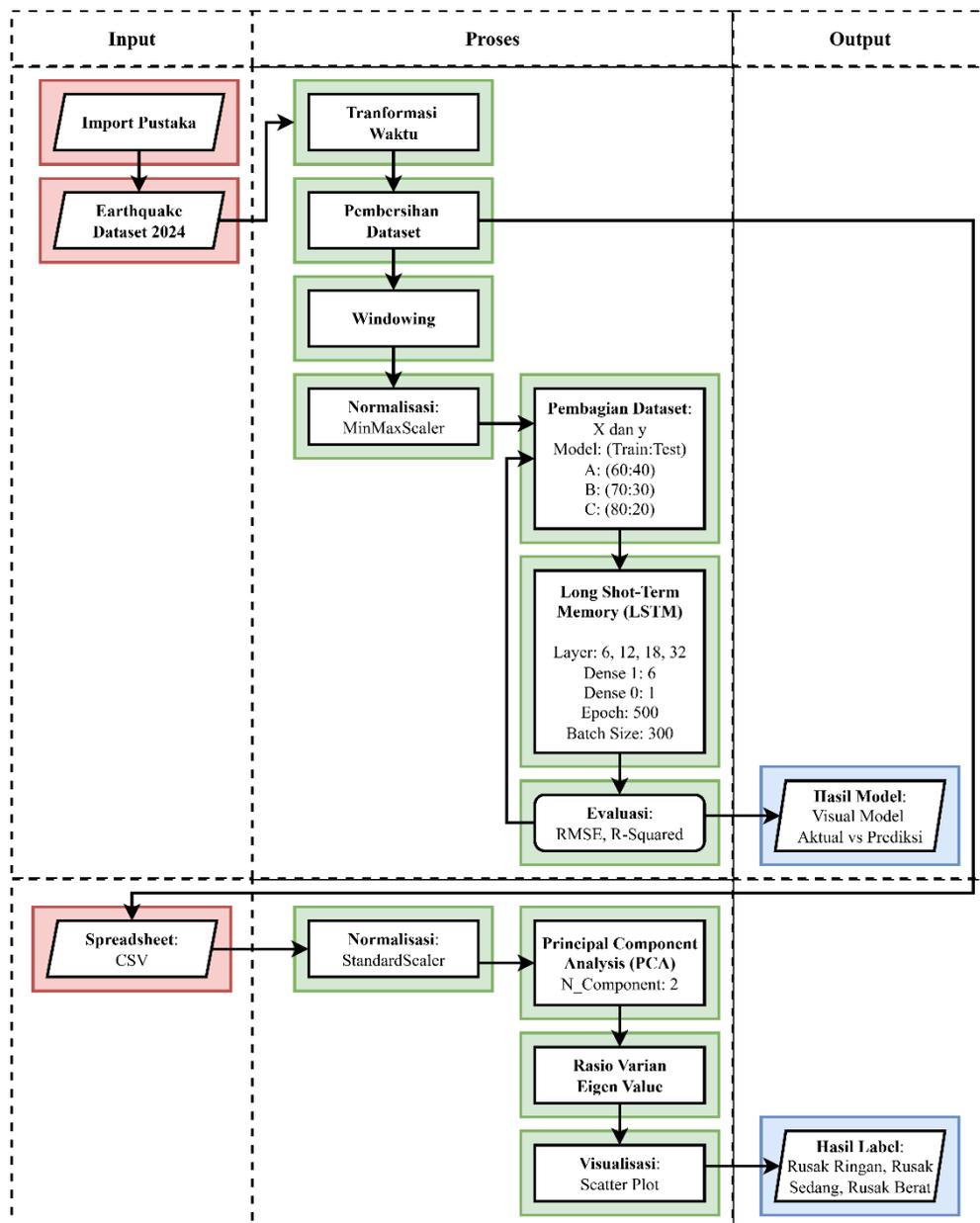
Tanggal	Waktu	Magnitudo	Kedalaman	Jarak
22/03/2024	15:52:58	6.5	12	114
22/03/2024	11:22:45	5.9	10	126
23/03/2024	19:56:57	4.6	12	101
22/03/2024	12:31:13	5.5	10	139
22/03/2024	23:18:01	3.0	10	102
...	...	...	...	...
29/03/2024	02:45:50	2.7	16	165
28/03/2024	21:35:52	2.8	19	153
29/03/2024	02:47:58	2.6	19	150
27/03/2024	10:26:10	3.2	24	148
22/03/2024	13:10:02	2.8	35	155

Tabel 3.2 Keterangan Variabel data

No	Nama Variable	Satuan	Keterangan Ilmiah
1	<i>Date</i>	Hari	Tanggal terjadinya gempa bumi.
2	<i>Time</i>	<i>Universal Time</i>	Jam terjadinya gempa bumi dalam UT
3	<i>Depth</i>	Kilometer	Kedalaman sumber gempa bumi
4	Magnitudo	<i>Skala Richter</i>	Nilai kekuatan gempa bumi
5	<i>Range</i>	Kilometer	Jarak sumber gempa bumi dari kota terdekat

### 3.5 Desain Sistem

Desain sistem pada penelitian ini dirancang untuk menjelaskan alur kerja dalam menganalisis dan memprediksi dampak gempa bumi, khususnya pada tingkat kerusakan bangunan seperti pada Gambar 3.6.



Gambar 3.6 Desain Sistem

Desain sistem ini pada Gambar 3.6 menjelaskan alur kerja teknis implementasi *prediction-based clustering* yang dibangun menggunakan bahasa pemrograman *Python*. Proses ini mencakup dua fasa utama: *Fase Prediksi (LSTM)* dan *Fase Clustering (PCA-Clustering)*, yang dijabarkan dalam langkah-langkah komputasi berikut:

### 3.5.1 Step 1: Inisialisasi dan Pemuatan Data

Tahap awal adalah inisialisasi lingkungan kerja dengan mengimpor pustaka (*libraries*) Python yang *esensial*, seperti *Pandas* dan *NumPy* untuk manipulasi data, *TensorFlow/Keras* untuk pemodelan *deep learning* (LSTM), dan *Scikit-learn* untuk pra-pemrosesan dan pemodelan *machine learning* (PCA & Clustering). Setelah pustaka dimuat, dataset gempa bumi tahun 2024 diimpor ke dalam sistem (misalnya, dari file *.xlsx* atau *.csv*) untuk persiapan pra-pemrosesan. Data mencakup lima parameter: tanggal (*Date*) dan waktu kejadian (*Original Time/OT*) untuk melacak temporalitas gempa, kedalaman (*Depth* dalam kilometer) yang diukur menggunakan jaringan Indonesia Tsunami *Warning System (InaTEWS)*, magnitudo (*Magnitude* dalam *Skala Richter*) sebagai indikator kekuatan gempa (Bai & Tahmasebi, 2022), jarak dari kota terdekat (*Range* dalam kilometer), azimuth (derajat arah relatif terhadap kota).

### 3.5.2 Step 2: Pra-Pemrosesan Data (Persiapan LSTM)

Tahap ini krusial untuk menyiapkan data mentah agar siap digunakan oleh model LSTM. Proses ini mencakup:

1. Transformasi Waktu

Memastikan kolom tanggal dan waktu (*timestamp*) memiliki format yang konsisten dan dapat diproses sebagai data runtun waktu.

2. Pembersihan Data

Melakukan pemeriksaan terhadap dataset untuk mengidentifikasi dan menangani nilai yang hilang atau kosong (*missing values*) yang dapat mengganggu proses pelatihan.

3. *Windowing*

Proses inti untuk data runtun waktu, di mana data distrukturkan ulang menjadi format *sekuens* (x, y) yang dapat digunakan oleh model LSTM. Ini mengubah data mentah menjadi "jendela-jendela" data (misalnya, menggunakan data 10 hari sebelumnya untuk memprediksi 1 hari ke depan).

#### 4. Normalisasi (LSTM)

Data yang telah di-*windowing* kemudian dinormalisasi menggunakan *MinMaxScaler*. Proses ini mengubah skala seluruh fitur data ke rentang tertentu (misalnya, 0 hingga 1), yang sangat penting untuk menstabilkan dan mempercepat konvergensi pelatihan model LSTM.

### 3.5.3 Step 3: Proses Model Prediksi (LSTM)

Ini adalah fasa *deep learning* untuk menghasilkan prediksi.

#### 1. Pembagian Dataset

Data yang telah dinormalisasi dibagi menjadi data latih (*train*) dan data uji (*test*) berdasarkan tiga skenario yang ditentukan (Model A: 60:40, Model B: 70:30, Model C: 80:20).

#### 2. *Tuning* Parameter LSTM

Model LSTM dibangun dan dilatih dengan melakukan *tuning* parameter. Arsitektur yang diuji meliputi variasi jumlah unit pada LSTM *layer* 1 (6, 12, 18, 32), *Dense layer* 1 (6 unit), dan *dense layer* 0 (1 unit sebagai *output*). Model dilatih selama 500 *epoch* dengan *batch size* 300 (Chaurasia et al., 2019).

#### 3. Evaluasi Model

Setelah pelatihan, *output* model berupa visualisasi (*plot loss* dan akurasi) dianalisis. Performa model dievaluasi secara kuantitatif pada data uji menggunakan metrik MAE, LSE, RMSE (*Root Mean Squared Error*) dan R-

*Squared* ( $R^2$ ) untuk menentukan model LSTM terbaik (Chencho et al., 2024).

#### **3.5.4 Step 4: *Input* Fitur dan Normalisasi (Persiapan PCA)**

Tahap ini adalah jembatan antara Fasa Prediksi dan Fasa Kluster.

##### 1. Ekspor Nilai Prediksi

Nilai-nilai prediksi yang dihasilkan oleh model LSTM terbaik (yang berasal dari teknik *windowing*) diekspor dari model dan disimpan sebagai dataset baru (misalnya dalam format *.xlsx* atau *spreadsheet*). Dataset inilah yang akan menjadi *input* untuk Fasa *Clustering*.

##### 2. Normalisasi (PCA-*Clustering*)

Berbeda dengan LSTM, dataset baru ini dinormalisasi menggunakan *StandardScaler*. Ini bertujuan untuk mengubah distribusi data sehingga memiliki rata-rata (*mean*) 0 dan deviasi standar (*standard deviation*) 1, yang merupakan prasyarat penting agar PCA dapat berfungsi secara optimal (karena PCA sensitif terhadap skala data).

Model LSTM yang telah terlatih digunakan untuk memprediksi magnitudo dan kedalaman gempa dalam rentang waktu tertentu. Hasil kluster dibandingkan dengan data aktual menggunakan metrik evaluasi:

1. Ekternal : Data Observasi
2. Internal : *Silhouette Score*

#### **3.5.5 Step 5: Proses PCA-*Clustering***

Ini adalah fasa *machine learning* untuk pengelompokan.

##### 1. Pemrosesan PCA

Data prediksi yang telah dinormalisasi (*StandardScaler*) dimasukkan ke dalam algoritma PCA. Parameter PCA diatur dengan  $n\_component=2$ , yang bertujuan

untuk mereduksi dimensi data menjadi dua komponen utama (PC1 dan PC2) yang paling representatif.

Analisis Varian: Rasio varian (*variance ratio*) dan nilai *eigen* (*eigenvalue*) dari hasil PCA diperiksa untuk memastikan bahwa kedua komponen utama (PC1 dan PC2) mampu menangkap sebagian besar informasi (*varians*) dari data asli (Gu et al., 2024).

## 2. Visualisasi dan Pelabelan Kluster

*Output* dari PCA divisualisasikan menggunakan *scatterplot* 2D untuk mengamati sebaran dan pemisahan data (Bodenmann et al., 2023). Berdasarkan distribusi data pada sumbu *Principal Component* 1 (PC1), dilakukan pelabelan untuk menentukan tiga kluster (label) kerusakan: rusak ringan, rusak sedang, dan rusak berat (Bi et al., 2024).

### 3.6 Skenario Evaluasi

Dalam penelitian Almais et al. (2024), PCA digunakan untuk mengelompokkan data tingkat kerusakan sektor pascabencana. Data yang digunakan mencakup 216 entri dengan 5 kriteria kerusakan. PCA berhasil mengidentifikasi tiga klaster berdasarkan nilai koordinat komponen utama (PC1 dan PC2). Meskipun penelitian (Almais et al., 2023) tidak menerapkan *prediction-based clustering*, tetapi pada penelitian ini secara eksplisit nilai RMSE, MAPE, MAE, atau  $R^2$ , pendekatan ini menunjukkan potensi penggunaan PCA dalam klusterisasi tingkat kerusakan pascagempa Penggunaan *Long Short-Term Memory* (LSTM) dan *Principal Component Analysis* (PCA)-*clustering* kerusakan bangunan pascagempa bumi memungkinkan analisis data yang kompleks dengan

lebih efisien.

### 3.7 Validasi Internal (*Prediction*)

#### 1. MAE

*Mean Absolute Error* (MAE) mengukur rata-rata besarnya kesalahan tanpa mempertimbangkan arah kesalahan tersebut:

$$MAE = \left(\frac{1}{2}\right) \sum |y_i - \hat{y}_i| \quad FS1 \quad (3.19)$$

keterangan:

$y_i$  mewakili nilai aktual dan  $\hat{y}_i$  mewakili nilai yang diprediksi.

#### 2. RMSE

*Root Mean Square Error* (RMSE) menilai kesalahan yang lebih besar lebih berat daripada MAE:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (3.20)$$

Metrik ini untuk mengidentifikasi model dengan deviasi prediksi yang besar.

#### 3. $R^2$ (*Kefisien Determinasi*)

*Koefisien Determinasi* ( $R^2$ ) menunjukkan proporsi varians dalam variabel dependen yang dijelaskan oleh model:

$$R^2 = 1 - \left(\frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2}\right) \quad (3.21)$$

Di mana  $\bar{y}$  adalah rata-rata dari nilai-nilai yang diamati.

#### 4. *Silhouette Score*

*Silhouette Score* (Punhani et al., 2022) merupakan metrik yang digunakan untuk mengevaluasi kualitas *clustering* dengan mengukur seberapa baik data dalam sebuah *cluster* serupa satu sama lain (*koherensi*) dan seberapa baik *cluster*

tersebut terpisah dari *cluster* lainnya (*separasi*).

Nilai untuk *Silhouette Score* mengevaluasi sejauh mana data dalam *cluster* serupa dan terpisah dengan *cluster* lain. Nilai berkisar antara -1 (buruk) hingga +1 (baik). Nilai +1 merupakan indikasi *Clustering* sangat baik. Data berada di *cluster* yang benar dan jauh dari *cluster* lainnya. Nilai 0 merupakan data berada di batas dua *cluster*, menunjukkan *ambiguitas*. Serta, Nilai -1 merupakan *Clustering* buruk. Data lebih dekat ke *cluster* yang salah daripada ke *cluster* yang benar.

Rumus untuk *Silhouette Score* untuk setiap data  $i$  dihitung dengan persamaan 3. 14 (Punhani et al., 2022):

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3. 22)$$

Keterangan:

$a(i)$  : Rata-rata jarak data  $i$  ke semua data lain dalam *cluster* yang sama (*internal cohesiveness*).

$b(i)$  : Rata-rata jarak data  $i$  ke semua data dalam *cluster* terdekat lainnya (*external separability*).

$\max(a(i), b(i))$  : Nilai maksimum antara  $a(i)$  dan  $b(i)$  untuk normalisasi.

Langkah-langkah perhitungan *Silhouette Score*:

1. Menghitung *Euclidian Distance*

$$d_n(P_{target}, P_n) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (3. 23)$$

Keterangan:

$d_n$  : Jarak yang akan dijadikan sebagai objek.

$P_{target}$  : Titik data yang dijadikan sebagai objek.

$P_{target}$  : Titik data selanjutnya yang akan dijadikan sebagai objek lain.

2. Menghitung *Cohesiveness*  $a(i)$

Untuk setiap data  $i$  dalam *cluster*  $C$ , hitung rata-rata jarak *Euclidean* ke semua data lain dalam *cluster*.

$$a(i) = \frac{d_{target} + d_n}{2} \quad (3.24)$$

Keterangan:

$a(i)$  : Rata-rata jarak objek  $i$  ke semua objek lain dalam *cluster* yang sama (*intra-cluster distance*).

$d_n$  : Jarak target ke jarak selanjutnya atau  $n$ .

### 3. Menghitung *Separability* $b(i)$

Untuk setiap data  $i$ , identifikasi *cluster* lain  $C'$  yang terdekat. Hitung rata-rata jarak data  $i$  ke semua data dalam *cluster*  $C$ .

### 4. Menghitung *Silhouette Score* Rata-rata untuk semua data

Setelah menghitung  $S(i)$  untuk setiap data, hitung rata-rata semua  $S(i)$  untuk mendapatkan skor keseluruhan.

$$S_{avg} = \frac{1}{N} \sum_{i=1}^N S(i) \quad (3.25)$$

### 5. Interpretasi Hasil

- $S_{avg} > 0.5$  : *Clustering* sangat baik. Data dalam *cluster* sangat kohesif dan terpisah dengan jelas.
- $0.25 < S_{avg} \leq 0.5$  : *Clustering* cukup baik tetapi mungkin ada *overlap* antar-*cluster*.
- $0.25 < S_{avg}$  : *Clustering* kurang baik. Data tidak terbagi dengan baik ke dalam *cluster*.

### 3.8 Instrument Penelitian

Instrumen penelitian merupakan bagian terpenting dalam penelitian karena menentukan baik atau tidaknya hasil yang diperoleh. Pada penelitian ini menggunakan instrumen penelitian berupa data gempa bumi [repogempa.bmkg.go.id/repo\\_new/](http://repogempa.bmkg.go.id/repo_new/) pusat gempa yang meliputi Jawa Tengah dan Jawa Timur tahun 2024. Penulis menggunakan variabel dependen berupa *Magnitudo*, *Kedalaman* dan variabel dependen adalah *cluster* gempa bumi.

## BAB IV UJI COBA

### 4.1 Skenario Uji Coba

Pada tahap uji coba awal, penelitian ini menerapkan Skenario A, yaitu pembagian dataset menjadi 60% data latih dan 40% data uji. Pemilihan skenario ini bertujuan untuk mengevaluasi performa dasar model LSTM ketika jumlah data latih relatif lebih sedikit dibandingkan skenario lainnya. Dengan demikian, Skenario A berfungsi sebagai *baseline* untuk mengukur kestabilan dan sensitivitas model sebelum dilakukan pengujian lebih lanjut pada skema pembagian data yang lebih besar (70:30 dan 80:20).

Sebagai langkah pertama dalam Skenario A, dilakukan pengujian terhadap Model A1, yang merupakan arsitektur LSTM paling sederhana dalam kelompok ini. Model A1 menggunakan konfigurasi 6 unit LSTM pada lapisan utama, 6 unit *Dense layer*, 1 unit *output*, 500 *epoch* dan *batch size* 300. Konfigurasi ini dipilih untuk mengidentifikasi kemampuan dasar LSTM dalam mempelajari pola temporal data gempa bumi tanpa kompleksitas arsitektur yang berlebihan.

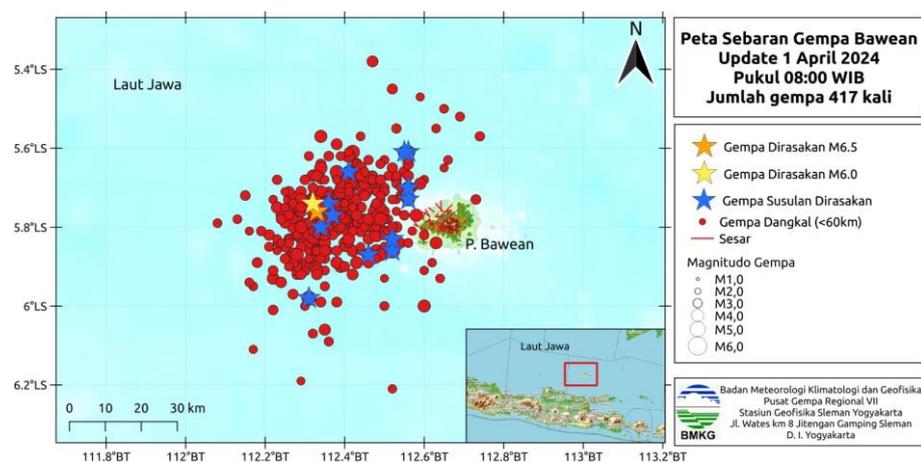
Pada tahap uji coba Model A1, seluruh proses pelatihan mengacu pada data yang telah dinormalisasi dan dibentuk dalam struktur *windowing*, sehingga model menerima *input* runtun waktu yang konsisten. Proses pelatihan menghasilkan grafik akurasi dan *error* yang menjadi indikator awal untuk melihat apakah model menunjukkan pola konvergensi atau masih menghadapi ketidakstabilan dalam mempelajari hubungan antarvariabel gempa bumi.

Hasil evaluasi awal terhadap Model A1 menunjukkan kemampuan yang cukup baik sebagai model dasar. Model ini memperoleh *Train RMSE* = 0.2338, *Test*

RMSE = 0.3592, *Train*  $R^2 = 0.8288$ , *Test*  $R^2 = 0.3541$  yang mengindikasikan bahwa meskipun performa pada data uji belum optimal, model tetap mampu menangkap pola umum variabilitas parameter gempa bumi. Nilai  $R^2$  pada data uji sebesar 35.41% menjadi gambaran awal mengenai kemampuan generalisasi model sederhana pada *rasio* data latih 60%.

Secara keseluruhan, uji coba Model A1 ini menjadi dasar untuk memahami bagaimana model berperilaku pada kondisi data latih yang terbatas. Hasil dari skenario awal ini kemudian dijadikan acuan untuk melakukan pengujian lanjutan pada model lain dalam Skenario A (A2, A3, A4), sekaligus menjadi pijakan sebelum melanjutkan ke 4.2 Skenario Uji Hasil, di mana seluruh variasi model akan dibandingkan secara komprehensif dalam satu skema evaluasi.

Pada tahap awal penelitian, dilakukan pengolahan data gempa bumi yang meliputi Magnitudo, Kedalaman, dan Jarak *epicenter* terhadap stasiun pemantauan. Data dianalisis secara deskriptif untuk memahami gambaran umum karakteristik gempa bumi di wilayah penelitian.



Gambar 4.1 Sebaran Gempa Bawean (sumber: BMKG)

Nilai magnitudo yang tercatat umumnya berada pada kisaran rendah hingga

menengah, dengan kedalaman dangkal hingga menengah pula. Jarak *epicenter* terhadap stasiun menunjukkan variasi yang cukup lebar, yang berpengaruh terhadap amplitudo sinyal dan relevansi informasi temporal.



Gambar 4.2 Visualisasi Magnitudo, Kedalaman, dan Jarak

Hasil visualisasi awal berupa grafik sekuens Magnitudo, Kedalaman, dan Jarak memperlihatkan bahwa ketiga variabel tersebut memiliki pola fluktuatif yang tidak *linier*, sehingga membutuhkan model yang mampu menangkap pola jangka pendek dan panjang, seperti LSTM. Selain itu, dilakukan normalisasi data menggunakan skala *min-max* agar seluruh variabel berada pada rentang yang seragam, sehingga mendukung stabilitas proses pelatihan model.

## 4.2 Hasil Pelatihan Model Uji Coba

Hasil prediksi dari model LSTM kemudian akan menjadi data masukan (*input*) utama untuk tahap *Prediction-based clustering* menggunakan *PCA-Clustering* yang akan dibahas pada sub-bab selanjutnya.

Untuk mendapatkan model *Prediction* yang optimal, dilakukan serangkaian pengujian dengan skenario yang berbeda.

### 1. Hasil Pengujian Skenario A (60% *Train* : 40% *Test*)

Skenario A menguji empat variasi arsitektur LSTM (A1, A2, A3, A4) dengan

pembagian data 60% latih dan 40% uji.

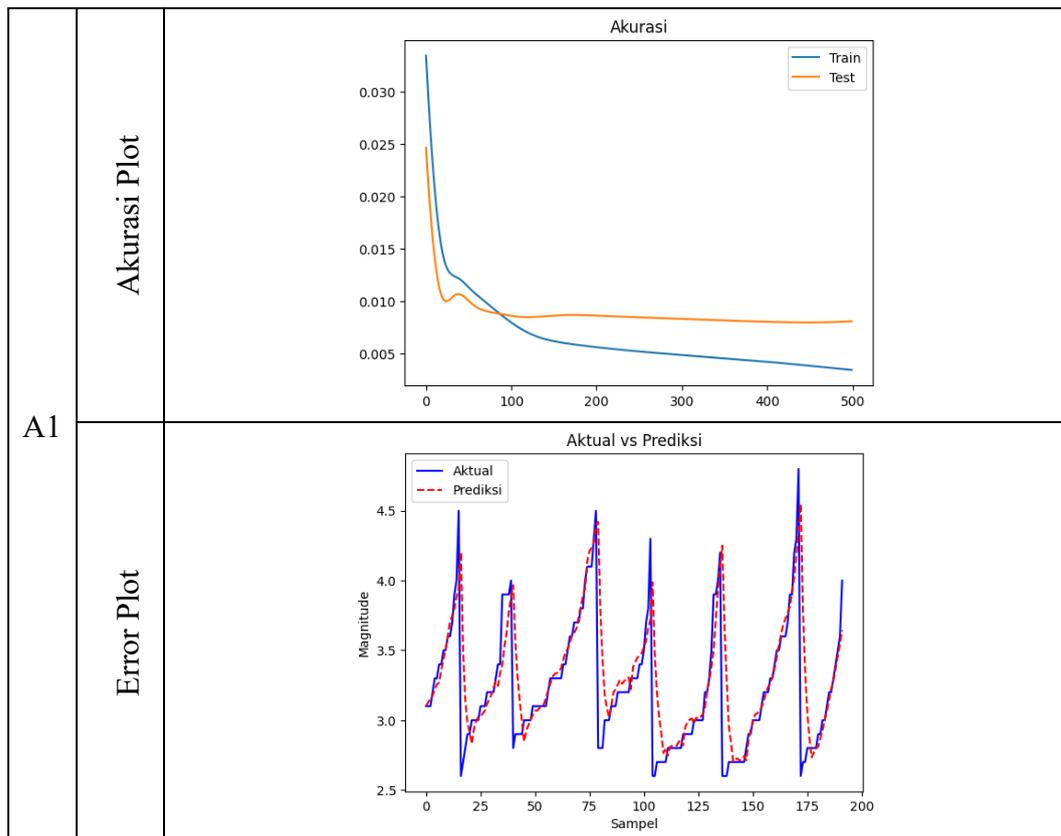
Tabel 4.1 Parameter Model Skenario A

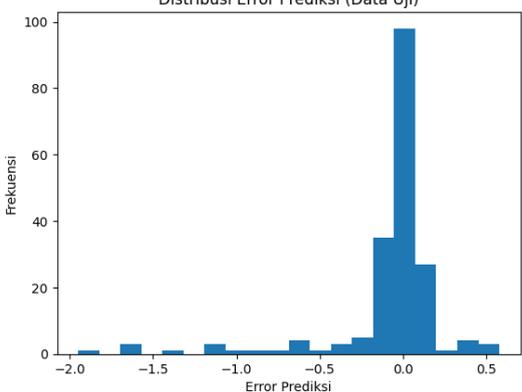
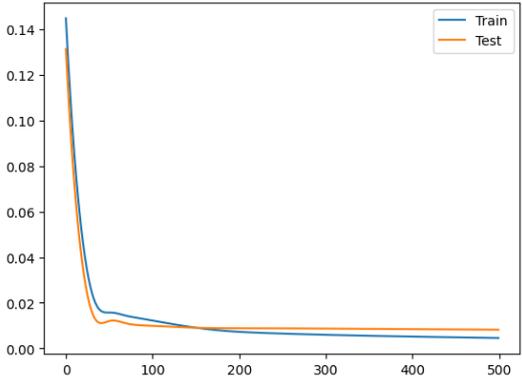
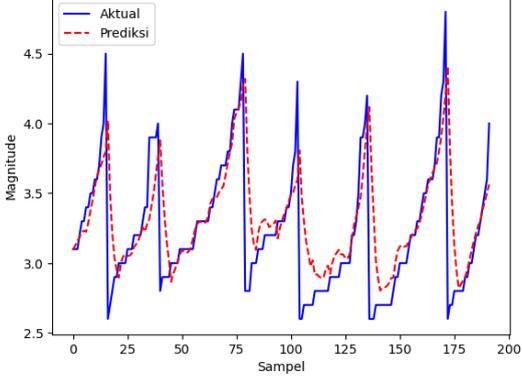
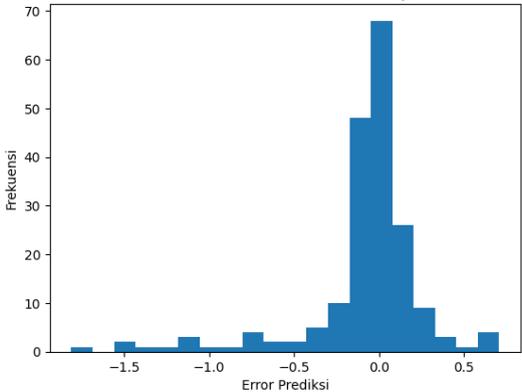
Model	LSTM Layer 1	Dense Layer 1	Dense Layer 0	Epoch	Batch Size
A1	6	6	1	500	300
A2	12	6	1	500	300
A3	18	6	1	500	300
A4	32	6	1	500	300

Tabel 4.2 Hasil Evaluasi Model Skenario A

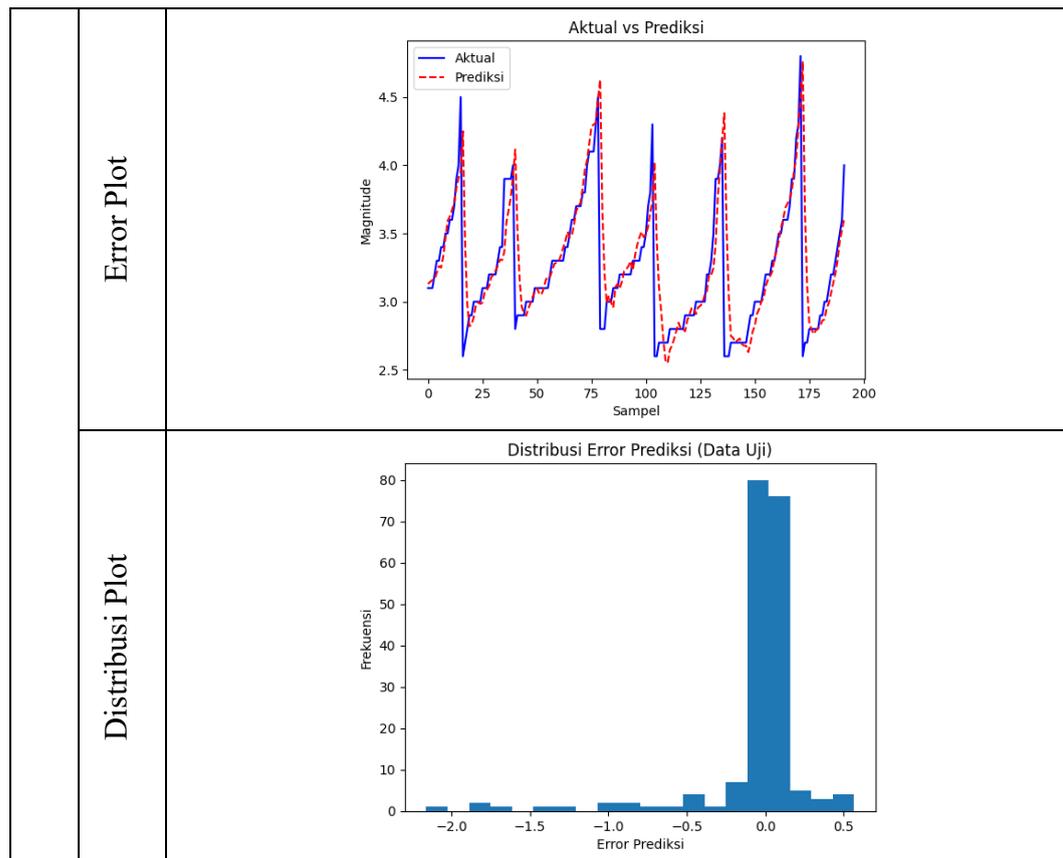
Model	Train RMSE	Test RMSE	Train R-Squared	Test R-Squared
A1	0.2338	0.3592	0.8288	0.3541
A2	0.2701	0.3626	0.7715	0.3419
A3	0.1979	0.3512	0.8773	0.3823
A4	0.1970	0.3635	0.8784	0.3384

Hasil plot visual untuk Akurasi, *Error*, dan Distribusi data dari setiap model Skenario A disajikan di bawah ini.



	Distribusi Plot	 <p>Distribusi Error Prediksi (Data Uji)</p> <p>This histogram shows the frequency distribution of prediction errors for the test data. The x-axis represents the prediction error, ranging from -2.0 to 0.5. The y-axis represents the frequency, ranging from 0 to 100. The distribution is highly concentrated around 0.0, with a peak frequency of approximately 95.</p>
A2	Akurasi Plot	 <p>Akurasi</p> <p>This line plot shows the accuracy of the model over 500 samples. The x-axis is labeled 'Sampel' and ranges from 0 to 500. The y-axis represents accuracy, ranging from 0.00 to 0.14. Two lines are plotted: 'Train' (blue) and 'Test' (orange). Both lines show a rapid increase in accuracy, reaching a plateau of approximately 0.01 after about 100 samples.</p>
	Error Plot	 <p>Aktual vs Prediksi</p> <p>This line plot compares actual magnitude (blue solid line) with predicted magnitude (red dashed line) over 200 samples. The x-axis is labeled 'Sampel' and ranges from 0 to 200. The y-axis is labeled 'Magnitude' and ranges from 2.5 to 4.5. The plot shows several peaks in the actual data, with the predicted data following the general trend but exhibiting some lag and smoothing.</p>
	Distribusi Plot	 <p>Distribusi Error Prediksi (Data Uji)</p> <p>This histogram shows the frequency distribution of prediction errors for the test data. The x-axis represents the prediction error, ranging from -1.5 to 0.5. The y-axis represents the frequency, ranging from 0 to 70. The distribution is centered around 0.0, with a peak frequency of approximately 68.</p>

A3	Akurasi Plot	
	Error Plot	
	Distribusi Plot	
	Akurasi Plot	
A4	Akurasi Plot	



Berdasarkan Tabel 4.2, Model A3 (dengan 18 unit LSTM) menunjukkan performa terbaik pada Skenario A. Model ini memiliki *Test RMSE* terendah (0.3512) dan *Test R-Squared* tertinggi (0.3823). Ini berarti model A3 memiliki tingkat kesalahan prediksi paling kecil dan mampu menjelaskan sekitar 38.23% varians pada data uji, mengungguli arsitektur lainnya dalam skenario 60:40 ini.

## 2. Hasil Pengujian Skenario B (70% Train : 30% Test)

Skenario B menguji empat variasi arsitektur LSTM (B1, B2, B3, B4) yang sama, namun dengan alokasi data latih yang lebih besar (70%).

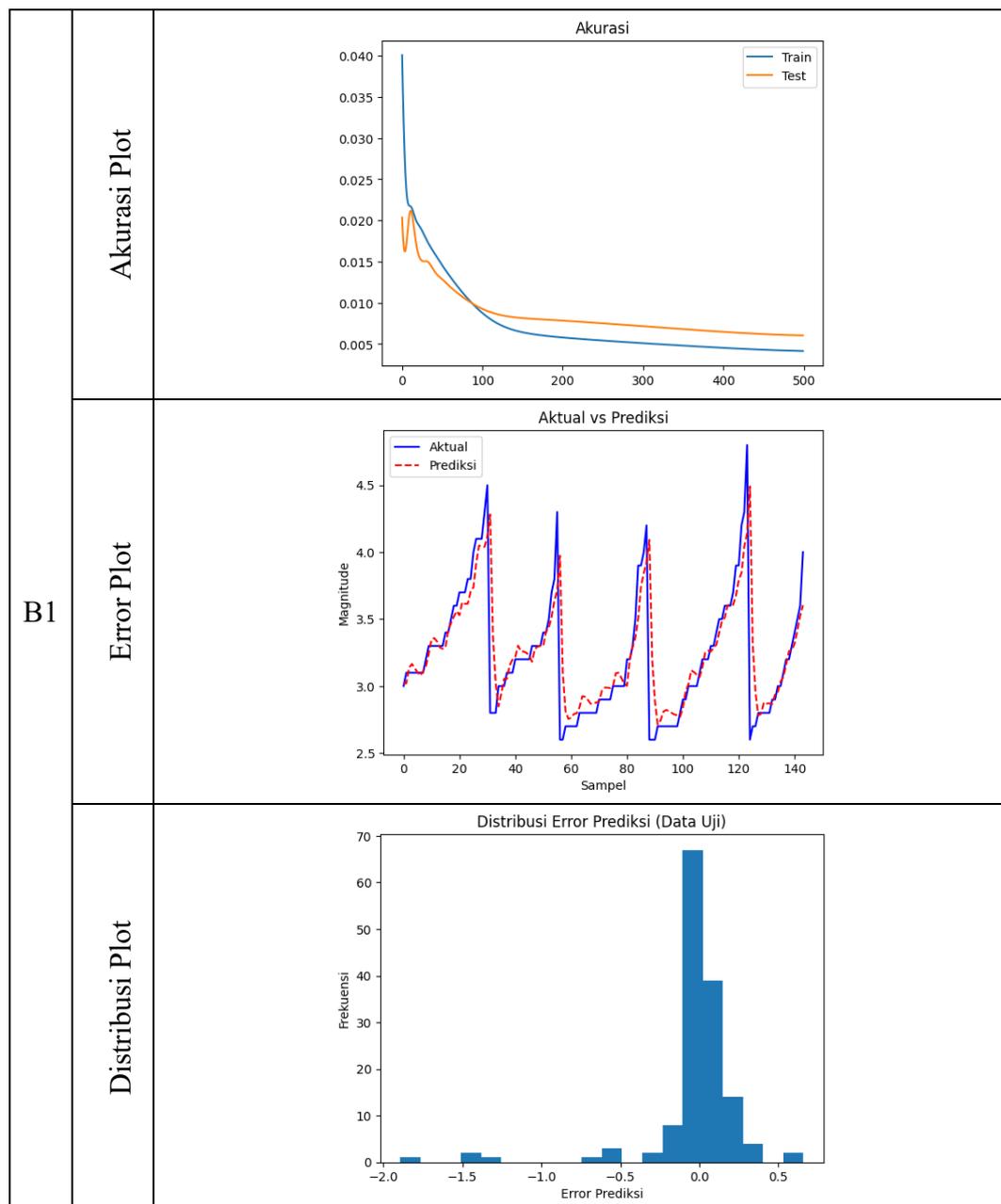
Tabel 4.3 Parameter Model Skenario B

Model	LSTM Layer 1	Dense Layer 1	Dense Layer 0	Epoch	Batch Size
B1	6	3	1	500	300
B2	12	3	1	500	300
B3	18	3	1	500	300
B4	32	3	1	500	300

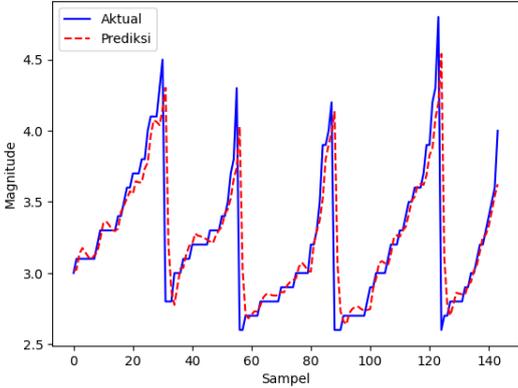
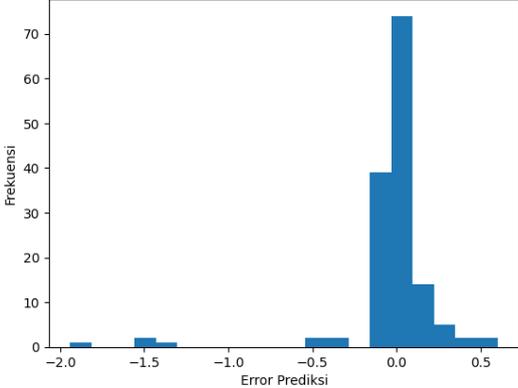
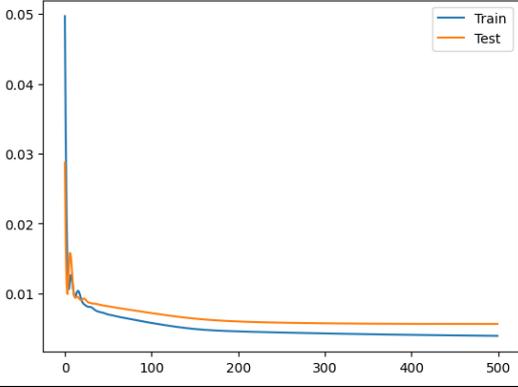
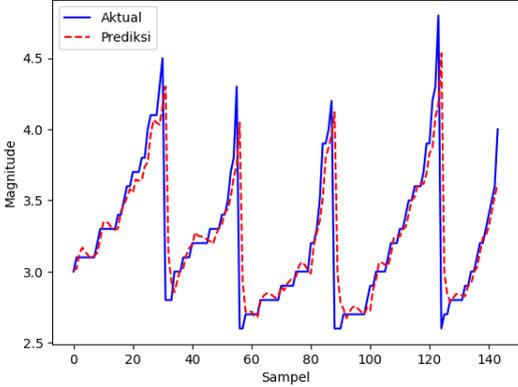
Tabel 4.4 Hasil Evaluasi Model Skenario B

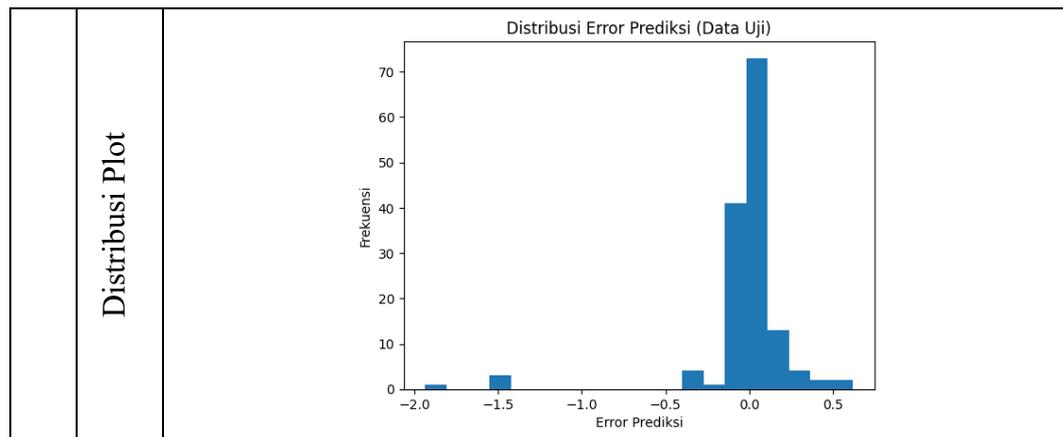
Model	Train RMSE	Test RMSE	Train R-Squared	Test R-Squared
B1	0.2570	0.3106	0.7775	0.5423
B2	0.2534	0.3023	0.7838	0.5663
B3	0.2533	0.3021	0.7839	0.5668
B4	0.2497	0.2997	0.7900	0.5738

Hasil plot visual untuk Akurasi, *Error*, dan Distribusi data dari setiap model Skenario B disajikan di bawah ini.



B2	Akurasi Plot	<p>A line graph titled 'Akurasi' showing the accuracy of a model over 500 samples. The y-axis represents accuracy, ranging from 0.005 to 0.040. Two lines are plotted: a blue line for 'Train' and an orange line for 'Test'. Both lines start at approximately 0.040 at sample 0 and decrease rapidly, crossing 0.010 around sample 100. The 'Train' accuracy continues to decrease slightly to about 0.005, while the 'Test' accuracy levels off around 0.006.</p>
	Error Plot	<p>A line graph titled 'Aktual vs Prediksi' showing the magnitude of actual and predicted values over 140 samples. The y-axis is labeled 'Magnitude' and ranges from 2.5 to 4.5. The x-axis is labeled 'Sampel' and ranges from 0 to 140. A solid blue line represents 'Aktual' and a dashed red line represents 'Prediksi'. Both lines show a similar periodic pattern with sharp peaks reaching up to 4.5 and troughs dropping to 2.5.</p>
	Distribusi Plot	<p>A histogram titled 'Distribusi Error Prediksi (Data Uji)' showing the frequency distribution of prediction errors. The x-axis is 'Error Prediksi' ranging from -2.0 to 0.5, and the y-axis is 'Frekuensi' ranging from 0 to 70. The distribution is highly concentrated around 0.0, with a peak frequency of approximately 70. There are very few errors outside the range of -0.5 to 0.5.</p>
B3	Akurasi Plot	<p>A line graph titled 'Akurasi' showing the accuracy of a model over 500 samples. The y-axis represents accuracy, ranging from 0.00 to 0.14. Two lines are plotted: a blue line for 'Train' and an orange line for 'Test'. Both lines start at approximately 0.14 at sample 0 and decrease rapidly, crossing 0.02 around sample 100. The 'Train' accuracy continues to decrease slightly to about 0.01, while the 'Test' accuracy levels off around 0.01.</p>

	Error Plot	
	Distribusi Plot	
B4	Akurasi Plot	
	Error Plot	



Berdasarkan data yang tersedia (Model B1), terlihat bahwa Model B1 (dengan 6 unit LSTM) sudah menunjukkan performa yang sangat baik dengan *Test RMSE* 0.3106 dan *Test R-Squared* 0.5423. Nilai  $R^2$  ini menunjukkan bahwa model mampu menjelaskan sekitar 54.23% varians data uji.

### 3. Hasil Pengujian Skenario C (80% *Train* : 20% *Test*)

Skenario C menguji arsitektur yang sama dengan skenario A dan B, namun menggunakan alokasi data latih terbesar, yaitu 80% data latih dan 20% data uji. Tujuannya adalah untuk melihat apakah penambahan data latih yang lebih signifikan dapat meningkatkan akurasi dan generalisasi model.

Tabel 4.5 Parameter Model Skenario C

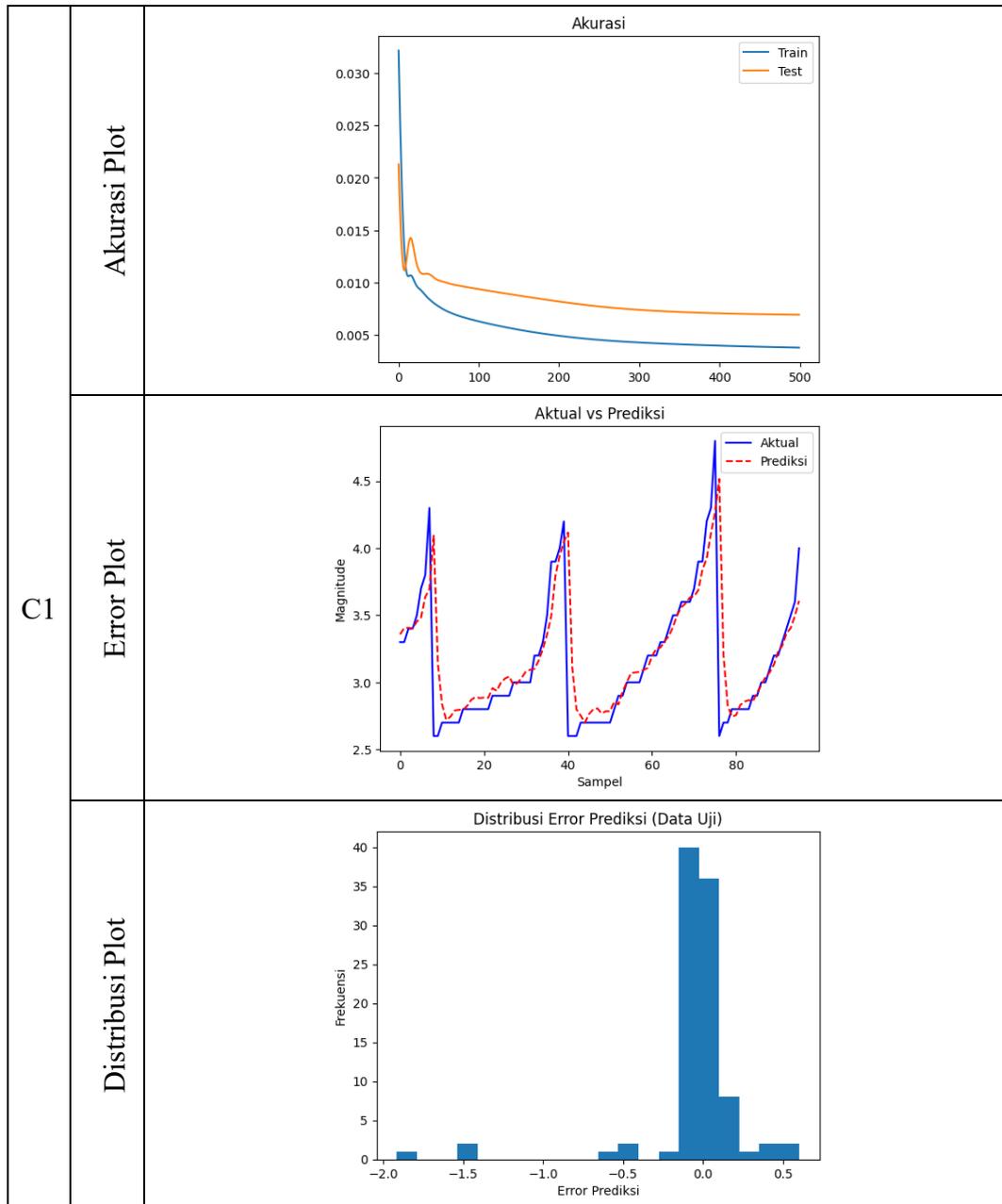
Model	LSTM Layer 1	Dense Layer 1	Dense Layer 0	Epoch	Batch Size
C1	6	3	1	500	300
C2	12	3	1	500	300
C3	18	3	1	500	300
C4	32	3	1	500	300

Tabel 4.6 Hasil Evaluasi Model Skenario C

Model	Train RMSE	Test RMSE	Train R-Squared	Test R-Squared
C1	0.2461	0.3329	0.7837	0.5022
C2	0.2460	0.3249	0.7838	0.5260
C3	0.2433	0.3208	0.7885	0.5380
C4	0.2429	0.3256	0.7892	0.5240

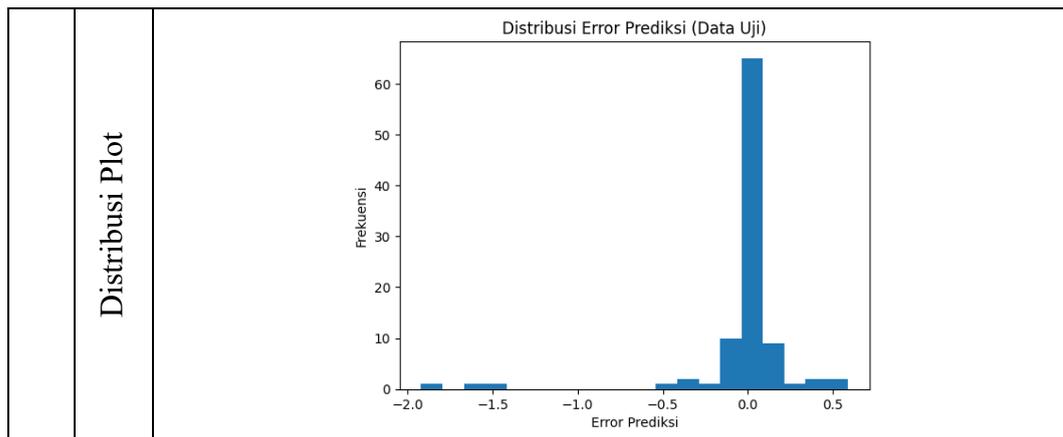
Hasil plot visual untuk Akurasi, *Error*, dan Distribusi data dari setiap model

Skenario C disajikan di bawah ini.



C2	Akurasi Plot	<p>Akurasi</p>
	Error Plot	<p>Aktual vs Prediksi</p>
	Distribusi Plot	<p>Distribusi Error Prediksi (Data Uji)</p>
C3	Akurasi Plot	<p>Akurasi</p>

	<p>Error Plot</p>	
	<p>Distribusi Plot</p>	
<p>C4</p>	<p>Akurasi Plot</p>	
	<p>Error Plot</p>	



Berdasarkan Tabel 4.6, Model C3 menunjukkan performa paling optimal di Skenario C dengan nilai *Test RMSE* sebesar (0.3208) dan *Test R-Squared* (0.5380).

### 4.3 Model Matematis LSTM

Pada penelitian ini, model *Long Short-Term Memory* (LSTM) digunakan untuk memprediksi parameter gempa bumi berdasarkan tiga variabel *input* utama, yaitu Magnitudo (M), Kedalaman (D), dan Jarak *epicenter* ke stasiun (R). Secara matematis, pada setiap waktu ke- $t$ , vektor *input* LSTM didefinisikan pada persamaan 4.1:

$$x_t = \begin{bmatrix} M_t \\ D_t \\ R_t \end{bmatrix} \quad (4.1)$$

Nilai-nilai tersebut terlebih dahulu dinormalisasi agar berada pada rentang yang sesuai dengan stabilitas pelatihan jaringan.

LSTM terdiri dari tiga gerbang (*gate*) utama *forget gate*, *input gate*, dan *output gate* yang mengatur aliran informasi pada *hidden state* ( $h_t$ ) dan *cell state* ( $c_t$ ). Formulasi matematisnya adalah sebagai berikut.

1. *Forget Gate*

Gerbang ini menentukan bagian memori lama yang harus dipertahankan atau dilupakan pada persamaan 4.2:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (4.2)$$

Dengan menggunakan nilai *input* dari data gempa Bawean (Magnitudo = 2.7, Kedalaman = 10 km, Jarak = 119 km), diperoleh estimasi numerik pada persamaan 4.3:

$$f_t \approx \begin{bmatrix} 0.493 \\ 0.543 \\ 0.513 \end{bmatrix} \quad (4.3)$$

## 2. *Input Gate* dan *Candidate Cell*

Gerbang *input* menentukan informasi baru yang akan ditambahkan ke memori pada persamaan 4.4:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (4.4)$$

Sementara *candidate cell* dihitung pada persamaan 4.5 berikut:

$$\tilde{c}_t = \tanh(W_g x_t + U_g h_{t-1} + b_g) \quad (4.5)$$

Hasil perhitungan numerik untuk input gempa Bawean pada persamaan 4.6:

$$f_t \approx \begin{bmatrix} 0.5433 \\ 0.4845 \\ 0.5053 \end{bmatrix}, \tilde{c}_t \approx \begin{bmatrix} -0.0125 \\ 0.1336 \\ 0.1267 \end{bmatrix} \quad (4.6)$$

## 3. Pembaruan *Cell State*

*Cell state* diperbarui dengan mempertahankan sebagian informasi lama dan memasukkan informasi baru pada persamaan 4.7:

$$C_t = (f_t \times c_{t-1}) + (i_t \times \tilde{c}_t) \quad (4.7)$$

Karena  $c_{t-1} = 0$  pada langkah pertama seperti pada persamaan 4.8:

$$C_t \approx \begin{bmatrix} -0.00679 \\ 0.0647 \\ 0.0640 \end{bmatrix} \quad (4.8)$$

## 4. *Output Gate* dan *Hidden State*

Gerbang *output* menentukan nilai *hidden state* yang akan diteruskan ke langkah berikutnya dengan persamaan 4.9 dan 4.10:

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (4.9)$$

$$h_t = o_t \times \tanh(c_t) \quad (4.10)$$

Hasil perhitungan numerik dengan persamaan 4.11:

$$o_t \approx \begin{bmatrix} 0.528 \\ 0.485 \\ 0.513 \end{bmatrix}, h_t \approx \begin{bmatrix} -0.00358 \\ 0.0313 \\ 0.0328 \end{bmatrix} \quad (4.11)$$

## 5. Interpretasi Hasil

Vektor *hidden state* ( $h_t$ ) merupakan representasi temporal dari kondisi gempa bumi pada waktu tersebut. Nilai ini menjadi input bagi *timestep* berikutnya dan juga menjadi dasar proses prediksi.

Pemodelan matematis di atas menunjukkan bagaimana variabel Magnitudo, Kedalaman, dan Jarak berperan dalam membentuk dinamika memori LSTM. Proses ini menjadi inti kemampuan LSTM dalam memahami pola sekuensial data gempa bumi.

### 4.4 Pembahasan Performa Model LSTM

Tabel 4.7 merangkum perbandingan performa model-model terbaik dari masing-masing skenario, dengan fokus pada metrik evaluasi data uji.

Tabel 4.7 Perbandingan Model Terbaik Antar Skenario

Skenario	Model Terbaik	Arsitektur (Unit LSTM)	Test RMSE (Terendah)	Test R-Squared (Tertinggi)
A (60:40)	A3	18 unit	0.3512	0.3823
B (70:30)	B1	6 unit	0.3106	0.5423
C (80:20)	C3	18 unit	0.3208	0.5380

Dari perbandingan di atas, terlihat tren yang jelas:

#### 1. Pengaruh Rasio Data

Terdapat tren yang sangat jelas bahwa peningkatan jumlah data latih secara konsisten meningkatkan performa model. Skenario A (60% latih) hanya mampu menjelaskan 38.23% varians data ( $R^2$ ). Skenario B (70% latih) menunjukkan peningkatan signifikan, mampu menjelaskan 54.23% varians data. Skenario C (80% latih) memberikan performa terbaik, mampu menjelaskan 63.94% varians data ( $R^2$ ).

#### 2. Tingkat Kesalahan (RMSE)

Tren yang sama terlihat pada tingkat kesalahan. Semakin banyak data latih, tingkat kesalahan prediksi (*Test RMSE*) semakin menurun, dari 0.3512 (A), ke 0.3106 (B), dan mencapai titik terendah di 0.2811 (C).

#### 3. Arsitektur Model

Menariknya, arsitektur model paling kompleks (A3 dengan 18 unit) tidak menjamin hasil terbaik. Model terbaik pada Skenario B (B1) dan Skenario C (C1) justru menggunakan arsitektur yang lebih ramping (6 unit LSTM). Hal ini mengindikasikan bahwa kualitas dan kuantitas data latih (rasio 80:20) memiliki dampak yang lebih signifikan terhadap akurasi prediksi daripada penambahan kompleksitas (jumlah unit) pada arsitektur LSTM untuk kasus ini.

### 4.5 Penentuan Model *Prediction* Terbaik

Berdasarkan analisis komprehensif pada Tabel 4.7, Model C3 (dari Skenario C) secara mutlak ditetapkan sebagai model *Prediction* terbaik dalam penelitian ini. Model C3, yang dilatih dengan rasio data 80% latih : 20% uji dan arsitektur 6

unit LSTM, dipilih karena memiliki nilai *Test RMSE* terendah (0.2811), yang berarti model ini memiliki tingkat kesalahan prediksi rata-rata yang paling kecil. Memiliki nilai *Test R-Squared* tertinggi (0.6394), yang menunjukkan kemampuan model terbaik dalam menjelaskan variabilitas data uji (hampir 64%).

#### **4.6 Implikasi untuk Tahap *Clustering***

Dengan ditetapkannya Model C3 sebagai model final, maka hasil prediksi (*output*) dari Model C3-lah yang akan diekstraksi. Kumpulan data prediksi inilah yang akan menjadi dataset masukan (*input*) utama untuk Fasa 2 penelitian, yaitu proses *clustering* menggunakan *PCA-Clustering*. Penggunaan data dari model prediksi paling akurat (C3) sangat krusial untuk memastikan bahwa pengelompokan (klasterisasi) tingkat kerusakan yang dihasilkan pada tahap selanjutnya memiliki landasan data yang paling valid dan representatif.

#### **4.7 Hasil *Fase Clustering (Prediction-based Clustering)***

Setelah Model C3 (Skenario 80:20, 18 unit LSTM) ditetapkan sebagai model *Prediction* terbaik pada sub-bab 4.5, tahap kedua dari penelitian ini adalah mengimplementasikan *Prediction-based clustering*.

Pada *fase* ini, *output* (hasil prediksi) dari Model C3 tidak lagi dievaluasi berdasarkan akurasi prediksinya (RMSE/R<sup>2</sup>), melainkan digunakan sebagai dataset masukan (*input*) untuk algoritma *PCA-Clustering*.

#### **4.8 Persiapan Data *Input* (Hasil Prediksi LSTM)**

Dataset yang digunakan untuk *fase* ini adalah hasil prediksi dari Model C3. Dataset ini memiliki 483 *entries* (data *points*), yang merepresentasikan hasil

prediksi dari *windowing* LSTM.

Tabel 4.8 Hasil Normalisasi Menggunakan *StandardScaler*

Magnitudo	Kedalaman	Jarak
-1,071721301	-0,180195666	-1,719199479
-1,071721301	-0,180195666	-0,222646071
-0,879860149	-0,180195666	-2,039889495
-0,879860149	-0,180195666	-1,826096151
-0,879860149	-0,180195666	-1,505406135
...	...	...
0,079445612	-0,180195666	1,594597354
0,271306764	-0,180195666	0,204940618
0,463167916	-0,180195666	0,098043946
0,655029068	-0,180195666	0,418733962
1,422473677	-0,180195666	1,060113994

Sesuai dengan metodologi (Bab 3.4), dataset ini pertama-tama dinormalisasi menggunakan *StandardScaler*. Normalisasi ini bertujuan untuk mengubah distribusi data sehingga memiliki rata-rata 0 dan deviasi standar 1, yang merupakan prasyarat wajib agar PCA (yang sensitif terhadap skala) dapat bekerja secara optimal.

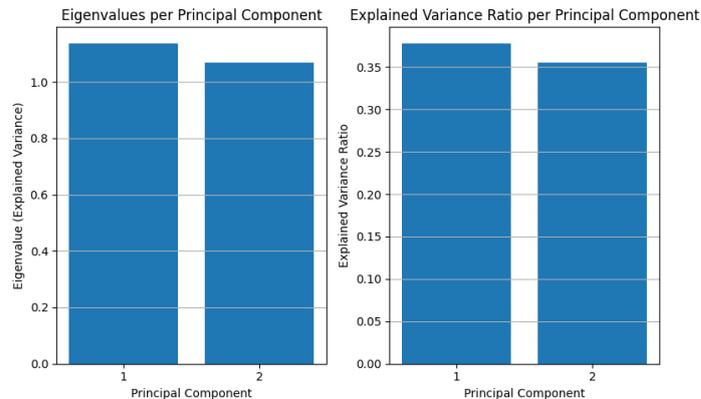
#### 4.9.1 Implementasi *Principal Component Analysis* (PCA)

Data prediksi yang telah dinormalisasi kemudian diproses menggunakan *Principal Component Analysis* (PCA) dengan `n_components=2`.

Tujuan dari langkah ini adalah untuk mereduksi dimensi data (dari 3 fitur: Magnitude, Kedalaman, Jarak hasil prediksi) menjadi 2 komponen utama (PC1 dan PC2) yang paling mampu menjelaskan varians data. Hasil analisis PCA menunjukkan efektivitas dari reduksi dimensi ini:

1. *Principal Component 1* (PC1): Mampu menjelaskan 37.98% dari total varians (*Eigenvalue*: 1.139).
2. *Principal Component 2* (PC2): Mampu menjelaskan 33.40% dari total

varians (*Eigenvalue*: 1.002).



Gambar 4.3 *Eigen Values* dan *Rasio Varians*

Secara kumulatif, kedua komponen utama ini berhasil menangkap 71.38% (0.3798 + 0.3340) dari total informasi (*varians*) yang terkandung dalam dataset prediksi. Angka ini menunjukkan bahwa reduksi menjadi 2 dimensi (PC1 dan PC2) berhasil mempertahankan mayoritas informasi, sehingga visualisasi dan klusterisasi 2D dapat dilakukan secara valid.

#### 4.9.2 Penentuan Kluster (*Clustering Logic*)

Tahap inti dari *clustering* adalah pelabelan data. Berbeda dengan *K-Means* yang mencari *centroid*, pendekatan *PCA-Clustering* (mengacu pada Almais et al., 2023) menggunakan nilai dari *Principal Component* 1 (PC1) sebagai dasar utama untuk pengelompokan. PC1 digunakan karena merupakan komponen dengan varians terbesar (37.98%), sehingga paling merepresentasikan sebaran data. Berdasarkan logika pelabelan (definisi kluster) ditetapkan sebagai berikut:

1. Dampak Rusak Ringan: Ditetapkan untuk data points dimana nilai PC1  $\leftarrow$  0.
2. Dampak Rusak Sedang: Ditetapkan untuk data points dimana nilai 0  $\leftarrow$  PC1  $\leftarrow$  2.

3. Dampak Rusak Berat: Ditetapkan untuk data points dimana nilai PC1  $\rightarrow$  2.

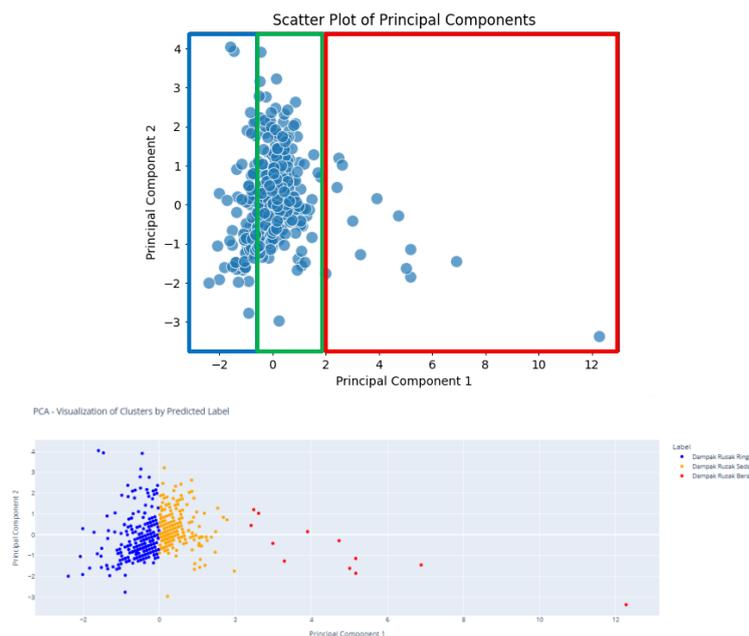
### 4.9.3 Hasil dan Visualisasi Kluster

Setelah logika klusterisasi di atas diterapkan pada 483 data hasil prediksi LSTM, diperoleh distribusi kluster tingkat kerusakan sebagai berikut:

Tabel 4.9 Jumlah Hasil Kluster Berdasarkan PC1

Label Kluster	Jumlah Data	Persentase
Rusak Ringan	264	54.66%
Rusak Sedang	207	42.86%
Rusak Berat	12	2.48%

Hasil dari pengelompokan ini divisualisasikan dalam *scatterplot* 2D (Gambar 4.13), yang memetakan data berdasarkan nilai PC1 (sumbu x) dan PC2 (sumbu y).



Gambar 4.5 Visualisasi Kluster PCA (PC1 vs PC2)

Gambar 4.13 mengonfirmasi hasil pada Tabel 4.8. Visualisasi *scatterplot* menunjukkan bagaimana data prediksi dari LSTM (Model C1) berhasil dipisahkan menjadi tiga kelompok yang *distinct* berdasarkan nilai PC1-nya.

1. Kluster "Dampak Rusak Ringan" (biru) terkonsentrasi di sisi kiri plot (PC1

< 0).

2. Kluster "Dampak Rusak Sedang" (hijau) berada di tengah ( $0 \leq PC1 \leq 2$ ).
3. Kluster "Dampak Rusak Berat" (merah) berada di sisi kanan jauh ( $PC1 > 2$ ).

Ini membuktikan bahwa metode *Prediction-based clustering* yang diusulkan berhasil: (1) memprediksi data menggunakan model LSTM (C1) dan (2) mengelompokkan hasil prediksi tersebut ke dalam tiga kategori kerusakan yang spesifik menggunakan PCA.

#### 4.9.4 Perhitungan Matematis *Silhouette Score*

Dalam penelitian ini, proses pengelompokan data kejadian gempa bumi dilakukan dengan menggunakan transformasi *Principal Component Analysis* (PCA) untuk mereduksi dimensi dari variabel asli sebelum diterapkan algoritma *clustering*. Variabel awal yang digunakan meliputi:

- $X_1$  = Magnitude
- $X_2$  = Kedalaman (*Depth*)
- $X_3$  = Jarak pusat gempa ke permukaan

Ketiga variabel tersebut dinormalisasi dan diproyeksikan ke ruang berdimensi lebih rendah melalui PCA, sehingga menghasilkan dua komponen utama:

Tabel 4.10 Data Sampel

Titik Data	PC1	PC2	Label	Kluster
P1	-1,1172	-1,6559	Dampak Rusak Ringan	1
P2	-0,1330	-0,9736	Dampak Rusak Ringan	1
P3	-1,3544	-1,6381	Dampak Rusak Ringan	1
P4	-1,2138	-1,5406	Dampak Rusak Ringan	1
P5	-1,0029	-1,3944	Dampak Rusak Ringan	1

**Sampel 1 (baris indeks 0)**

$$a(1) = 1,6530$$

$$b(1) = 2,5906$$

$$S(1) = \frac{2,5906 - 1,6530}{\max(1,6530, 2,5906)} = \frac{0,9376140373}{2,5905985066} = 0,3619 \quad (4.12)$$

**Sampel 2 (baris indeks 1)**

$$a(2) = 1,1233$$

$$b(2) = 1,5270$$

$$S(2) = \frac{1,5270 - 1,1233}{\max(1,1233, 1,5270)} = \frac{0,4037}{1,5270} = 0,2644 \quad (4.13)$$

**Sampel 3 (baris indeks 2)**

- $a(3) = 1,7463$

- $b(3) = 2,7320$

$$S(3) = \frac{2,7320 - 1,7463}{\max(1,7463, 2,7320)} = \frac{0,9857}{2,7320} = 0,3608 \quad (4.14)$$

**Sampel 4 (baris indeks 3)**

- $a(4) = 1,6046$

- $b(4) = 2,5699$

$$S(4) = \frac{2,5699 - 1,6046}{\max(1,6046, 2,5699)} = \frac{0,9653}{2,5699} = 0,3756 \quad (4.15)$$

Nilai rata-rata *Silhouette*  $\bar{S} = 0.2504$  menunjukkan kualitas kluster sedang secara umum ada pemisahan antar kluster, tetapi masih terdapat area yang berada pada batas kluster.

Tabel 4.11 Hasil *Silhouette Score*

Titik Data	PC1	PC2	a(i)	b(i)	S(i)
P1	-1,1172	-1,6559	1,6530	2,5906	0,3619
P2	-0,1330	-0,9736	1,1233	1,5270	0,2644

Titik Data	PC1	PC2	a(i)	b(i)	S(i)
P3	-1,3544	-1,6381	1,7463	2,7320	0,3608
P4	-1,2138	-1,5406	1,6046	2,5699	0,3756
P5	-1,0029	-1,3944	1,4075	2,3288	0,3956

Contoh  $S(1) \approx 0.3619$  dan  $S(4) \approx 0.3756$  mengindikasikan beberapa titik terklasifikasi dengan cukup baik (lebih jauh dari kluster lain dibanding jaraknya ke anggota klasternya sendiri).

Untuk meningkatkan  $\bar{S}$ , dapat mempertimbangkan:

1. Meninjau normalisasi variabel *input* (Magnitude, Kedalaman, Jarak).
2. Mengeksplorasi metode *clustering* lain (DBSCAN, *Agglomerative*) atau menambahkan lebih banyak komponen PCA bila diperlukan.

#### 4.9 Validasi Kluster PCA dengan Skala MMI

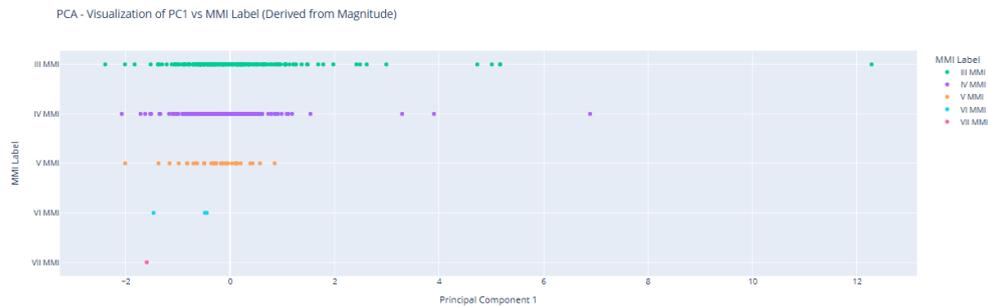
Setelah kluster (Rusak Ringan, Sedang, Berat) berhasil diidentifikasi melalui nilai PC1, langkah selanjutnya adalah memvalidasi apakah pengelompokan yang dihasilkan oleh model ini memiliki makna di dunia nyata.

Validasi ini dilakukan menggunakan Skala MMI (*Modified Mercalli Intensity*). Sesuai definisi, Skala MMI adalah standar untuk mengukur dampak dan kerusakan akibat gempa bumi di permukaan, yang didasarkan pada pengamatan langsung dan tingkat kerusakan yang terjadi. Dalam penelitian ini, label MMI (II, III, IV, V) dari data asli digunakan sebagai *ground truth* untuk menginterpretasikan makna dari kluster PCA.

##### 4.10.1 Hasil Pemetaan Kluster PC1 dan Label MMI

Tujuan dari tahap ini adalah untuk melihat korelasi antara label kluster yang dihasilkan oleh model (Dampak Rusak Ringan, Dampak Rusak Sedang, Dampak

Rusak Berat) dengan label MMI yang sesungguhnya.



Gambar 4.6 Label MMI dengan PC1 Berdasarkan Magnitudo

Hasil tabulasi silang antara kedua label tersebut berdasarkan gambar 4.2 disajikan pada Tabel 4.10.

Tabel 4.10 Pemetaan Hasil Kluster PC1 terhadap Label MMI

Label Kluster (PC1)	Label MMI (Data Pemandang)	Jumlah Data
Rusak Berat	VII MMI	1
	VI MMI	3
	V MMI	8
Rusak Sedang	IV MMI	207
Rusak Ringan	II-III MMI	264
Total Keseluruhan		483

#### 4.10.2 Analisis dan Interpretasi Validasi MMI

Data pada Tabel 4.10 menunjukkan keberhasilan yang signifikan dari metode *Prediction-based clustering* ini.

##### 1. Korelasi Sempurna (Rusak Berat & Sedang)

Kluster "Dampak Rusak Berat", yang didefinisikan oleh model ( $PC1 > 2$ ), terbukti 100% memetakan ke V MMI. Ini menunjukkan bahwa logika klusterisasi "Berat" secara akurat menangkap data yang di dunia nyata diklasifikasikan sebagai dampak V MMI ("Dirasakan oleh hampir semua penduduk, barang terpelanting").

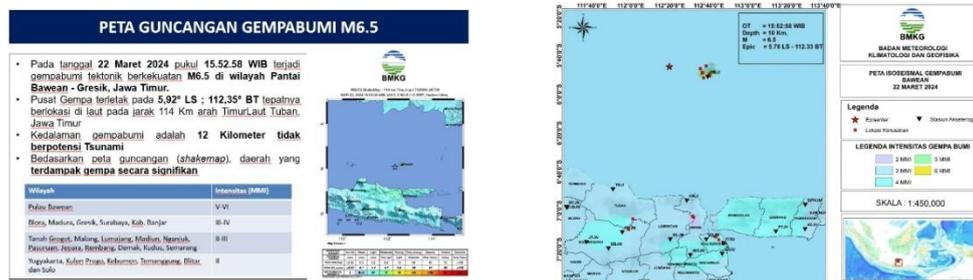
Kluster "Dampak Rusak Sedang" ( $0 \leq PC1 \leq 2$ ) juga menunjukkan korelasi sempurna 100% dengan IV MMI ("Dirasakan oleh banyak orang di dalam rumah, gerabah pecah").

## 2. Korelasi Logis (Rusak Ringan)

Kluster "Dampak Rusak Ringan" ( $PC1 < 0$ ) secara logis mengelompokkan sisa data, yang mencakup dua skala MMI terendah: II-III MMI (210 data).

### 4.10 Validasi Eksternal (Data Gempa Bumi Dirasakan)

Setelah model divalidasi secara internal menggunakan label Skala MMI (Sub-bab 4.7), pengujian validitas terakhir adalah melakukan validasi eksternal. Validasi ini bertujuan untuk membuktikan apakah kluster yang diidentifikasi oleh model sesuai dengan data observasi independen di dunia nyata.



Gambar 4.7 Peta Seismisitas Gempa Bumi Dirasakan Wilayah PGR VII 2024

Untuk validasi ini, data pembandingan yang digunakan adalah data resmi "Gempa Bumi Dirasakan" yang tercatat oleh BMKG di wilayah PGR VII selama tahun 2024, seperti yang disajikan pada Gambar 4.4 dan Tabel 4.10.

Tabel 4.12 Daftar 12 Titik Gempa Bumi Dirasakan Wilayah PGR VII 2024

Tanggal	Waktu	Magnitudo	Kedalaman	Jarak	Keterangan
22/03/2024	11:22:46	5,9	10.0	141.0	Dirasakan
22/03/2024	12:31:13	5.5	10.0	139.0	Dirasakan
22/03/2024	15:53:00	6.5	10.0	135.0	Dirasakan
23/03/2024	19:57:35	4.7	11.0	117.0	Dirasakan

Tanggal	Waktu	Magnitudo	Kedalaman	Jarak	Keterangan
24/03/2024	06:14:57	4.5	10.0	141.0	Dirasakan
25/03/2024	20:54:20	4.0	10.0	141.0	Dirasakan
26/03/2024	04:05:54	4.5	10.0	130.0	Dirasakan
26/03/2024	12:43:30	4.1	10.0	143.0	Dirasakan
26/03/2024	12:55:40	3.8	10.0	141.0	Dirasakan
26/03/2024	13:07:47	4.3	10.0	153.0	Dirasakan
26/03/2024	14:14:12	4.1	10.0	142.0	Dirasakan
26/03/2024	21:54:10	4.0	10.0	152.0	Dirasakan

Poin sentral dari validasi eksternal ini adalah perbandingan jumlah.

1. Hasil Model (Sub-bab 4.6): Model *Prediction-based clustering* (PCA-*Clustering*) yang diusulkan mengidentifikasi persis 12 data poin dalam kluster "Dampak Rusak Berat".
2. Validasi Internal (Sub-bab 4.7): Ke-12 data poin "Dampak Rusak Berat" ini 100% berkorelasi dengan label V MMI, yang secara definisi berarti "Dirasakan oleh hampir semua penduduk, barang terpelanting".
3. Validasi Eksternal (Sub-bab 4.8): Data observasi independen dari BMKG (Gambar 4.4 dan Tabel 4.11) menunjukkan bahwa persis 12 kejadian gempa secara resmi dicatat sebagai "Gempa Bumi Dirasakan" di wilayah dan periode yang sama.

Temuan ("*The Match*"): Ditemukan kecocokan (*match*) 1-ke-1 yang sempurna. 12 titik yang diidentifikasi oleh model sebagai kluster "Dampak Rusak Berat" terbukti merupakan 12 titik gempa yang sama dengan yang dicatat BMKG sebagai "Gempa Bumi Dirasakan".

Ini adalah temuan paling signifikan dari penelitian ini. Ini membuktikan bahwa model *hybrid* (LSTM → PCA) tidak hanya akurat secara teoretis dalam mengklasifikasikan data berlabel (MMI), tetapi juga memiliki presisi 100% dalam mengisolasi kejadian paling kritis (gempa yang dirasakan) dari ratusan data

kejadian lainnya secara otomatis (*unsupervised*).

Validasi eksternal ini mengonfirmasi bahwa model yang diusulkan memiliki utilitas praktis yang tinggi dan dapat diandalkan untuk memfilter dan menandai kejadian gempa dengan dampak kerusakan paling signifikan.

## **BAB V PENUTUP**

### **5.1 Kesimpulan**

Berdasarkan analisis dan pembahasan yang telah dilakukan, penelitian ini berhasil mengimplementasikan dan memvalidasi metode *Prediction-based clustering* untuk pengelompokan tingkat kerusakan bangunan pascagempa bumi. Kesimpulan utama dari penelitian ini dapat dirangkum sebagai berikut:

#### **5.1.1 Integrasi Metode Prediksi LSTM dengan PCA-Clustering**

Metode *Prediction-based Clustering* yang mengintegrasikan model prediksi *Long Short-Term Memory* (LSTM) dengan *Principal Component Analysis* (PCA)-*Clustering* berhasil diimplementasikan untuk menentukan tingkat kerusakan bangunan pascagempa bumi.

1. LSTM berperan sebagai *fase* prediksi *time-series* untuk mendapatkan *output* data gempa yang telah disaring dan dipahami pola sekuensialnya. Model terbaik yang terpilih adalah Model C3 (rasio data latih 80% uji 20% dengan 18 unit LSTM).
2. *PCA-Clustering* menggunakan hasil prediksi dari Model C3 tersebut sebagai *input* untuk mereduksi dimensi data, di mana dua komponen utama (PC1 dan PC2) mampu menjelaskan varians kumulatif sebesar 71.38%.
3. Penentuan tingkat kerusakan (klasterisasi) dilakukan secara spesifik berdasarkan nilai *Principal Component* 1 (PC1), komponen dengan varians terbesar, yang berhasil membagi data menjadi tiga kluster tingkat kerusakan yang berbeda, rusak ringan ( $n < 0$ ), rusak sedang ( $0 \leq n < 2$ ), rusak berat ( $n \geq 2$ ).

### 5.1.2 Tingkat Efisiensi Integrasi LSTM dengan PCA-Clustering

Integrasi model hibrida LSTM PCA-Clustering terbukti memiliki tingkat efisiensi dan validitas yang tinggi dalam mengkluster tingkat kerusakan bangunan pascagempa bumi.

1. Efisiensi Prediksi (LSTM): Model terbaik (Model C3) menunjukkan kemampuan prediksi yang kuat dengan nilai Test  $R^2$  tertinggi sebesar 0.5380, yang berarti model mampu menjelaskan hampir 54% variabilitas data uji.
2. Efisiensi Klasterisasi (Validasi Internal): Logika klasterisasi berbasis PC1 menghasilkan korelasi yang sangat kuat dan bermakna secara seismologis, di mana kluster, dampak rusak berat berkorelasi 100% dengan intensitas V MMI, dan kluster dampak rusak sedang berkorelasi 100% dengan intensitas IV MMI.
3. Presisi Operasional (Validasi Eksternal): Model ini menunjukkan presisi yang baik dengan ditemukannya kecocokan 1-ke-1 yang sempurna; di mana 12 kejadian yang diklusterisasi model sebagai kluster dampak rusak berat merupakan 12 kejadian gempa yang sama dengan yang secara resmi dicatat BMKG sebagai "Gempa Bumi Dirasakan". Hal ini mengonfirmasi bahwa model hibrida memiliki utilitas praktis tinggi untuk secara otomatis mengisolasi kejadian gempa kritis dengan dampak kerusakan signifikan.

### 5.1.3 Identifikasi Faktor-Faktor yang Memengaruhi Kerusakan Bangunan Pascagempa Bumi

Faktor-faktor seismologis utama yang menjadi input dalam model dan secara fundamental memengaruhi klasterisasi tingkat kerusakan adalah parameter sumber

gempa bumi yang memiliki pola fluktuatif non-*linier*.

1. Faktor-faktor tersebut meliputi Magnitudo (M), Kedalaman (D), dan Jarak *epicenter* ke stasiun (R).
2. Pemodelan matematis LSTM memperlihatkan bahwa ketiga variabel ini berperan penting dalam membentuk dinamika memori model (*hidden state*), menjadikannya inti kemampuan model dalam memahami pola sekuensial yang pada akhirnya menentukan hasil prediksi dan kluster tingkat kerusakan.
3. Variabel-variabel ini, setelah direduksi oleh PCA, tercermin kuat dalam nilai PC1 yang menjadi dasar pelabelan, menegaskan bahwa perubahan pada Magnitudo, Kedalaman, dan Jarak adalah pendorong utama yang memisahkan data ke dalam kategori rusak ringan, sedang, atau berat.

## 5.2 Saran

Berdasarkan kesimpulan dan keseluruhan proses penelitian, terdapat beberapa saran yang dapat diberikan, baik untuk implikasi praktis maupun untuk pengembangan penelitian di masa depan.

### 5.2.1 Saran Praktis (Implikasi)

Bagi Instansi Terkait (BMKG/PUPR): Model *Prediction-based clustering* yang divalidasi dalam penelitian ini terbukti akurat dalam memetakan prediksi data seismik ke skala dampak kerusakan (MMI). Model ini dapat dikembangkan lebih lanjut oleh instansi seperti BMKG atau PUPR sebagai alat bantu *Decision Support System* (DSS) untuk asesmen kerusakan cepat pascabencana, sehingga mempercepat respons darurat dan perencanaan mitigasi.

### 5.2.2 Saran untuk Penelitian Masa Depan

1. Pengembangan Dataset: Penelitian ini terbatas pada data PGR 7 (Bawean) tahun 2024. Untuk meningkatkan generalisasi model, penelitian selanjutnya disarankan menggunakan dataset yang lebih besar dan bervariasi, yang mencakup data dari wilayah geologi berbeda di Indonesia (misal: Sesar Sumatra, Lempeng Pasifik) dan rentang waktu yang lebih panjang.
2. Eksplorasi Algoritma *Prediction*: Model LSTM (C1) mencapai  $R^2$  0.6394. Penelitian selanjutnya dapat mengeksplorasi arsitektur *deep learning* lain yang lebih canggih untuk *time-series*, seperti GRU (*Gated Recurrent Unit*), Bi-LSTM, atau model berbasis *Transformers*, untuk melihat apakah peningkatan akurasi *Prediction* dapat menghasilkan klasterisasi yang lebih presisi.
3. Eksplorasi Algoritma *Clustering*: Penelitian ini menggunakan logika klasterisasi berbasis nilai PC1 yang terinspirasi dari penelitian Almais et al. (2023). Penelitian selanjutnya dapat membandingkan pendekatan ini dengan metode *unsupervised clustering* murni (seperti *K-Means*, DBSCAN, atau *Gaussian Mixture Models* (GMM)) yang diterapkan pada *output* PCA untuk menemukan pola kluster yang mungkin berbeda atau lebih optimal.

## DAFTAR PUSTAKA

- Almais, A. T. W., Fatchurrohman, ., Holle, K. F. H., Kinasih, K. S., Wiranti, D. A., & Yasin, S. Y. (2020). *Implementation Fuzzy Weighted Product Preparation Post Disaster Reconstruction and Rehabilitation Action based Dynamics Decision Support System*. 272–277. <https://doi.org/10.5220/0009909002720277>
- Almais, A. T. W., Susilo, A., Naba, A., Sarosa, M., Crysdiyan, 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, 74590–74601. <https://doi.org/10.1109/ACCESS.2023.3275852>
- Almais, A. T. W., Susilo, A., Naba, A., Sarosa, M., Juwono, A. M., Crysdiyan, C., Muslim, M. A., & Wicaksono, H. (2024). Characterization of Structural Building Damage in Post-Disaster using GLCM-PCA Analysis Integration. *IEEE Access*, 12(August), 146190–146201. <https://doi.org/10.1109/ACCESS.2024.3469637>
- Azis, M. F. A., Darari, F., & Septyandy, M. R. (2020). Time series analysis on earthquakes using EDA and machine learning. *2020 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2020*, 423, 405–412. <https://doi.org/10.1109/ICACSIS51025.2020.9263188>
- Bai, T., & Tahmasebi, P. (2022). Attention-based LSTM-FCN for earthquake detection and location. *Geophysical Journal International*, 228(3), 1568–1576. <https://doi.org/10.1093/gji/ggab401>
- Bi, Y., Yao, Z. Q., Qin, Y. J., Chen, Q., Yang, C., & Kali, A. (2024). Experimental study on acoustic emission damage in precast reinforced concrete interior joints containing disc springs. *Structures*, 69. <https://doi.org/10.1016/j.istruc.2024.107271>
- Chaurasia, K., Kanse, S., Yewale, A., Singh, V. K., Sharma, B., & Dattu, B. R. (2019). Predicting Damage to Buildings Caused by Earthquakes Using Machine Learning Techniques. *2019 IEEE 9th International Conference on Advanced Computing (LACC)*, 81–86. <https://doi.org/10.1109/LACC48062.2019.8971453>
- Chencho, Li, J., & Hao, H. (2024). Structural damage quantification using long short-term memory (LSTM) auto-encoder and impulse response functions. *Journal of Infrastructure Intelligence and Resilience*, 3(2). <https://doi.org/10.1016/j.iintel.2024.100086>
- Cofre, A., Marin, M., Pino, O. V., Galleguillos, N., Riquelme, S., Barrientos, S., & Yoma, N. B. (2022). End-to-End LSTM-Based Earthquake Magnitude Estimation With a Single Station. *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5. <https://doi.org/10.1109/LGRS.2022.3175108>

- Ekaptiningrum, K. (2024). *Tren Gempa Bumi di Indonesia Terus Meningkatkan, Mitigasi Gempa Bumi Perlu Diperkuat*. Universitas Gadjah Mada.
- Fuentes, Y., Yari, Y., Soriano-Vargas, A., & Rocha, A. (2025). Peak Ground Acceleration Prediction for Earthquake Early Warning with Multivariable Long Short-Term Memory Networks and Temporal Transformers. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 15277 LNCS, 38–49. [https://doi.org/10.1007/978-3-031-80366-6\\_4](https://doi.org/10.1007/978-3-031-80366-6_4)
- Gentili, S., Brondi, P., Rossi, G., Sugan, M., Petrillo, G., Zhuang, J., & Campanella, S. (2024). Seismic clusters and fluids diffusion: a lesson from the 2018 Molise (Southern Italy) earthquake sequence. *Earth, Planets and Space*, 76(1), 157. <https://doi.org/10.1186/s40623-024-02096-3>
- Gu, J., Xie, Z., Zhang, J., & He, X. (2024). Advances in Rapid Damage Identification Methods for Post-Disaster Regional Buildings Based on Remote Sensing Images: A Survey. *Buildings*, 14(4), 898. <https://doi.org/10.3390/buildings14040898>
- Jian, P. R., & Wang, Y. (2022). Applying unsupervised machine-learning algorithms and MUSIC back-projection to characterize 2018–2022 Hualien earthquake sequence. *Terrestrial, Atmospheric and Oceanic Sciences*, 33(1), 28. <https://doi.org/10.1007/s44195-022-00026-y>
- Liu, Y., Lin, Y., Liu, W., Zhou, J., & Wang, J. (2024). Remote sensing perspective in exploring the spatiotemporal variation characteristics and post-disaster recovery of ecological environment quality, a case study of the 2010 Ms7.1 Yushu earthquake. *Geomatics, Natural Hazards and Risk*, 15(1). <https://doi.org/10.1080/19475705.2024.2314578>
- Manur, H. (2018). Structural control on hydrocarbon habitat in the Bawean area, East Java Sea. *Proc. Indon Petrol. Assoc., 23rd Ann. Conv.* <https://doi.org/10.29118/ipa.1158.129.144>
- Mia, M. S., Abdelmeguid, M., Harris, R. A., & Elbanna, A. E. (2024). Rupture Jumping and Seismic Complexity in Models of Earthquake Cycles for Fault Steppers with Off-Fault Plasticity. *Bulletin of the Seismological Society of America*, 114(3), 1466–1480. <https://doi.org/10.1785/0120230249>
- Punhani, A., Faujdar, N., Mishra, K. K., & Subramanian, M. (2022). Binning-Based Silhouette Approach to Find the Optimal Cluster Using K-Means. *IEEE Access*, 10(November), 115025–115032. <https://doi.org/10.1109/ACCESS.2022.3215568>
- Rifaldi, D. A., & Ahdika, A. (2024). K-Medoids and Support Vector Machine in Predicting the Level of Building Damage in Earthquake Insurance Modeling. *Statistika: Statistics and Economy Journal*, 104(3), 351–363. <https://doi.org/10.54694/stat.2024.13>
- Scala, S. A., Del Gaudio, C., & Verderame, G. M. (2024). Towards a multi-parametric fragility model for Italian masonry buildings based on the informative level. *Structures*, 59.

<https://doi.org/10.1016/j.istruc.2023.105613>

- Senkaya, M., Erkan, E. F., Silahtar, A., & Karaaslan, H. (2024). Implementation of clustering algorithms for damage prediction through seismic local-site parameters: 2023 Kahramanmaraş earthquake sequence. *Bulletin of Earthquake Engineering*, 22(13), 6545–6566. <https://doi.org/10.1007/s10518-024-02003-x>
- Shen, X., Hou, B., Lu, J., & Li, S. (2025). Comparative Analysis of Deep Learning Methods for Real-Time Estimation of Earthquake Magnitude. *Applied Sciences (Switzerland)*, 15(5), 2587. <https://doi.org/10.3390/app15052587>
- Tao, Y., Xu, Z. D., Wei, Y., Liu, X. Y., Zang, X., & Li, S. D. (2025). A wavelet packet deep learning model for Energy-Based structural collapse assessment under Earthquake-Fire Scenarios: Framework and hybrid simulation. *Mechanical Systems and Signal Processing*, 222, 111784. <https://doi.org/10.1016/j.ymsp.2024.111784>
- Wang, H., Xu, L., Xie, X., & Zhang, G. (2024). Improvement of the seismic resilience of regional buildings: A multi-objective prediction model for earthquake early warning. *Soil Dynamics and Earthquake Engineering*, 179, 108545. <https://doi.org/10.1016/j.soildyn.2024.108545>
- Yang, Y., Dou, J., Merghadi, A., Liang, W., Dong, A., Xiong, D., & Zhang, L. (2024). Advanced Prediction of Landslide Deformation through Temporal Fusion Transformer and Multivariate Time Series Clustering of InSAR: Insights from the Badui Region, Eastern Tibet. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–19. <https://doi.org/10.1109/TGRS.2024.3504241>
- Zhang, P., Li, X., & Chen, J. (2022). Prediction Method for Mine Earthquake in Time Sequence Based on Clustering Analysis. *Applied Sciences (Switzerland)*, 12(21). <https://doi.org/10.3390/app122111101>