

BAB III

METODOLOGI PENELITIAN

Bab ini menjelaskan secara rinci metodologi yang digunakan dalam penelitian ini, mulai dari gambaran umum objek penelitian, metode penelitian, alur penelitian, teknik pengumpulan data, hingga variabel penelitian. Penjelasan pada bab ini disusun secara sistematis agar pembaca dapat memahami proses dan langkah-langkah yang dilakukan selama penelitian berlangsung. Metodologi yang digunakan dirancang untuk memastikan keabsahan dan keandalan data serta memudahkan proses analisis dan interpretasi hasil penelitian. Dengan menguraikan setiap tahapan secara terstruktur, diharapkan pembaca dapat memahami bagaimana penelitian ini dilakukan dan dapat mengikuti prosedur yang digunakan sesuai dengan konteks penelitian yang relevan.

3.1 Gambaran Umum Objek Penelitian

Objek penelitian ini berupa dataset peristiwa gempa bumi di Indonesia yang diperoleh dari platform Kaggle, dengan sumber utama berasal dari Earthquake Repository yang dikelola oleh Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) serta Earthquake Catalog yang dikelola oleh U.S. Geological Survey (USGS). Dataset ini merupakan kompilasi data observasional yang menggabungkan catatan kejadian gempa dari kedua lembaga tersebut dan dipublikasikan oleh pengguna Kaggle bernama kekavigi untuk keperluan analisis terbuka. Dataset diunduh pada 25 Agustus 2025, dengan cakupan periode kejadian gempa bumi mulai dari 1 November 2008 hingga Oktober 2025. Seluruh data dalam dataset ini merepresentasikan hasil observasi seismik yang didokumentasikan secara rutin oleh BMKG dan USGS[50].

Dataset mencakup wilayah geografis antara 6° Lintang Utara hingga 11° Lintang Selatan serta 94° hingga 142° Bujur Timur, dengan rentang magnitudo antara 1 hingga 9,5 skala magnitudo dan kedalaman hiposenter yang bervariasi hingga ratusan kilometer. Secara keseluruhan, dataset penelitian pada Tabel 3.1 terdiri dari 131.833 *record* (baris) dengan 37 atribut (kolom) yang

memuat informasi spasial, temporal, dan seismik. Atribut-atribut tersebut antara lain mencakup *time* (waktu kejadian), *latitude* dan *longitude* (koordinat episenter), *depth* (kedalaman hiposenter), *magnitude* (magnitudo), *magType* (jenis skala magnitudo), *place* (lokasi sekitar episenter), serta sejumlah parameter mekanisme sumber gempa (*strike*, *dip*, *rake*).

Berikut ini Tabel 3.1 memberikan gambaran mengenai struktur data yang digunakan, berikut ditampilkan 5 *record* pertama dari dataset:

Tabel 3. 1 Dataset Gempa Bumi di Indonesia

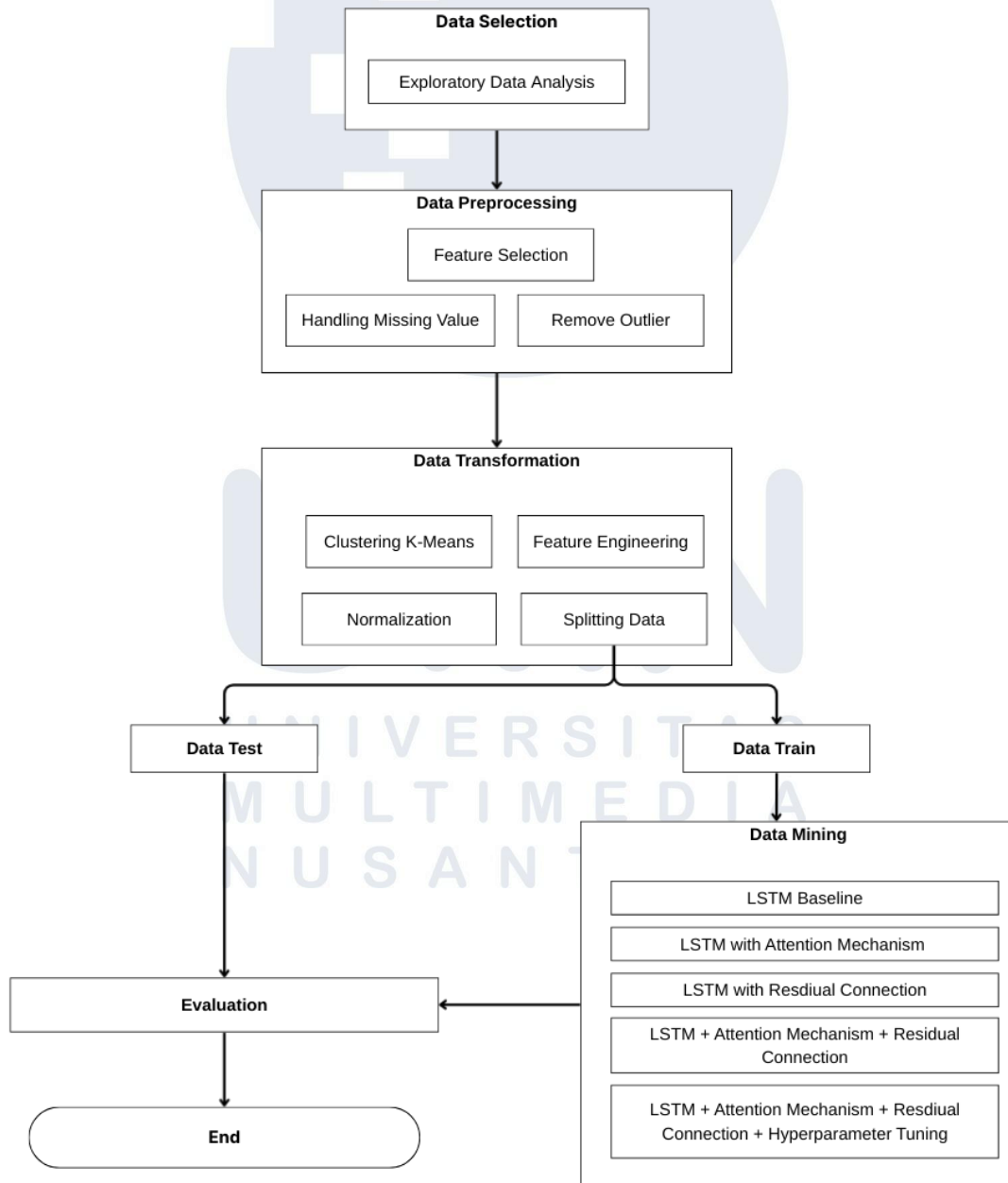
No	datetime	lat	lon	depth	mag	remark	...	misfit
1.	01/11/2008 21:02:43.058	-9.18	119.06	10	4.9	Sumba Region - Indonesia	...	NaN
2.	01/11/2008 20:58:50.248	-6.55	129.64	10	4.6	Banda Sea	...	NaN
3.	01/11/2008 17:43:12.941	-7.01	106.63	121	3.7	Java - Indonesia	...	NaN
4.	01/11/2008 16:24:14.755	-3.30	127.85	10	3.2	Seram - Indonesia	...	NaN
5.	01/11/2008 16:20:37.327	-6.41	129.54	70	4.3	Banda Sea	...	NaN
...
13 1.8 33	01/10/2025 00:00:35.181	-8.87	118.95	10	2.4	Sumba wa Region - Indonesia	...	NaN

3.2 Metode Penelitian

3.2.1 Alur Penelitian

Penelitian ini melakukan beberapa tahapan yang disusun secara terstruktur untuk mencapai tujuan yang telah ditetapkan. Diagram alur

penelitian pada Gambar 3.1 di bawah ini memberikan gambaran menyeluruh mengenai serangkaian proses yang dilalui.



Gambar 3. 1 Alur Penelitian

Berdasarkan Gambar 3.1, proses penelitian ini dibagi menjadi beberapa tahapan utama yang saling berhubungan. Berikut adalah penjelasan dari setiap tahapan yang dilakukan:

1) *Data Selection*

Tahapan awal penelitian ini adalah data selection, yaitu pengumpulan dan pemilihan dataset yang relevan dengan tujuan penelitian. Dataset diperoleh dari Kaggle.com, yang merupakan gabungan data gempa bumi dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) dan United States Geological Survey (USGS). Pada tahap ini dilakukan pemahaman awal terhadap karakteristik dan struktur data, meliputi jenis atribut seperti waktu, lokasi, kedalaman, dan magnitudo, rentang nilai data, serta identifikasi awal potensi permasalahan kualitas data seperti nilai hilang atau tidak valid. Proses data selection berfokus pada penyaringan dan pemilihan atribut yang memiliki keterkaitan langsung dengan tujuan pemodelan tanpa melakukan perubahan nilai data, sehingga data yang digunakan siap untuk diproses pada tahap selanjutnya.

a. *Exploratory Data Analysis*

Exploratory Data Analysis (EDA) dilakukan untuk menggali lebih dalam karakteristik dataset gempa bumi secara visual dan statistik. Pada tahap ini, analisis meliputi penyusunan *correlation matrix* untuk mengetahui hubungan antar variabel, visualisasi sebaran gempa berdasarkan koordinat geografis guna memahami pola spasial, serta analisis distribusi magnitudo gempa untuk mengidentifikasi tingkat intensitas kejadian. Selain itu, dilakukan pemetaan wilayah dengan aktivitas gempa tertinggi untuk mengamati konsentrasi kejadian gempa pada area tertentu. Melalui EDA, diperoleh gambaran awal mengenai pola temporal dan spasial yang membantu dalam memahami karakteristik data serta menjadi dasar dalam menentukan strategi pemodelan yang lebih tepat.

2) *Data Preprocessing*

Pada tahap data preprocessing, dilakukan pembersihan data untuk meningkatkan kualitas dataset yang akan digunakan.

Proses ini meliputi:

a. *Feature Selection*

Tahap ini dilakukan menggunakan metode filter berbasis korelasi, dengan mengacu pada hasil matriks korelasi pada tahap EDA. Fitur-fitur yang memiliki nilai korelasi absolut lebih dari 0,2 terhadap magnitudo dipilih sebagai fitur utama. Meskipun beberapa atribut seperti *datetime*, *latitude*, *longitude*, dan *location* memiliki korelasi rendah terhadap magnitudo, atribut tersebut tetap dipertahankan karena penelitian ini berfokus pada pemodelan data spasial-temporal, sehingga informasi waktu dan lokasi tetap relevan dalam proses prediksi.

b. *Handling missing value*, yaitu mengisi nilai kosong pada atribut numerik menggunakan nilai median karena median lebih tahan terhadap pengaruh *outlier*, sedangkan pada atribut kategorikal nilai kosong diisi dengan label “*unknown*” agar data tetap dapat digunakan dalam proses analisis dan pemodelan [51].

c. *Remove outlier*, yaitu menghapus data ekstrem menggunakan metode *Interquartile Range* (IQR). Metode ini dipilih karena data gempa bumi memiliki variasi yang tinggi, sehingga IQR mampu menentukan batas yang wajar dalam mendeteksi nilai ekstrem tanpa menghilangkan pola asli distribusi data gempa bumi.

3) *Data Transforming*

Pada tahap ini, data yang telah melalui proses pembersihan selanjutnya ditransformasikan agar memiliki skala, struktur, dan

representasi fitur yang sesuai dengan kebutuhan algoritma *data mining* dan *machine learning*. Dalam penelitian ini, *data transformation* dilakukan melalui beberapa proses utama, yaitu:

a. Clustering Spatial Data

Tahap berikutnya adalah pengelompokan data berdasarkan dimensi spasial. *Clustering K-Means* dilakukan untuk memetakan distribusi gempa bumi sesuai lokasi geografis, sehingga pola sebaran gempa dapat diketahui dengan lebih jelas[52]. Proses ini mengelompokkan koordinat gempa ke dalam beberapa wilayah berdasarkan kedekatan spasial, sehingga membentuk fitur kategorikal baru (*spatial_cluster*) yang merepresentasikan zona seismik tertentu. Fitur ini memberikan informasi spasial tingkat tinggi yang membantu model menangkap pola gempa yang spesifik pada wilayah tertentu dan memperkaya konteks dalam proses prediksi.

b. Feature Engineering

Tahapan ini dilakukan untuk memperkaya representasi data agar model dapat mengenali pola spasial dan temporal secara lebih efektif. Pada tahap ini dilakukan pemecahan fitur tanggal menjadi komponen tahun, bulan, hari, dan hari dalam minggu, serta transformasi fitur kategorikal *spatial_cluster* melalui teknik one-hot encoding untuk memungkinkan pemrosesan oleh model LSTM[10]. Hasilnya, dataset yang semula terdiri dari beberapa variabel utama diperluas menjadi 11 fitur yang lebih informatif, sehingga meningkatkan kemampuan model dalam memahami hubungan antarparameter geofisika.

c. Normalization

Data yang telah dipilih kemudian dinormalisasi agar seluruh fitur berada dalam skala yang sama. Penelitian ini menggunakan metode Min-Max Scaling yang mampu menyesuaikan semua variabel ke dalam rentang nilai tertentu[53]. Normalisasi ini dilakukan untuk mencegah adanya dominasi dari fitur dengan nilai yang lebih besar, mempercepat proses pelatihan, serta meningkatkan stabilitas model LSTM[54].

d. *Splitting Data*

Setelah data siap, dataset dibagi menjadi dua bagian utama, yaitu data latih (*training data*) dan data uji (*testing data*). Data latih digunakan untuk membantu model mempelajari pola-pola dalam data, sementara data uji digunakan untuk menilai kemampuan model dalam memprediksi nilai pada data yang belum pernah dilihat sebelumnya. Pembagian dataset ini dilakukan untuk memastikan proses evaluasi bersifat objektif dan bebas dari bias. Metode *chronological split* diterapkan dengan menggunakan 80% data pada periode awal sebagai data latih dan 20% data pada periode akhir sebagai data uji[10]. Baik data latih maupun data uji memiliki karakteristik yang konsisten karena keduanya berasal dari sumber dataset yang sama serta menggunakan variabel input dan output yang identik. Dengan demikian, tidak ada perbedaan dalam pengaturan data, kriteria inklusi, jenis prediktor, maupun outcome yang diprediksi.

4) Modeling

Tahap pemodelan merupakan bagian utama penelitian ini, di mana arsitektur *Long Short-Term Memory* (LSTM) dibangun untuk memprediksi magnitudo gempa berdasarkan data spasial-

temporal. LSTM dipilih karena kemampuannya dalam menangani data sekuensial dan pola dinamis pada kejadian gempa. Penelitian ini mengembangkan lima skenario pemodelan:

1. Model 1: LSTM dasar berbasis data spasial-temporal.
2. Model 2: LSTM dengan *attention mechanism*
3. Model 3: LSTM dengan *residual connection*
4. Model 4: LSTM dengan kombinasi *attention mechanism* dan *residual connection*.
5. Model 5: LSTM dengan *attention mechanism*, *residual connection*, dan *hyperparameter tuning*.

Pengembangan dimulai dari LSTM dasar sebagai baseline. Mekanisme *attention* dan *residual connection* ditambahkan untuk meningkatkan kemampuan model dalam menangkap pola penting dan menjaga stabilitas prediksi. Setelah memperoleh arsitektur yang lebih kuat, *hyperparameter tuning* diterapkan untuk mengoptimalkan performa model. Pendekatan ini disusun bertahap agar setiap peningkatan dapat dievaluasi secara jelas, sehingga terlihat model mana yang memberikan hasil paling optimal dalam memprediksi magnitudo gempa.

5) *Evaluation*

Performa model yang telah disesuaikan kemudian diukur menggunakan metrik evaluasi yang relevan[8][10], seperti *Mean Square Error* (MSE), *Mean Absolute Error* (MAE) dan *Root Mean Square Error* (RMSE). Hasil dari metrik ini akan memberikan penilaian kuantitatif mengenai seberapa akurat model dalam memprediksi magnitudo gempa. Analisis terhadap hasil evaluasi ini menjadi dasar untuk menarik kesimpulan dan menentukan keberhasilan model

6) *Comparing*

Tahap terakhir adalah proses perbandingan hasil antara model yang digunakan. Perbandingan dilakukan baik antarvarian model LSTM maupun dengan pendekatan lain untuk menilai efektivitas masing-masing. Melalui proses ini dapat ditentukan model yang paling optimal serta memberikan justifikasi ilmiah dalam pemilihan pendekatan yang digunakan.

Pada penelitian ini, tidak dilakukan uji stasioneritas dan differencing secara eksplisit. Hal ini dikarenakan metode yang digunakan adalah *Long Short-Term Memory* (LSTM), yang tidak mensyaratkan data bersifat stasioner seperti pada pendekatan *time series* klasik, misalnya ARIMA. Sebagai gantinya, data diproses melalui tahap normalisasi serta pembentukan data sekuens (*sliding window*) agar model dapat mempelajari pola temporal secara langsung dari data historis. Uji asumsi time series lainnya seperti *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) tidak menjadi fokus utama penelitian, karena evaluasi model lebih diarahkan pada kinerja prediksi menggunakan metrik RMSE dan MAE.

3.2.2 Metode Pengembangan Sistem / Metode Data Mining

Penelitian ini mengadopsi kerangka kerja KDD dengan langkah-langkah yang terperinci. Kerangka kerja ini memberikan struktur yang terorganisir, memfasilitasi perencanaan dan eksekusi setiap langkah dengan jelas. Fungsinya yang paling mencolok adalah membantu peneliti dalam memahami dan mengelola kompleksitas proses penelitian dari awal hingga akhir. Namun, pada penelitian ini peneliti hanya sampai pada tahap evaluasi dengan tujuan melihat akurasi dari hasil prediksi peneliti. Pemilihan kerangka kerja CRISP-DM dihasilkan dari membandingkan tiga metode *data mining*. Hal ini agar melihat pada Tabel 3.2 mengenai perbedaan pendekatan, kelebihan dan keterbatasannya dalam penelitian ini[55][56][57][58].

Tabel 3. 2 Perbandingan Metode Data Mining

Metode	CRISP-DM	SEMMA	KDD
Kelebihan	<p>1) Struktur proses yang jelas dan sistematis, cocok untuk proyek yang memerlukan data <i>preprocessing</i> mendalam.</p> <p>2) Fleksibel untuk menggabungkan data spasial dan temporal.</p> <p>3) Memiliki tahapan evaluasi model yang terintegrasi.</p>	<p>1) Fokus pada eksplorasi data yang mendalam, memudahkan pencarian pola awal dari data gempa.</p> <p>2) Cocok untuk eksperimen model cepat.</p>	<p>1) Menekankan pada penemuan pengetahuan baru, cocok untuk mengidentifikasi korelasi antar parameter gempa.</p> <p>2) Mencakup semua proses dari seleksi hingga penggunaan hasil.</p>
Kekurangan	<p>1) Dokumentasi dan analisis tahap awal memerlukan waktu lebih lama.</p> <p>2) Tidak fokus pada otomatisasi proses.</p>	<p>1) Kurang menekankan pemahaman konteks domain (seperti geofisika).</p> <p>2) Tahapan terbatas pada eksplorasi dan pemodelan, sehingga perlu metode tambahan untuk implementasi</p>	<p>1) Prosesnya lebih kompleks dan memakan waktu lama untuk data berukuran besar.</p> <p>2) Kurang fleksibel jika ingin iterasi model cepat.</p>

Metode	CRISP-DM	SEMMA	KDD
Tahapan	<i>Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, Deployment</i>	<i>Sample, Explore, Modify, Model, Assess</i>	<i>Selection, Preprocessing, Transformation, Data Mining, Interpretation/Evaluation, Usage</i>

Berdasarkan Tabel 3.2, penelitian ini menggunakan *Knowledge Discovery in Databases* (KDD) sebagai metodologi penelitian karena memiliki alur proses yang sistematis dan terdefinisi dengan baik, sehingga memudahkan pengelolaan penelitian dari tahap awal hingga tahap akhir. Pemilihan KDD mendukung tujuan penelitian ini, tahapan *data preparation* di dalamnya sangat relevan dengan kebutuhan pemodelan *Long Short-Term Memory* (LSTM) yang memanfaatkan data spasial-temporal, serta memungkinkan pengembangan, optimasi, dan evaluasi berbagai arsitektur LSTM secara terstruktur. Dengan alur KDD yang konsisten, penelitian ini dapat menganalisis kinerja model dasar maupun model LSTM yang diperluas dengan mekanisme *attention*, *residual connection*, serta *hyperparameter tuning*, sekaligus membandingkan performanya secara objektif untuk memperoleh model dengan kinerja terbaik. Adapun tahapan KDD yang diterapkan dalam penelitian ini adalah sebagai berikut:

a. *Data Selection*

Tahap ini dilakukan untuk memilih dataset gempa bumi yang sesuai dengan tujuan penelitian, baik dari segi sumber data maupun atribut yang digunakan, sehingga data yang dipilih relevan untuk proses analisis dan pemodelan.

b. *Data Preprocessing*

Tahap ini meliputi proses pembersihan data, seperti penghapusan atribut yang tidak relevan, penanganan nilai hilang (*missing value*), serta penghapusan outlier untuk memastikan kualitas dan konsistensi data sebelum dilakukan transformasi.

c. *Data Transformation*

Pada tahap ini, data yang telah dibersihkan ditransformasikan ke dalam bentuk yang sesuai untuk pemodelan, antara lain melalui normalisasi pada variabel numerik seperti koordinat geografis dan fitur lainnya, serta pembentukan fitur tambahan yang mendukung proses pembelajaran model.

d. *Data Mining (Modeling)*

Tahap ini merupakan inti dari proses KDD, yaitu penerapan algoritma LSTM untuk membangun model prediksi gempa bumi berdasarkan data temporal yang telah dipersiapkan pada tahap sebelumnya.

e. *Evaluation*

Model yang dihasilkan dievaluasi menggunakan metrik kinerja seperti *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil evaluasi ini digunakan untuk menilai tingkat akurasi dan performa model, serta menentukan apakah model telah memenuhi kriteria kinerja yang diharapkan atau masih memerlukan penyesuaian parameter.

3.3 Teknik Pengumpulan Data

Penelitian ini menggunakan teknik pengumpulan data sekunder yang diperoleh dari *Preliminary Earthquake Catalog* melalui platform Kaggle. Dataset ini merupakan hasil penggabungan data dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) sebagai lembaga pemerintah non-departemen di Indonesia, serta *Earthquake Catalog* yang dikelola oleh United States Geological Survey (USGS). Rentang waktu data mencakup periode

November 2008 hingga Oktober 2025. Dataset memuat sejumlah variabel yang relevan untuk analisis, antara lain *Date* (tanggal kejadian), *Time* (waktu kejadian), *Latitude* (lintang episentrum), *Longitude* (bujur episentrum), *Depth* (kedalaman gempa dalam kilometer), *Magnitude* (kekuatan gempa), dan *location* (keterangan wilayah kejadian berdasarkan *regions*). Pada beberapa kejadian tertentu, dataset juga menyertakan keterangan geografis pada *location*; dan atribut inti, seperti *magnitude* dan *mag_type* (misalnya MLv atau mb) yang mengukur kekuatan gempa, serta *phasecount* dan *azimuth_gap* yang menjadi indikator kualitas penentuan lokasi hiposentrum. Penggunaan data sekunder ini memungkinkan penelitian untuk melakukan analisis pola kejadian gempa bumi secara historis serta mengembangkan model prediksi magnitudo berbasis *Long Short-Term Memory* (LSTM) dengan memanfaatkan data spasial dan temporal yang telah terstandardisasi.

3.4 Variabel Penelitian

Dalam penelitian ini, variabel yang digunakan dibagi menjadi dua kategori utama, yaitu variabel input (prediktor) dan variabel output (target). Pembagian ini dimaksudkan untuk memberikan kejelasan mengenai peran masing-masing variabel dalam proses pemodelan menggunakan metode *Long Short-Term Memory* (LSTM). Dataset yang digunakan memuat catatan peristiwa gempa bumi mulai dari Januari 2024 hingga Oktober 2025. Pada dataset ditemukan banyak variabel atau kolom yaitu tanggal kejadian (*datetime*) koordinat lintang episentrum (*latitude*), koordinat bujur episentrum (*longitude*), kedalaman gempa (*depth*) dalam kilometer, magnitudo gempa (*magnitude*), serta keterangan wilayah tempat kejadian (*location*). Selain itu, dalam beberapa kejadian juga tersedia data mekanisme fokus. Meski demikian, fokus penelitian ini adalah pada variabel spasial-temporal, yang melibatkan data koordinat (lintang dan bujur), kedalaman, waktu (tanggal dan timestamp), serta magnitudo sebagai target utama dalam pemodelan LSTM. Pengelompokan variabel ini memberikan kejelasan fungsi masing-masing dalam proses pelatihan dan evaluasi model. Variabel-variabel utama yang digunakan dalam penelitian ini meliputi :

3.4.1 Variabel Input

Variabel input merupakan sekumpulan faktor yang digunakan sebagai masukan bagi model prediksi. Variabel-variabel ini dipilih karena dianggap relevan dan berpengaruh terhadap karakteristik kejadian gempa bumi. Adapun variabel input dalam penelitian ini meliputi:

1. Latitude (Lintang)

Variabel ini menunjukkan posisi geografis episenter gempa pada sumbu lintang. Lintang menjadi indikator penting karena distribusi gempa di Indonesia erat kaitannya dengan letak geografis yang berada di pertemuan tiga lempeng besar dunia, yaitu Lempeng Indo-Australia, Eurasia, dan Pasifik. Nilai lintang membantu model dalam mengenali pola sebaran kejadian gempa secara spasial, khususnya dalam membedakan kawasan rawan gempa di bagian utara maupun selatan Indonesia.

2. Longitude (Bujur)

Koordinat bujur memberikan informasi posisi episenter gempa pada sumbu horizontal bumi. Dengan adanya variabel bujur, model dapat memetakan distribusi kejadian gempa dari barat ke timur wilayah Indonesia. Kombinasi lintang dan bujur berperan penting dalam membangun konteks spasial, sehingga model mampu memahami keterkaitan antara lokasi geografis dan intensitas gempa yang terjadi.

3. Depth (Kedalaman)

Kedalaman hiposenter atau titik pusat gempa bumi di bawah permukaan bumi merupakan salah satu faktor krusial yang memengaruhi dampak gempa. Gempa dangkal (kedalaman < 60 km) umumnya menimbulkan kerusakan lebih besar dibandingkan gempa dalam (kedalaman > 300 km). Oleh karena itu, variabel kedalaman dimasukkan sebagai prediktor untuk membantu model

dalam membedakan potensi kekuatan gempa berdasarkan letak vertikalnya.

4. Time (Waktu Kejadian)

Variabel waktu direpresentasikan dalam bentuk data spasial-temporal, yang mencakup tahun, bulan, hari, hingga jam dan detik terjadinya gempa. Aspek temporal ini memungkinkan model untuk menangkap pola urutan kejadian gempa dalam rentang waktu panjang. Data temporal juga penting untuk mempelajari tren maupun siklus kejadian gempa, sehingga model tidak hanya belajar dari distribusi lokasi tetapi juga dari urutan peristiwa seismik sepanjang periode penelitian.

Dengan keempat variabel input ini (latitude, longitude, depth, dan time), penelitian ini menekankan pentingnya pendekatan spasial-temporal. Artinya, data yang digunakan tidak hanya menjelaskan di mana gempa terjadi, tetapi juga kapan gempa tersebut terjadi. Kombinasi kedua dimensi ini diharapkan dapat meningkatkan performa model prediksi berbasis LSTM karena LSTM unggul dalam mengolah data berurutan dan mampu mengenali pola yang kompleks.

3.4.2 Variabel Output

Variabel output merupakan hasil yang diprediksi oleh model. Dalam penelitian ini, variabel output adalah magnitudo gempa bumi (mag). Magnitudo merupakan ukuran kekuatan gempa yang bersifat kuantitatif, biasanya menggunakan skala magnitudo lokal (ML) atau skala moment magnitudo (Mw). Variabel ini dipilih sebagai target utama karena magnitudo berfungsi sebagai indikator kritis untuk menilai potensi dampak dan tingkat keparahan gempa bumi. Misalnya, gempa dengan magnitudo rendah (≤ 4) umumnya tidak menimbulkan kerusakan signifikan, sedangkan gempa dengan magnitudo ≥ 6 berpotensi menyebabkan kerusakan parah dan bahkan tsunami. Dengan menjadikan magnitudo sebagai variabel target, penelitian ini berupaya mengembangkan model prediksi yang dapat memberikan estimasi

kekuatan gempa di masa depan berdasarkan pola spasial-temporal dari data historis. Keberhasilan model dalam memprediksi magnitudo dengan akurat diharapkan dapat memberikan kontribusi pada pemahaman risiko bencana seismik dan mendukung sistem peringatan dini gempa di Indonesia.

Selain itu agar menjaga kualitas dan konsistensi data yang digunakan dalam pemodelan prediksi magnitudo gempa, penelitian ini memberlakukan pembatasan nilai variabel kontinu berdasarkan karakteristik geofisika wilayah Indonesia. Periode data temporal meliputi kejadian gempa mulai Januari 2024 hingga Oktober 2025, yang mencerminkan cakupan historis yang cukup komprehensif. Variabel spasial dibatasi pada rentang geografis 6° Lintang Utara hingga 11° Lintang Selatan dan 94° hingga 142° Bujur Timur, sesuai dengan batas wilayah Indonesia yang mengalami aktivitas seismik signifikan. Selain itu, variabel magnitudo berinterval 1.0 hingga 9.5 untuk mengakomodasi seluruh spektrum kekuatan gempa yang mungkin terjadi, sedangkan kedalaman hiposenter diatur dalam rentang 0 sampai sekitar 700 km, yang merefleksikan klasifikasi gempa dangkal hingga dalam. Pembatasan ini esensial untuk memastikan keberlanjutan validitas data serta memperkuat keakuratan hasil model, sekaligus menjaga relevansi penelitian terhadap konteks geofisika seismik di Indonesia.

3.5 Teknik Analisis Data

Analisis data dalam penelitian ini dilakukan secara bertahap mengikuti alur CRISP-DM. Proses dimulai dengan memahami konteks penelitian serta mengumpulkan dataset gempa bumi dari BMKG dan USGS melalui Kaggle. Data kemudian diperiksa untuk mengenali karakteristik, rentang nilai, serta potensi masalah kualitasnya. Setelah itu dilakukan pembersihan data untuk menghapus nilai hilang dan kolom yang tidak relevan.

Tahap berikutnya adalah eksplorasi dataset guna mengidentifikasi distribusi, pola spasial-temporal, dan hubungan antarvariabel sebagai dasar

perancangan model. Pemilihan fitur dilakukan secara manual berbasis pengetahuan domain, diikuti pembuatan fitur baru seperti pecahan tanggal dan encoding cluster spasial untuk memperkaya representasi data. Seluruh fitur kemudian dinormalisasi menggunakan Min-Max Scaling agar berada pada skala yang seragam.

Data yang telah siap dipisahkan secara kronologis menjadi data latih dan data uji dengan proporsi 80:20. Model LSTM kemudian dibangun melalui lima skenario pengembangan yang melibatkan penambahan attention mechanism, residual connection, serta hyperparameter tuning. Kinerja setiap model diukur menggunakan MSE, RMSE, dan MAE. Hasil evaluasi selanjutnya dibandingkan untuk menentukan model paling optimal dalam memprediksi magnitudo gempa.

