

## Penerapan Data Sains Untuk Klasifikasi Wilayah Yang Terdampak Kerusakan Gempa Bumi Dengan Metode C.45

Claudia Swastikawati <sup>1)</sup>, Muhamad Agoeng Pamoengkas <sup>2)</sup>, Alfian Cahyo Wahyudi <sup>3)</sup>, Kusrini <sup>4)</sup>

<sup>1)</sup> PJJ Informatika, Pascasarjana, Universitas AMIKOM Yogyakarta

<sup>2)</sup> PJJ Informatika, Pascasarjana, Universitas AMIKOM Yogyakarta

<sup>3)</sup> PJJ Informatika, Pascasarjana, Universitas AMIKOM Yogyakarta

<sup>4)</sup> PJJ Informatika, Pascasarjana, Universitas AMIKOM Yogyakarta

<sup>1)</sup> [claudia.swastikawati@students.amikom.ac.id](mailto:claudia.swastikawati@students.amikom.ac.id)

<sup>2)</sup> [agoeng@students.amikom.ac.id](mailto:agoeng@students.amikom.ac.id)

<sup>3)</sup> [alfian.wahyudi@students.amikom.ac.id](mailto:alfian.wahyudi@students.amikom.ac.id)

<sup>4)</sup> [kusrini@amikom.ac.id](mailto:kusrini@amikom.ac.id)

### Abstrak

Di Indonesia, gempa bumi sering terjadi karena interaksi plat tektonik. Energi seismik yang dicatat oleh seismograf diukur pada skala Richter (SR). Data dari BMKG dan BNPB selama semester pertama 2022 menunjukkan getaran energi seismik dari skala kecil hingga besar di sekitar Indonesia. Penelitian ini menggunakan metode C4.5 untuk klasifikasi dampak gempa, mengolah data historis mengenai wilayah, magnitudo, kedalaman, dan tingkat kerusakan. Proses penelitian mencakup pengumpulan data, pra-pemrosesan, pemilihan model, pelatihan model, dan penilaian kinerja menggunakan akurasi, presisi, recall, dan skor F1. Dataset berisi 62 catatan gempa dari BMKG dan dampak kerusakan di Jawa dari BNPB. Model C4.5 mencapai akurasi 62% pada tuning 8:2, meskipun menghadapi tantangan seperti variasi kondisi geografis, struktur bangunan, dan jumlah kejadian dalam satu hari. Hasil penelitian ini memberikan kontribusi berupa gabungan data bencana dari [www.bmkg.go.id](http://www.bmkg.go.id) dan kerusakan dari <https://gis.bnpb.go.id/>. Selain itu, model C4.5 dapat dilakukan untuk skala kerusakan rumah akibat gempa bumi. Penelitian ini menghasilkan model prediksi yang akurat untuk mengklasifikasikan tingkat kerusakan bangunan, memberikan wawasan praktis untuk mitigasi bencana, dan mendukung pemerintah serta lembaga penanggulangan bencana dalam meningkatkan sistem peringatan dini dan distribusi bantuan.

**Kata kunci :** Algoritma C4.5, Gempa bumi, Model klasifikasi, Mitigasi bencana

### Abstract

*In Indonesia, earthquakes frequently occur due to tectonic plate interactions. Seismic energy recorded by seismographs is measured on the Richter scale (SR). Data from BMKG and BNPB during the first half of 2022 show seismic energy vibrations ranging from small to large scales around Indonesia. This study uses the C4.5 method to classify earthquake impacts by processing historical data on regions, magnitudes, depths, and previous damage levels. The research process includes data collection, preprocessing, model selection, model training, and performance evaluation using accuracy, precision, recall, and F1 score metrics. The dataset contains 62 earthquake records from BMKG and damage impacts in Java from BNPB. The C4.5 model achieved 62% accuracy with an 8:2 tuning ratio, despite challenges such as geographical variations, building structures, inconsistent locations, and the number of events in a single day. This study provides a dataset combining disaster data from [www.bmkg.go.id](http://www.bmkg.go.id) and damage data from <https://gis.bnpb.go.id/>. Additionally, the C4.5 model can be applied to assess the scale of damage to houses caused by earthquakes. This research has developed an accurate predictive model for classifying the extent of building damage, providing practical insights for disaster mitigation and supporting the government and disaster response agencies in enhancing early warning systems and aid distribution.*

**Keywords:** C4.5 algorithm, earthquake, classification model, disaster mitigation

## **1. PENDAHULUAN**

Indonesia merupakan negara yang sangat rentan terhadap bencana gempa bumi karena terletak di kawasan Cincin Api Pasifik, tempat bertemunya tiga lempeng tektonik besar: Indo-Australia, Eurasia, dan Pasifik. Aktivitas seismik yang tinggi sering menyebabkan gempa bumi di berbagai wilayah Indonesia (SRIKANTH, et al., 2022; Wahyu Pratama et al., 2024). Berdasarkan data dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) serta Badan Nasional Penanggulangan Bencana (BNPB), gempa bumi dengan berbagai magnitudo terjadi sepanjang tahun, menimbulkan kerusakan signifikan pada bangunan dan infrastruktur serta mengakibatkan korban jiwa (A. Gaba et al., 2019; K. Chaurasia, 2019).

Penanganan bencana gempa bumi memerlukan penilaian cepat dan akurat terhadap tingkat kerusakan yang ditimbulkan. Hal ini penting untuk mendukung distribusi bantuan, perencanaan rehabilitasi, dan rekonstruksi pasca bencana (SRIKANTH, et al., 2022; Wahyu Pratama et al., 2024). Kompleksitas faktor-faktor yang mempengaruhi kerusakan bangunan, seperti intensitas gempa, kedalaman gempa, jenis tanah, serta kualitas konstruksi, membuat penilaian ini menjadi tugas yang menantang (K. Chaurasia, 2019; Oktarina et al., 2020).

Seiring dengan perkembangan teknologi, penerapan ilmu data dan pembelajaran mesin (machine learning) dalam analisis dampak bencana menunjukkan potensi besar. Pembelajaran mesin memungkinkan analisis data dalam skala besar dan identifikasi pola yang mungkin tidak terlihat melalui metode tradisional (Wahyu Pratama et al., 2024). Algoritma C4.5 adalah salah satu yang sering digunakan dalam klasifikasi dan prediksi karena kemampuannya menangani data kontinu dan diskrit serta menghasilkan model pohon keputusan yang mudah diinterpretasikan (Oktarina et al., 2020).

Penelitian ini bertujuan mengembangkan model prediksi untuk mengklasifikasikan tingkat kerusakan bangunan akibat gempa bumi menggunakan algoritma C4.5. Data yang digunakan mencakup informasi historis mengenai gempa bumi, seperti magnitudo, kedalaman, dan lokasi, serta tingkat kerusakan yang dilaporkan. Dengan data tersebut dapat diambil tujuan penelitian ini adalah untuk mengumpulkan dan memproses data gempa bumi serta kerusakan bangunan dari sumber terpercaya seperti BMKG dan BNPB, yang menyediakan data relevan mengenai kejadian gempa bumi dan dampaknya di Indonesia.

Selain itu, penelitian ini bertujuan untuk membangun model klasifikasi menggunakan algoritma C4.5, yang dipilih karena kemampuannya menangani data kontinu dan diskrit serta menghasilkan model pohon keputusan yang mudah diinterpretasikan (K. Chaurasia, 2019; Oktarina et al., 2020). Kinerja model ini akan dievaluasi menggunakan metrik akurasi, presisi, recall, dan skor F1 untuk memastikan keandalan prediksi yang dihasilkan (Oktarina et al., 2020). Akhirnya, penelitian ini juga bertujuan memberikan rekomendasi praktis untuk mitigasi bencana dan alokasi sumber daya yang lebih efektif selama dan setelah gempa bumi. Hasil penelitian ini diharapkan dapat mendukung upaya pemerintah dan lembaga penanggulangan bencana dalam meningkatkan sistem peringatan dini dan distribusi bantuan.

Melalui penelitian ini, diharapkan dapat dihasilkan model prediksi yang andal dan akurat dalam mengklasifikasikan tingkat kerusakan bangunan. Model ini diharapkan mendukung upaya pemerintah dan lembaga penanggulangan bencana dalam memperbaiki sistem peringatan dini dan distribusi bantuan, serta berkontribusi pada peningkatan keselamatan dan kesejahteraan masyarakat di daerah rawan gempa (K. Chaurasia, 2019; Oktarina et al., 2020). Selain itu, penelitian ini juga diharapkan dapat memberikan wawasan lebih mendalam tentang potensi penggunaan teknik pembelajaran mesin dalam mitigasi bencana di masa depan.

Dengan demikian, penelitian ini tidak hanya memiliki kontribusi teoretis dalam bidang ilmu data dan pembelajaran mesin, tetapi juga memberikan manfaat praktis yang signifikan dalam penanganan bencana gempa bumi di Indonesia. Pengembangan lebih lanjut dari model ini, termasuk peningkatan kualitas data dan penggunaan teknik pemrosesan data yang lebih canggih, dapat semakin meningkatkan akurasi dan keandalan prediksi, sehingga mendukung upaya mitigasi bencana yang lebih efektif dan efisien (Wahyu Pratama et al., 2024).

## 2. KAJIAN PUSTAKA

Gempa bumi merupakan salah satu bencana alam yang paling merusak, menyebabkan kerusakan signifikan pada bangunan, infrastruktur, dan hilangnya nyawa. Penilaian yang akurat dan tepat waktu terhadap kerusakan yang disebabkan oleh gempa bumi sangat penting untuk respons bencana dan upaya mitigasi yang efektif. Penilaian keparahan gempa bumi merupakan bidang yang kompleks dan memiliki kesulitan tersendiri untuk dilakukan secara tepat dan cepat walaupun sudah banyak dikembangkan teknik pengklasifikasian, belum serta merta dapat digunakan sebagai tolak ukur. Penerapan teknik pembelajaran mesin telah menunjukkan potensi besar dalam meningkatkan prediksi dan klasifikasi kerusakan akibat gempa bumi, khususnya menggunakan algoritma C4.5 (A. Gaba et al., 2019; SRIKANTH, et al., 2022).

### 2.1. Penggunaan Algoritma C4.5

Algoritma C4.5 adalah pengklasifikasi pohon keputusan yang dikenal karena kemampuannya menangani data kontinu dan diskrit, membuatnya sangat efektif untuk membangun pohon keputusan yang dapat diinterpretasikan. Metode ini telah digunakan untuk menganalisis berbagai faktor yang berkontribusi terhadap kerusakan gempa bumi, seperti bahan bangunan, kualitas konstruksi, dan kedekatan dengan episentrum gempa (K. Chaurasia, 2019; Oktarina et al., 2020). Dengan memanfaatkan kekuatan pembelajaran mesin, khususnya algoritma C4.5, memungkinkan pengembangan model yang dapat memprediksi tingkat kerusakan pada bangunan setelah gempa bumi. Model-model ini menggunakan berbagai kriteria termasuk usia bangunan, jenis fondasi, jumlah lantai, dan bahan konstruksi (Mangalathu et al., 2020). Dengan mengklasifikasikan bangunan berdasarkan skala kerusakan, menjadi mungkin untuk mengidentifikasi area berisiko tinggi dan memprioritaskannya untuk operasi penyelamatan dan bantuan, yang pada akhirnya mengurangi korban jiwa dan mengoptimalkan alokasi sumber daya.

### 2.2. Dataset dan Akurasi Model

Penelitian ini menghasilkan dataset berupa data dampak bencana dan hasil prediksi skala kerusakan akibat gempa bumi. Penelitian dilakukan dengan model C4.5 berdasarkan dataset sebanyak 62 yang diperoleh dari catatan kejadian gempa bumi BMKG dan dampak kerusakan yang tercatat di wilayah Jawa pada portal BNPB. Model yang dikembangkan mencapai tingkat akurasi sebesar 62%, pada tuning 8:2, yang menunjukkan kemampuannya dalam memprediksi pola kerusakan yang terjadi. Namun, hasil ini juga mengungkap beberapa tantangan yang dihadapi dalam pengembangan model, seperti variasi yang signifikan dalam kondisi geografis, struktur bangunan di wilayah yang luas, ketidak konsistenan titik lokasi, dan jumlah kejadian dalam hari yang sama sangat berpengaruh terhadap jumlah kerusakan akibat dampak gempa bumi.

### 2.3. Penggunaan Teknik Pembelajaran Mesin Lainnya

Penelitian sebelumnya telah mengeksplorasi penggunaan algoritma pembelajaran mesin, khususnya pohon keputusan dan algoritma C4.5, untuk prediksi kerusakan akibat gempa. (Sinaga et al., 2023) mengembangkan aplikasi web menggunakan C4.5 untuk memprediksi dampak gempa bumi di Indonesia berdasarkan Skala Intensitas Gempa BMKG dengan perolehan akurasi mencapai 80%, precision = 78.41%, dan recall = 75% dengan model terbaik pada split data 9:1 (A. Gaba et al., 2019). Penelitian lain, (Tantyoko et al., 2023), membandingkan algoritma Random Forest dan feature selection untuk memprediksi tingkat kerusakan bangunan pasca gempa mencapai akurasi hingga 70.83.

### 2.4. Keunggulan dan Tantangan Penggunaan Model Pembelajaran Mesin

Studi-studi ini menunjukkan potensi teknik pembelajaran mesin, khususnya C4.5 dan metode berbasis pohon keputusan lainnya, dalam meningkatkan akurasi prediksi kerusakan gempa bumi dan menyediakan alat berharga bagi upaya mitigasi bencana alam (Kubo et al., 2024; Mangalathu et al., 2020). Efektivitas model pembelajaran mesin dalam prediksi kerusakan akibat gempa bumi telah didokumentasikan dengan baik. Misalnya, studi-studi telah menyoroti penggunaan algoritma pohon keputusan dalam memprediksi kerusakan pada bangunan beton bertulang, menunjukkan potensi model-model ini dalam skenario dunia nyata.

### 2.5. Pengembangan Model Lebih Lanjut

Mousavi dan Beroza (2019) menunjukkan penerapan pembelajaran mesin untuk estimasi magnitudo gempa bumi, yang merupakan faktor kritis dalam menilai potensi kerusakan (Mousavi & Beroza, 2020). Selanjutnya, Ghimire et al. (2022) telah menguji berbagai model pembelajaran mesin untuk prediksi kerusakan seismik, menekankan pentingnya kualitas data dan pemilihan fitur dalam meningkatkan akurasi prediksi (Ghimire et al., 2022). Miura et al. (2020) mengembangkan metodologi berbasis CNN untuk identifikasi otomatis kerusakan bangunan dari citra udara pasca bencana, menunjukkan akurasi sekitar 95% dalam mengklasifikasikan kerusakan bangunan (Miura et al., 2020). Selain itu, Ding et al. (2022) mengusulkan metode efisien untuk mendeteksi bangunan yang runtuh menggunakan citra UAV dan algoritma Faster R-CNN, mencapai tingkat presisi dan penarikan yang tinggi (Ding et al., 2022).

Secara keseluruhan, integrasi teknik pembelajaran mesin seperti algoritma C4.5 dalam klasifikasi kerusakan akibat gempa bumi menawarkan solusi untuk manajemen bencana. Ini memfasilitasi penilaian cepat terhadap area yang terdampak, memungkinkan pengambilan keputusan yang tepat dan informasi dalam keadaan pasca gempa bumi (Irawan & Hermawan Hasibuan, 2020). Pendekatan ini tidak hanya meningkatkan efisiensi respons darurat tetapi juga berkontribusi pada pengembangan infrastruktur dan komunitas yang lebih tangguh (Cheng et al., 2023; Mousavi & Beroza, 2020; Oktarina et al., 2020).

## **2.6. Novelty**

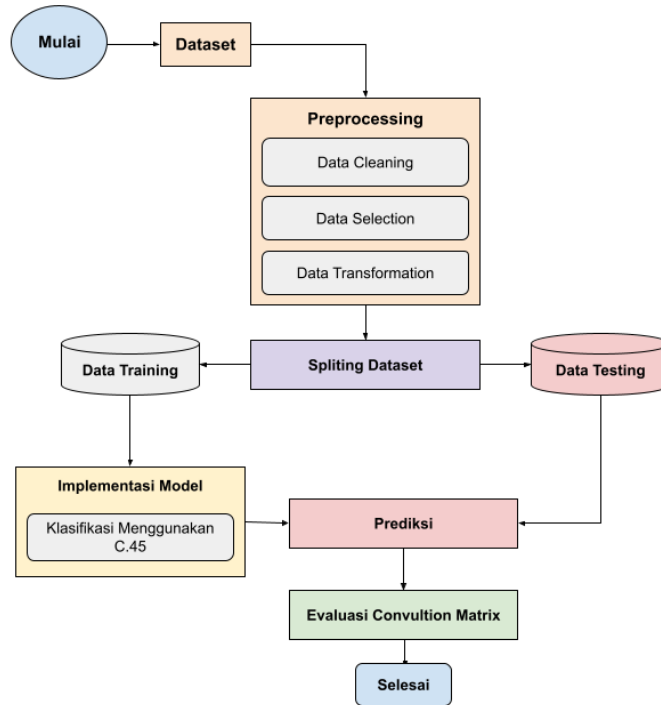
Penelitian ini memperkenalkan beberapa aspek baru dan berbeda dibandingkan dengan penelitian sebelumnya. Salah satu perbedaan utama adalah fokus pada penggunaan dataset terbaru yang mencakup catatan gempa bumi dan dampak kerusakan di wilayah Jawa selama semester pertama tahun 2022. Berbeda dengan penelitian sebelumnya yang menggunakan dataset yang lebih umum atau mencakup periode waktu yang lebih luas (Tantyoko et al., 2023), penelitian ini memberikan fokus khusus pada wilayah dan periode tertentu, sehingga memberikan konteks yang lebih relevan dan spesifik. Selain itu, penelitian ini menggunakan algoritma C4.5 dengan tuning khusus 8:2, sebuah pendekatan yang belum banyak dieksplorasi dalam konteks prediksi kerusakan gempa bumi di Indonesia. Hal ini memungkinkan evaluasi yang lebih mendalam terhadap kemampuan model dalam mengklasifikasikan kerusakan bangunan berdasarkan variasi geografis, struktur bangunan, dan faktor-faktor lokal lainnya (Sinaga et al., 2023).

Selain itu, penelitian ini menggunakan data gabungan dari dua sumber utama, yaitu BMKG dan BNPB, yang memberikan informasi yang lebih komprehensif dan detail mengenai kejadian gempa dan dampaknya. Ini memungkinkan pengembangan model prediksi yang lebih akurat dan spesifik untuk konteks Indonesia, terutama untuk wilayah Jawa yang sangat rentan terhadap gempa bumi. Penelitian ini juga menekankan pentingnya pertimbangan faktor-faktor lokal seperti variasi kondisi geografis dan struktur bangunan dalam pengembangan model prediksi. Pendekatan ini memberikan kontribusi baru tidak hanya dalam hal dataset dan metode analisis, tetapi juga dalam pendekatan kontekstual yang lebih mendalam terhadap prediksi kerusakan gempa bumi. Ini diharapkan dapat menjadi dasar untuk penelitian lebih lanjut dan pengembangan model yang lebih baik untuk mitigasi bencana di masa depan.

## **3. METODOLOGI PENELITIAN**

Pada bagian ini akan membahas metode penelitian yang digunakan dalam penelitian ini, hal ini sangat penting karena sebagai runtutan proses untuk mencapai hasil yang diharapkan. Penelitian membahas mengenai prediksi kerusakan bangunan akibat gempa dengan menggunakan metode Algoritma C4.5, yakni sebuah metode pohon keputusan yang diterapkan untuk mengklasifikasikan tingkat intensitas gempa berdasarkan Skala Intensitas Gempa BMKG (Sinaga et al., 2023). Teknik pengumpulan data dilakukan dengan memanfaatkan dokumentasi dari website BMKG dan BNPB, serta melakukan studi literatur (Sinaga et al., 2023). Pendekatan jaringan Bayesian yang lebih baik, dengan mengintegrasikan model cloud, telah menunjukkan akurasi tinggi dalam memprediksi kerusakan bangunan, dengan struktur bangunan dan material pondasi diidentifikasi sebagai faktor kritis (Cheng et al., 2023). Model jaringan Bayesian juga telah diterapkan untuk mengetahui risiko kerusakan bangunan akibat gempa bumi di Indonesia, mengingat lokasinya yang

berada di pertemuan tiga lempeng tektonik besar (Tantyoko et al., 2023). Cara-cara tersebut bertujuan untuk meminimalkan potensi kerugian dengan memperkirakan kemungkinan terjadinya kerusakan bangunan di daerah rawan gempa. Metode penelitian ini akan mencakup langkah-langkah yang akan diambil dalam pengumpulan data, preprocessing data, mengimplementasikan algoritma, menganalisis hasil. Berikut bagan alir yang akan dilakukan dalam penelitian ini.



Gambar 1. Alur Penelitian

### 3.1. Dataset

Dataset dikumpulkan dari dua sumber yakni Badan Meteorologi Klimatologi dan Geofisika (BMKG) dan Badan Nasional Penanggulangan Bencana (BNPB) yang di scrapping dari website BMKG dan BNPB dalam dua tahun yakni pada periode tahun 2022 sampai tahun 2023. Data yang didapat bertipe time series yang mempunyai 14 variabel dan terdiri dari no, tanggal, kabupaten, propinsi, magnitudo, waktu, lintang bujur, kedalaman, meninggal, hilang, terluka, rumah rusak, fasum rusak. Untuk mengurangi kelas yang begitu banyak maka, penelitian ini akan menjadikan atribut rumah rusak sebagai label dan akan dilakukan pengkategorian atribut menjadi rumah rusak “sedikit”, “banyak”, dan “tidak ada”. Tabel 1 menampilkan data dampak bencana dari BPBD yang dapat diakses melalui portal data bencana <https://gis.bnpb.go.id/>, dimana memberikan data detail mengenai variabel-variabel dampak bencana yang digunakan dalam penelitian ini. Selanjutnya, Tabel 2 berisi dataset dari BMKG yang dapat diakses melalui situs <https://dataonline.bmkg.go.id/>, memberikan informasi tentang kejadian gempa dan karakteristiknya yang relevan untuk analisis. Tabel 3, sebagai dataset gabungan yang berasal dari Tabel 1 dan cross-check dengan Tabel 2, menunjukkan integrasi data dari kedua sumber tersebut untuk analisis yang komprehensif.


Tabel 1. Dataset mentah dari BNPB

No.	Kode Identitas Gempa	ID Kabupaten	Tanggal Kejadian	Kejadian	Lokasi	Kabupaten	Provinsi	Kronologi & Dokumentasi	Penyebab	Meninggal	Hilang	Terluka	Rumah Rusak	Rumah Terendam	Fasum Rusak
1	320410820230120311	3204	31/12/2023	GEMPA BUMI	Kec. Arjasari Ds. Pinggirsari Kec. Cicalengka Ds. Nagrog	BANDUNG	JAWA BARAT	Dokumentasi	Gempa Pertama Kekuatan : 4.1 SR Tanggal : 31-Dec-2023 Waktu Gempa : 14:35:34 WIB Lintang : 6.84 LS Bujur : 107.93 BT Kedalaman : 10 Km Gempa Kedua : Kekuatan : 4.8 SR Tanggal : 31-Dec-2023 Waktu Gempa : 20:34:24 WIB Lintang : 6.85 LS Bujur : 107.94 BT Kedalaman : 5 Km				3		1



2	3211 1082023 012031 1	3211	31/12/2023	GEMPA BUMI	Kec. Sumedang Selatan Ds. Kotakulon Ds. Cipameungpeuk Kec. Tanjung Medar Ds. Sukatani Kec. Sumedang Utara Ds. Rancamulya Ds. Padasuka Ds. Kotakelar Ds. Margamukti Ds. Talun Ds. Kebon jati	SU ME DA NG	JAWA BARAT	Dokumentasi	Gempa Pertama Kekuatan : 4.1 SR Tanggal : 31-Dec-2023 Waktu Gempa : 14:35:34 WIB Lintang : 6.84 LS Bujur : 107.93 BT Kedalaman : 10 Km Gempa Kedua : Kekuatan : 4.8 SR Tanggal : 31-Dec-2023 Waktu Gempa : 20:34:24 WIB Lintang : 6.85 LS Bujur : 107.94 BT Kedalaman : 5 Km	18	1350	61
...	...	...	...	...	...	...	...	...	...	...	...	...

Tabel 2. Data set dari BMKG

 Badan Meteorologi Klimatologi dan Geofisika (BMKG) Sistem Peringatan Dini Tsunami Indonesia (InaTEWS) Lintang : -11 s/d 6 Bujur : 94 s/d 142 Kedalaman : 1.00 s/d 1000.00 Magnitude : 1.00 s/d 9.50 Rentang Waktu : 01-01-2022 s/d 31-12-2023				
Tanggal (GMT)	Lintang (°)	Bujur (°)	Kedalaman (km)	Magnitudo (SR)
2022-01-01 23:50:06	-8.13	122.38421	10	2.94
2022-01-01 23:48:17	-8.85	124.10889	76.3	3.37
2022-01-01 22:30:50	-3.59	140.21089	11.5	3.8
2022-01-01 22:15:37	-8.03	122.34876	10	3.48
...	...	...	...	...

Tabel 3. Dataset Gabungan

No	Tanggal	Kabupaten	Provinsi	Magnitudo	Waktu	Lintang	Bujur	Kedalaman	Meninggal	Hilang	Terluka	Rumahnya rusak	Fasus rusak
1	12/31/2023	Bandung	Jawa Barat	4.1	14:35:34	6.84	107.93	10.0	0	0	0	3	1
2	12/31/2023	Bandung	Jawa Barat	4.8	20:34:24	6.85	107.94	5.0	0	0	0	0	0
3	12/31/2023	Sumedang	Jawa Barat	4.1	14:35:34	6.84	107.93	10.0	0	0	18	1350	61
4	12/31/2023	Sumedang	Jawa Barat	4.8	20:34:24	6.85	107.94	5.0	0	0	0	0	0
5	12/31/2023	Subang	Jawa Barat	4.1	14:35:34	6.84	107.93	10.0	0	0	0	37	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...

### 3.2. Preprocessing

Preprocessing yang kami lakukan adalah tahapan krusial dalam persiapan data mentah yang kami peroleh dari web scraping BMKG dan BNPB sebelum melakukan pelatihan model. Proses ini mencakup pembersihan data, seleksi data, dan transformasi data. File Excel yang diperoleh berisi detail lengkap mengenai kejadian bencana seperti kode identifikasi, kronologi, dokumentasi kerusakan, dan penyebab yang dalam kasus ini adalah gempa bumi. File tersebut dapat diakses pada laman <https://gis.bnpb.go.id/>. Semua informasi masih terkumpul dalam satu file, memerlukan pemisahan dan pembersihan data agar dapat digunakan dalam proses berikutnya. Tindakan persiapan data melibatkan ekstraksi detail kejadian gempa bumi ke dalam tabel yang terstruktur, dengan fokus pada peristiwa yang terjadi di Pulau Jawa. Langkah selanjutnya meliputi pembersihan data dan proses scaling untuk memudahkan implementasi model prediksi.

Setelah memperoleh data yang dibutuhkan kami menggunakan matriks korelasi, Gambar 2, untuk menemukan hubungan antar atribut dalam mempengaruhi prediksi dampak kerusakan akibat gempa bumi.



Gambar 2. Matriks Korelasi

Dari matrik korelasi diatas diperoleh hubungan antar atribut yang paling berpengaruh untuk memprediksi jumlah kerusakan yakni rumah rusak, fasum rusak, magnitudo, kedalaman, dan provinsi.

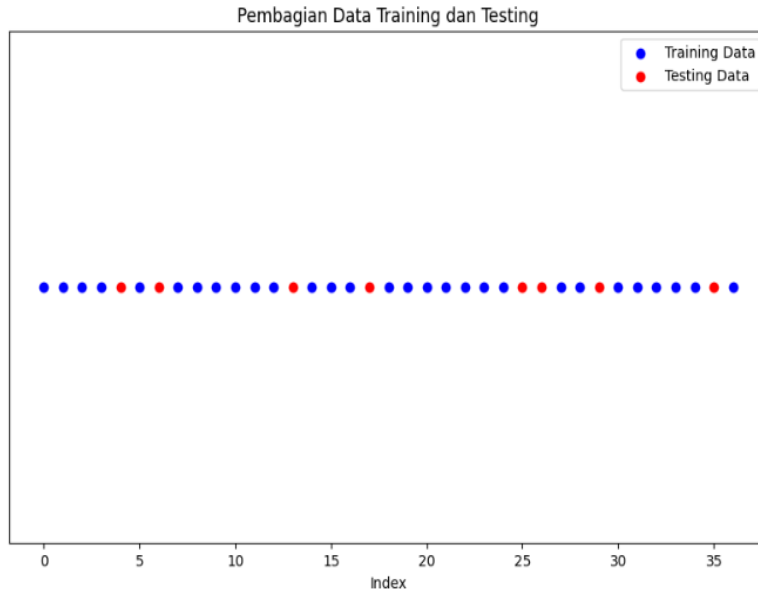
Kami mengidentifikasi bahwa beberapa gempa terjadi pada tanggal yang sama. Hal ini menyebabkan adanya perulangan kejadian gempa yang signifikan mempengaruhi tingkat kerusakan yang tercatat. Untuk memperkuat analisis kami, kami memutuskan untuk menambahkan kolom yang mencatat jumlah kejadian gempa bumi tersebut. Dengan demikian, kami menghasilkan data akhir yang terdiri dari 36 kejadian gempa yang tercatat di wilayah Jawa. Tabel Dataset jadi yang telah melalui preprocessing tersedia pada Tabel 4. (Sebagian)

Tabel 4. Tabel dataset setelah preprocessing

Jumlah kejadian	Provinsi	Magnitudo (SR)	Kedalaman (KM)	Rumah rusak
2	JAWA BARAT	Rendah	Sedang	Sedikit
2	JAWA BARAT	Rendah	Sedang	Banyak
2	JAWA BARAT	Rendah	Sedang	Banyak
1	JAWA BARAT	Sedang	Sedang	Sedikit
1	JAWA TENGAH	Rendah	Sedang	Banyak
1	JAWA BARAT	Rendah	Sedang	Banyak
1	JAWA BARAT	Rendah	Sedang	Banyak
1	JAWA BARAT	Rendah	Sedang	Banyak
1	JAWA BARAT	Rendah	Sedang	Banyak
...	...	...	...	...

### 3.3. Splitting Dataset

Untuk memperoleh hasil evaluasi model yang akurat, maka splitting dataset sangat diperlukan agar model dapat melakukan pelatihan dan pengujian pada data yang belum pernah dilihat sebelumnya. Ini akan memberikan gambaran yang realistis tentang bagaimana model akan berkinerja terhadap data baru. Dalam penelitian ini dilakukan splitting data 80% untuk training dan 20% untuk pengujian. Gambar 3 memberikan gambaran jelas tentang bagaimana dataset awal dibagi menjadi dua bagian, dengan 29 data training digunakan untuk melatih model dan 8 data testing sisanya untuk menguji performa model..



Gambar 3. Pembagian data *training* dan data *testing*

### 3.4. Klasifikasi C.45

Model C.45 dipilih karena kemampuannya dalam menangani data kontinu dan diskrit sangat baik dan paling banyak digunakan dalam berbagai aplikasi machine learning dan data mining (Irawan & Hermawan Hasibuan, 2020). Dalam kelebihanannya yaitu output dalam bentuk aturan hal ini yang sering lebih mudah dipahami dan digunakan oleh manusia.

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang dapat mengatasi nilai yang hilang (missing value) dan variabel kontinu dalam pembuatan pohon Keputusan (Irawan & Hermawan Hasibuan, 2020). Algoritma ini menggunakan Gain Ratio untuk memilih atribut terbaik dalam membangun pohon keputusan berdasarkan data  $D$  yang terdiri dari instance-instance  $t_1, t_2, \dots, t_n$ , kelas-kelas  $C = \{C_1, C_2, \dots, C_m\}$ , dan himpunan atribut  $A = \{A_1, A_2, \dots, A_h\}$  (Sinaga et al., 2023). Langkah-langkah membangun pohon keputusan menggunakan metode Decision Tree dengan algoritma C4.5 adalah sebagai berikut (Sinaga et al., 2023) :

1. Menghitung Entropi Awal (Entropy(D)):

Hitung entropi dari data  $D$  menggunakan rumus:

$$info(D) = - \sum_{i=1}^m p_i \log_2 p_i \quad (1)[12]$$

$$p_i = \frac{freq(C_i, D)}{|D|} \quad (2)[12]$$

2. Menghitung Entropi untuk setiap atribut

$$info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times info(D_j) \quad (3)[12]$$

3. Memilih Atribut Terbaik untuk Splitting:

Hitung nilai Gain Ratio

$$Gain(A) = info(D) - info_A(D) \quad (4)[12]$$



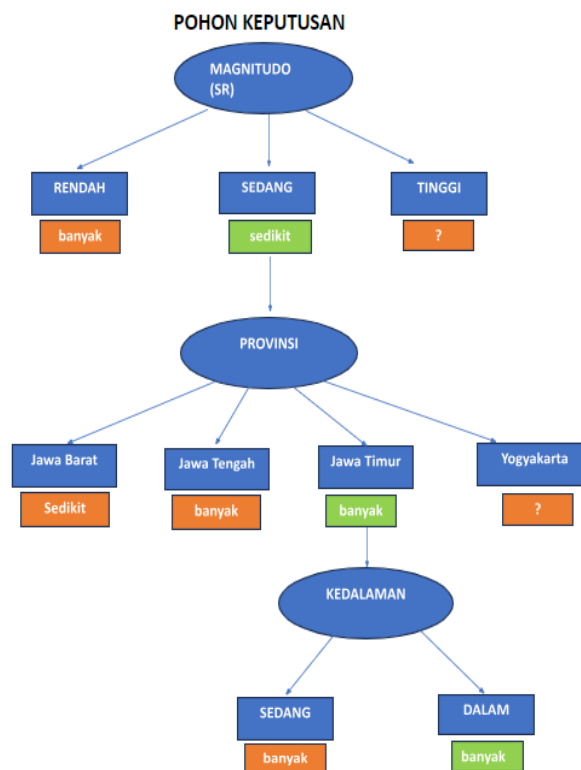
$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)}$$

(5)[12]

Pilih atribut dengan nilai Gain Ratio tertinggi sebagai atribut terbaik untuk splitting.

Proses pembuatan node dalam algoritma pohon keputusan C4.5 melibatkan pengambilan keputusan pada setiap node berdasarkan atribut terbaik, secara rekursif membagi data menjadi subset untuk setiap nilai atribut terbaik, dan berhenti ketika kondisi tertentu terpenuhi, seperti semua instance dalam sebuah node memiliki kelas yang sama atau tidak ada lagi atribut yang tersedia untuk dipisah guna mencegah *overfitting* atau *cropping* berdasarkan kriteria yang telah ditentukan (Irawan & Hermawan Hasibuan, 2020). Proses berulang ini menghasilkan konstruksi pohon keputusan yang dapat digunakan untuk tugas klasifikasi atau prediksi pada data baru.

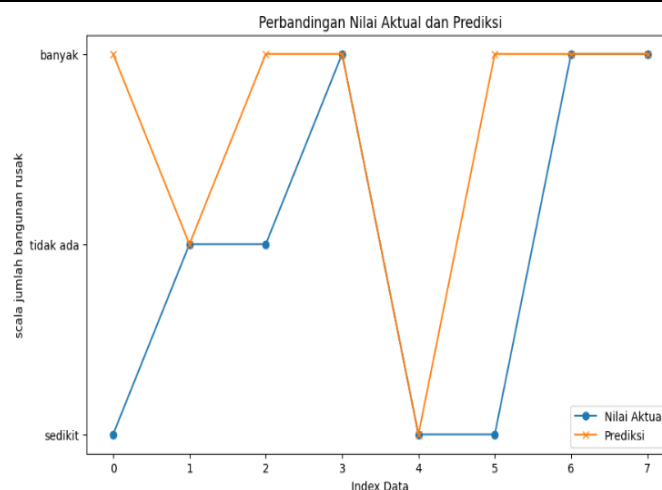
Kesimpulannya, algoritma pohon keputusan C4.5 memainkan peran penting dalam membangun pohon keputusan dengan memilih atribut terbaik di setiap node, membagi data secara rekursif, dan berhenti berdasarkan kondisi tertentu. Pemanfaatan entropi sebagai fungsi biaya membedakannya dari algoritma lain dan berkontribusi terhadap efektivitasnya dalam proses pengambilan keputusan. Pohon keputusan menunjukkan bagaimana C4.5 memilih atribut di setiap langkah, seperti yang terlihat pada Gambar 4. Pohon keputusan tersebut memperlihatkan proses pemilihan atribut, pembagian data, dan kondisi berhenti yang diterapkan untuk mencapai hasil akhir yang optimal.



Gambar 4. Pohon keputusan model terbaik

Berdasarkan Gambar 4 dapat disimpulkan bahwa atribut kekuatan gempa, yang diukur dengan magnitudo, merupakan faktor utama yang mempengaruhi dampak gempa yang terjadi di wilayah Pulau Jawa.

Pada sistem prediksi skala kerusakan bangunan terhadap bencana gempa bumi telah dilakukan penerapan model C4.5 dengan atribut tanggal kejadian, kedalaman, provinsi, dan jumlah rumah serta fasilitas umum yang rusak. Data yang diperoleh melalui berbagai tahapan preprocessing adalah 32 catatan bencana. Hasil prediksi dapat dilihat pada gambar 5.



Gambar 5. Grafik Hasil Prediksi

Dari 8 data testing yang digunakan, model berhasil memprediksi dengan benar pada 5 data dan salah pada 3 data lainnya, mencapai tingkat akurasi sebesar 62%. Selain itu, nilai precision model yang tercatat sebesar 0.83 menunjukkan kemampuan model dalam mengidentifikasi secara tepat, sedangkan recall sebesar 0.61 menunjukkan kemampuan model dalam menangkap sebagian besar kasus positif yang sebenarnya.

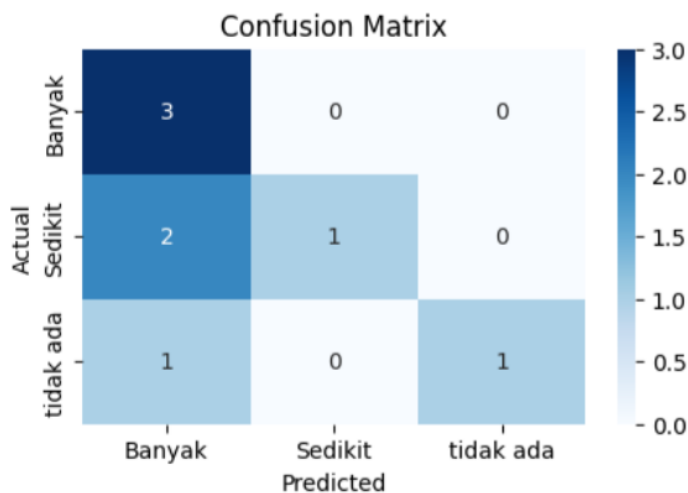
#### 4. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan dataset berupa data dampak bencana dan hasil prediksi skala kerusakan akibat gempa bumi. Penelitian dilakukan dengan model C4.5 berdasarkan dataset sebanyak 62 yang diperoleh dari catatan kejadian gempabumi BMKG dan dampak kerusakan yang tercatat di wilayah Jawa pada portal BNPB. Model yang dikembangkan mencapai tingkat akurasi sebesar 62%, pada tuning 8:2, yang menunjukkan kemampuannya dalam memprediksi pola kerusakan yang terjadi. Namun, hasil ini juga mengungkap beberapa tantangan yang dihadapi dalam pengembangan model, seperti variasi yang signifikan dalam kondisi geografis, struktur bangunan di wilayah yang luas, ketidak konsistenan titik lokasi, dan jumlah kejadian dalam hari yang sama sangat berpengaruh terhadap jumlah kerusakan akibat dampak gempa bumi.

Tantangan dalam pengembangan model prediksi kerusakan akibat gempa bumi di Pulau Jawa mencakup beberapa faktor utama. Perbedaan kondisi geografis dan jenis struktur bangunan di wilayah yang luas dapat mempengaruhi hasil prediksi, dimana gempa dengan magnitudo yang sama dapat menyebabkan tingkat kerusakan yang berbeda tergantung pada struktur tanah dan kualitas konstruksi bangunan. Ketidakonsistenan dalam titik lokasi kejadian gempa, terutama jika terjadi berdekatan atau dalam waktu singkat, juga mempersulit penentuan dampak dari setiap kejadian individu, menambah kompleksitas dalam pengumpulan data kerusakan (Sinaga et al., 2023). Selain itu, variabilitas dalam jumlah gempa yang terjadi pada hari yang sama memperumit penilaian kerusakan, karena model harus mampu menangani situasi di mana beberapa gempa dapat terjadi hampir bersamaan, yang bisa mempengaruhi tingkat kerusakan secara kumulatif.

Penelitian sebelumnya telah mengeksplorasi penggunaan algoritma pembelajaran mesin, khususnya pohon keputusan dan algoritma C4.5, untuk prediksi kerusakan akibat gempa. (Sinaga et al., 2023). Penelitian mengembangkan aplikasi web menggunakan C4.5 untuk memprediksi dampak gempa bumi di Indonesia berdasarkan Skala Intensitas Gempa BMKG dengan perolehan accuracy mencapai 80% , *precision* = 78.41% dan *recall* = 75% dengan model terbaik pada split data 9:1. Penelitian lain, (Tantyoko et al., 2023), membandingkan algoritma Random Forest dan Feature Selection untuk memprediksi tingkat kerusakan bangunan pasca gempa mencapai akurasi hingga 70,83%. Studi-studi ini menunjukkan potensi teknik pembelajaran mesin, khususnya C4.5 dan metode berbasis pohon keputusan lainnya, dalam meningkatkan akurasi prediksi kerusakan gempa bumi dan menyediakan alat berharga bagi upaya mitigasi bencana alam.

Penulis kemudian menggunakan fitur "*Matrix Confusion*" untuk menunjukkan hasil pengujiannya secara keseluruhan, hasilnya sudah baik pada uji coba teknik C.45 seperti ditunjukkan pada Gambar 6.



Gambar 6. *Confusion Matrix* uji coba dengan C.45

Evaluasi kinerja model dilakukan menggunakan *Confusion Matrix*, yang memberikan gambaran tentang kemampuan model dalam mengklasifikasikan data ke dalam kategori yang benar. Hasil evaluasi menunjukkan bahwa model berhasil memprediksi dengan benar 3 dari 6 kasus sebagai "Banyak," mengindikasikan kemampuan yang baik dalam mengidentifikasi kerusakan yang signifikan. Namun, untuk kategori "Sedikit," model hanya memprediksi dengan benar 1 dari 3 kasus, sementara sisanya salah diklasifikasikan sebagai "Banyak," menunjukkan adanya kesalahan dalam menangani kerusakan sedang. Selain itu, dari 2 kasus yang sebenarnya tidak ada kerusakan, model hanya memprediksi dengan benar 1 kasus, dengan satu kesalahan prediksi yang masuk dalam kategori "Banyak." *Confusion Matrix* yang ditunjukkan pada Gambar 6 di atas mengungkap bahwa model mengalami kesulitan dalam membedakan antara kategori "Sedikit" dan "Tidak ada," yang dapat menunjukkan adanya potensi *overfitting* pada kategori dengan data yang lebih banyak atau lebih representative (Mangalathu et al., 2020).

Rekomendasi untuk pengembangan lebih lanjut mencakup beberapa aspek penting untuk meningkatkan akurasi dan efektivitas model prediksi kerusakan akibat gempa bumi. Kompleksitas kerusakan yang bervariasi dari satu gempa ke gempa lainnya mempengaruhi tingkat akurasi model. Perbedaan ini dapat disebabkan oleh berbagai faktor, termasuk kondisi geologi, kualitas infrastruktur, dan kepadatan penduduk. Penambahan data dapat membantu model mengenali pola yang lebih kompleks dan meningkatkan akurasi prediksi. Selain itu, mempertimbangkan faktor-faktor tambahan tersebut dapat membantu dalam meningkatkan ketepatan prediksi. Penggunaan teknik pemrosesan data yang lebih canggih, seperti normalisasi data dan fitur engineering, juga diharapkan dapat mengatasi variasi dalam data dan meningkatkan kinerja model.

Meskipun ada keterbatasan, seperti ukuran dataset yang terbatas, penelitian ini memberikan kontribusi penting dalam pemahaman terhadap potensi kerusakan akibat gempa bumi di wilayah Jawa. Model prediksi yang dikembangkan dapat digunakan oleh lembaga penanggulangan bencana, seperti BMKG dan BNPB, untuk meningkatkan sistem peringatan dini dan strategi mitigasi. Dengan demikian, diharapkan dapat meminimalkan kerugian dan korban jiwa dalam kejadian gempa bumi di masa mendatang, serta memperkuat upaya-upaya mitigasi bencana dan respons yang lebih efektif. Pengembangan lebih lanjut dengan menambahkan data yang lebih luas dan mempertimbangkan faktor-faktor tambahan diharapkan dapat meningkatkan keakuratan model prediksi.

## 5. PENUTUP

Penggunaan model C4.5 dalam memprediksi dampak kerusakan akibat gempa bumi dapat memberikan wawasan yang berharga bagi upaya mitigasi dan respons bencana di wilayah Jawa. Meskipun akurasi prediksi sebesar 62% kurang maksimal, dikarenakan keterbatasan data real di semester pertama tahun 2022, dengan beberapa improvisasi skema dan optimasi model penelitian ini telah memberikan kontribusi dataset dan dapat menjadi landasan yang baik untuk pengembangan model prediksi lebih lanjut. Pengembangan melalui penambahan data, peningkatan fitur, atau penerapan teknik pemrosesan data yang lebih canggih dapat meningkatkan performa model untuk prediksi yang lebih akurat di masa depan.

## DAFTAR PUSTAKA

- A. Gaba, A. J. R. S. Y. A. and M. M., RV College of Engineering. Department of Computer Science & Engineering, RV College of Engineering. Department of Information Science & Engineering, RV College of Engineering. Department of Master of Computer Applications, Florida International University, Institute of Electrical and Electronics Engineers, Institute of Electrical and Electronics Engineers. Bangalore Section, & IEEE Communications Society. Bangalore Chapter. (2019). *Analysis and Prediction of Earthquake Impact: A Machine Learning approach*. Cheng, Z., Peng, C., & Chen, M. (2023). Real-Time Seismic Intensity Measurements Prediction for Earthquake Early Warning: A Systematic Literature Review. In *Sensors* (Vol. 23, Issue 11). MDPI. <https://doi.org/10.3390/s23115052>
- Ding, J., Zhang, J., Zhan, Z., Tang, X., & Wang, X. (2022). A Precision Efficient Method for Collapsed Building Detection in Post-Earthquake UAV Images Based on the Improved NMS Algorithm and Faster R-CNN. *Remote Sensing*, 14(3). <https://doi.org/10.3390/rs14030663>
- Ghimire, S., Guéguen, P., Giffard-Roisin, S., & Schorlemmer, D. (2022). Testing machine learning models for seismic damage prediction at a regional scale using building-damage dataset compiled after the 2015 Gorkha Nepal earthquake. *Earthquake Spectra*, 38(4), 2970–2993. <https://doi.org/10.1177/87552930221106495>
- Irawan, L., & Hermawan Hasibuan, L. (2020). ANALISA PREDIKSI EFEK KERUSAKAN GEMPA DARI MAGNITUDO (SKALA RICHTER) DENGAN METODE ALGORITMA ID3 MENGGUNAKAN APLIKASI DATA MINING ORANGE. *Jurnal Teknologi Informasi : Jurnal Keilmuan Dan Aplikasi Bidang Teknik Informatika*, 14(2), 189–201. <https://doi.org/10.47111/JTI>
- K. Chaurasia, S. K. A. Y. V. K. S. B. S. R. D. (2019). *Predicting Damage to Buildings Caused by Earthquakes Using Machine Learning Techniques*. IEEE.
- Kubo, H., Naoi, M., & Kano, M. (2024). Recent advances in earthquake seismology using machine learning. *Earth, Planets and Space*, 76(1). <https://doi.org/10.1186/s40623-024-01982-0>
- Mangalathu, S., Sun, H., Nweke, C. C., Yi, Z., & Burton, H. V. (2020). Classifying earthquake damage to buildings using machine learning. *Earthquake Spectra*, 36(1), 183–208. <https://doi.org/10.1177/8755293019878137>
- Miura, H., Aridome, T., & Matsuoka, M. (2020). Deep learning-based identification of collapsed, non-collapsed and blue tarp-covered buildings from post-disaster aerial images. *Remote Sensing*, 12(12). <https://doi.org/10.3390/rs12121924>

- Mousavi, S. M., & Beroza, G. C. (2020). A Machine-Learning Approach for Earthquake Magnitude Estimation. *Geophysical Research Letters*, 47(1). <https://doi.org/10.1029/2019GL085976>
- Oktarina, R., Bahagia, S. N., Diawati, L., & Pribadi, K. S. (2020). Artificial neural network for predicting earthquake casualties and damages in Indonesia. *IOP Conference Series: Earth and Environmental Science*, 426(1). <https://doi.org/10.1088/1755-1315/426/1/012156>
- Sinaga, D. P. S., Rini Marwati, & Bambang Avip Priatna Martadiputra. (2023). Aplikasi Web Prediksi Dampak Gempa di Indonesia Menggunakan Metode Decision Tree dengan Algoritma C4.5. *JMT : Jurnal Matematika Dan Terapan*, 5(2), 97–108. <https://doi.org/10.21009/jmt.5.2.5>
- SRIKANTH, Mr. PARUCHURU., SIVA SAI, Mr. N. H., & GOKULKRISHNAN, Mr. S. (2022). ANALYSIS AND PREDICTION OF EARTHQUAKE IMPACT-A MACHINE LEARNING APPROACH. *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, 06(03). <https://doi.org/10.55041/ijrsrem12008>
- Tantyoko, H., Kartika Sari, D., & Wijaya, A. R. (2023). PREDIKSI POTENSIAL GEMPA BUMI INDONESIA MENGGUNAKAN METODE RANDOM FOREST DAN FEATURE SELECTION. In *Idealis: Indonesia Journal Information System* (Vol. 6, Issue 2). <http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/indexHenriTantyoko><http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index>
- Wahyu Pratama, R., Herry Chrisnanto, Y., Abdillah Informatika, G., Jenderal Achmad Yani Cimahi Jl Terusan Jend Sudirman, U., Cimahi Sel, K., Cimahi, K., & Barat, J. (2024). KLASIFIKASI EFEK KERUSAKAN GEMPA BUMI BERDASARKAN SKALA MODIFIED MERCALLI INTENSITY MENGGUNAKAN ALGORITMA MULTICLASS SUPPORT VECTOR MACHINE. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 2). <https://www.bmkg.go.id/gempabumi/skala->