

## KLASIFIKASI DATA GEMPA BUMI DI PULAU JAWA MENGGUNAKAN ALGORITMA EXTREME GRADIENT BOOSTING

Adam Kharis Pratama, Herdi Ashaury, Fajri Rakhmat Umbara

Informatika, Universitas Jenderal Achmad Yani Cimahi

Jl. Terusan Jend. Sudirman, Cibeber, Kec. Cimahi Sel., Kota Cimahi, Jawa Barat 40531

adam.kharis@student.unjani.ac.id

### ABSTRAK

Gempa bumi adalah fenomena yang terjadi akibat pergerakan lempeng tektonik dan dapat memiliki dampak merusak. Kedalaman hiposentrum memiliki peran penting dalam menentukan karakteristik gempa. Proses klasifikasi kedalaman hiposentrum penting untuk memahami potensi gempa di wilayah Indonesia. Dalam menghadapi data gempa yang besar dan kompleks, serta tantangan dalam mengidentifikasi pola dari kedalaman hiposentrum, terdapat permasalahan dalam proses klasifikasi. Penelitian ini memfokuskan pada penerapan algoritma Extreme Gradient Boosting (XGBoost) dalam mengklasifikasikan data gempa bumi berdasarkan kedalaman hiposentrum di Pulau Jawa. Beberapa penelitian sebelumnya menggunakan XGBoost untuk klasifikasi menunjukkan kemampuan algoritma ini dalam berbagai konteks. Hasil eksperimen menunjukkan bahwa model XGBoost dengan penyetelan parameter menghasilkan akurasi 99.58%. Keputusan ini didukung oleh kemampuan model untuk mengatasi masalah ketidakseimbangan data dan menghasilkan prediksi akurat. Meskipun akurasi sempurna belum tercapai, hasil ini realistis dan memiliki potensi untuk mengklasifikasikan data gempa dengan akurasi yang tinggi.

**Kata kunci:** Gempa Bumi, Extreme Gradient, Hiposentrum, Klasifikasi

### 1. PENDAHULUAN

Indonesia merupakan wilayah yang rentan akan bencana alam seperti gempa bumi dan tsunami [1]. Bencana gempa sering terjadi di Indonesia karena negara Indonesia termasuk ke dalam lokasi Ring of Fire [2]. Bahkan dari data gempa yang telah terjadi didunia 90% terjadi di kawasan tersebut [3]. Dari seluruh wilayah di Indonesia, Pulau Jawa merupakan salah satu wilayah di Indonesia yang rentan terhadap gempa bumi. Hal ini disebabkan oleh letak geografis Pulau Jawa yang berada di dekat pertemuan antara lempeng Indo-Australia dan lempeng Eurasia. Kondisi ini menyebabkan adanya aktivitas sesar aktif di daerah tersebut, sehingga sering terjadi gempa bumi [4].

Gempa bumi merupakan proses terjadinya pergeseran lapisan tanah pada permukaan bumi akibat pergeseran lempeng tektonik. Gempa bumi yang disebabkan oleh aktivitas pergerakan lempeng tektonik disebut gempa bumi tektonik [5]. Dalam menentukan karakteristik gempa bumi, Hiposentrum memiliki peran penting dalam menentukan besaran gempa. Jarak hiposentrum ke permukaan bumi dapat mempengaruhi tingkat kekuatan gempa yang dirasakan. Gempa bumi dengan hiposentrum yang dangkal, yaitu dekat dengan permukaan bumi, cenderung memiliki efek yang lebih merusak daripada gempa bumi dengan hiposentrum yang dalam. Hiposentrum adalah pusat gempa yang merupakan sumber gempa bumi, terletak pada kedalaman tertentu di dalam lapisan kerak bumi. Kekuatan suatu gempa dipengaruhi oleh lokasi hiposentrumnya, yang bisa terletak pada kedalaman yang beragam. Gempa bumi dapat dikelompokkan menjadi tiga jenis berdasarkan kedalaman hiposentrumnya, yaitu gempa dalam, gempa menengah, dan gempa dangkal [6].

Klasifikasi kekuatan gempa berdasarkan kedalaman hiposentrum diperlukan untuk mengetahui potensi kekuatan gempa yang tersebar di wilayah Indonesia. Hasil dari proses klasifikasi seringkali mengalami masalah yaitu ketidaktepatan dalam klasifikasi [7]. Klasifikasi adalah proses pengelompokan data atau objek ke dalam kelas atau kategori yang berbeda berdasarkan atribut tertentu. Dalam machine learning, klasifikasi digunakan untuk mengklasifikasikan input baru berdasarkan data latih yang telah diklasifikasikan sebelumnya [8]. Dalam proses klasifikasi, data mining memperhatikan fitur-fitur yang relevan dari setiap gempa bumi dan mencari pola atau kombinasi fitur yang membedakan jenis gempa. Hal ini memungkinkan untuk membangun model yang dapat mengenali pola-pola tersebut dalam data gempa bumi baru dan mengklasifikasikannya dengan akurasi yang tinggi [9].

Klasifikasi dapat digunakan dengan algoritma klasifikasi seperti Naïve Bayes, Decision Tree atau Support Vector Machine, dari beberapa algoritma tersebut terdapat algoritma terbaru dari pengembangan Decision Tree yang merupakan hasil evolusi dari Gradient Boosting, itu adalah Extreme Gradient Boosting (XGBoost). Algoritma ini mengenali pola-pola yang mungkin tidak dapat ditangkap oleh satu pohon saja tapi dengan menggabungkan prediksi dari banyak pohon lemah, XGBoost menciptakan model yang kuat dan mampu mengatasi masalah overfitting. Dalam proses ini, XGBoost menggunakan teknik regularisasi dan mekanisme learning rate yang dapat diatur untuk mengontrol seberapa besar setiap iterasi memengaruhi hasil akhir. Melalui langkah-langkah ini, XGBoost secara adaptif dan iteratif membangun model yang lebih baik dari waktu ke waktu,

menghasilkan prediksi akhir yang akurat dan handal untuk berbagai tugas klasifikasi dan regresi [10].

Beberapa penelitian yang menggunakan XGBoost untuk klasifikasi telah dilakukan, antara lain Penelitian mengenai Klasifikasi Malware menggunakan XGBoost, dalam penelitian ini model dioptimalkan dengan penghapusan fitur yang noise serta dengan memanfaatkan fungsi penting fitur penyetelan XGBoost dan hyperparameter dalam memperoleh akurasi sebesar 97% [11]. Lalu pada penelitian selanjutnya mengenai analisis performa XGBoost dalam mengatasi data tidakimbang pada data credit card dengan menghasilkan akurasi sebesar 77,6%, pada penelitian ini dilakukan perbandingan akurasi menggunakan teknik oversampling untuk menangani data tidakimbang. Hasilnya, akurasi tidak jauh berbeda performanya dalam penerapan teknik oversampling menggunakan Synthetic Minority Oversampling Technique (SMOTE) dengan akurasi 78,73% dan Adaptive Synthetic (ADASYN) 77,84%. Hal ini dikarenakan algoritma XGBoost didasarkan pada tree boosting, ini menghindari overfitting data yang tidak seimbang [12].

Dari permasalahan dan penjelasan yang telah disebutkan di atas maka dalam penelitian ini akan menerapkan algoritma Extreme Gradient Boosting (XGBoost) Pada Klasifikasi Data Gempa Bumi di Pulau Jawa berdasarkan hiposentrumnya. Fokus Dalam penelitian ini hanya pada data gempa bumi di Pulau Jawa.

## 2. TINJAUAN PUSTAKA

### 2.1. Data Mining

Data mining adalah suatu proses yang bertujuan untuk menemukan pola-pola yang menarik dan memperoleh pengetahuan baru dari sejumlah besar data yang tersedia. Sumber data yang digunakan dalam proses ini dapat berupa database, data warehouse, situs web, atau repositori informasi lainnya. Data mining adalah proses menggunakan algoritma, model statistik, dan metode lain untuk mengekstraksi informasi dari kumpulan data yang banyak (Big Data). Ini melibatkan penemuan pola, hubungan, dan wawasan dalam data yang dapat digunakan untuk membuat prediksi, mengidentifikasi tren, dan mendukung pengambilan keputusan. Data mining sering menggunakan teknik dari pembelajaran mesin, statistik, dan manajemen basis data untuk menganalisis data [13].

Data mining merupakan komponen penting dalam proses Knowledge Discovery in Database (KDD) yang bertujuan untuk menggali pola atau model dari data menggunakan algoritma khusus. Data mining memiliki peranan krusial dalam proses knowledge discovery in database terutama dalam hal ekstraksi dan perhitungan data yang sedang dianalisis. Proses KDD melibatkan penggunaan database dengan volume besar yang mencakup pemilihan data, preprocessing (subsampling), transformasi data, penerapan metode data mining (algoritma) untuk menghitung pola data, dan evaluasi pola yang

dihasilkan dari proses data mining. Hasil dari proses ini diharapkan dapat menjadi informasi atau pengetahuan baru yang bermanfaat [14].

### 2.2. Klasifikasi

Klasifikasi adalah proses pengelompokan data atau objek baru diklasifikasikan berdasarkan data latih yang telah diklasifikasikan sebelumnya. Klasifikasi digunakan dalam berbagai bidang, termasuk pemasaran, kredit, pengelolaan risiko, dan pengambilan keputusan. Dalam proses klasifikasi, data akan dikelompokkan ke dalam kelas atau kategori yang berbeda berdasarkan atribut tertentu [15].

### 2.3. Gempa Bumi

Gempa bumi adalah salah satu jenis bencana geologis yang terjadi akibat gaya-gaya yang berasal dari dalam bumi. Gempa bumi merupakan fenomena alam yang muncul akibat pelepasan energi yang tiba-tiba dari dalam bumi, yang ditandai dengan terjadinya patahan pada lapisan batuan di kerak bumi. Energi yang mengakibatkan terjadinya gempa bumi ini berasal dari pergerakan lempeng tektonik. Energi yang terkumpul kemudian dilepaskan ke segala arah dalam bentuk gelombang gempa bumi, dan dampaknya dapat dirasakan hingga ke permukaan bumi [16].

### 2.4. Hiposentrum

Hiposentrum adalah lokasi sebenarnya di mana gempa bumi terjadi, dan juga merupakan sumber dari gelombang tubuh gempa. Hiposentrum gempa bumi dapat berada puluhan hingga ratusan kilometer di bawah permukaan. Kekuatan gempa bumi bergantung pada sejauh mana tekanan menumpuk pada area yang menonjol sebelum akhirnya pecah. Jika area tersebut pecah atau mengalami perubahan bentuk sebelum tegangan yang signifikan terakumulasi, maka gempa bumi yang terjadi tidak akan signifikan. Gempa bumi dibagi menjadi 3 golongan berdasarkan kedalamannya:

1. Gempa bumi dalam. Hiposentrum gempa bumi dalam lebih dari 300 km di bawah permukaan bumi (terjadi di dalam kerak bumi). Gempa bumi ini biasanya tidak terlalu membahayakan.
2. Gempa bumi menengah. Hiposentrum gempa bumi menengah berada diantara 60 km hingga 300 km di bawah permukaan bumi. Getaran yang diakibatkan gempa bumi golongan ini lebih terasa dibandingkan gempa bumi dalam dan dapat mengakibatkan kerusakan ringan.
3. Gempa bumi dangkal. Hiposentrum dari gempa bumi dangkal kurang dari 60 km dari permukaan bumi. Gempa bumi ini pada umumnya menyebabkan kerusakan besar [17].

### 2.5. Extreme Gradient Boosting

XGBoost adalah sebuah algoritma ansambel *machine-learning* yang menggunakan *decision tree* dan framework *Gradient Boosting*. XGBoost adalah peningkatan pada algoritma GBM. Perbedaan

utamanya adalah XGBoost menggunakan model yang lebih teratur, yang membantu mencegah overfitting. XGBoost menggunakan teknik boosting yang dapat menangani data yang besar dan tidak linear serta menyediakan fitur-fitur yang memungkinkan untuk mengatasi overfitting, seperti regularisasi dan pemilihan fitur yang optimal. Pada algoritma ini, decision tree dibuat dalam bentuk berurutan. Bobot memainkan peran penting dalam XGBoost. Bobot diberikan ke semua variabel independen yang kemudian dimasukkan ke dalam pohon keputusan yang memprediksi hasil. Bobot variabel yang diprediksi salah oleh pohon dinaikkan dan variabel ini kemudian dimasukkan ke decision tree kedua. Pengklasifikasi/prediktor individu ini kemudian disatukan untuk memberikan model yang kuat dan lebih tepat.

## 2.6. Evaluasi Model

Kinerja model dapat diukur dengan membandingkan hasil prediksi dengan data actual menggunakan confusion matrix. Confusion matrix adalah sebuah tabel yang digunakan untuk mengevaluasi performa model dalam tugas klasifikasi. Tabel ini berfungsi untuk membandingkan prediksi model dengan label actual dari data uji yang digunakan. Confusion matrix memiliki empat sel utama yaitu True Positive (TP), True Negative (TN), False Positive (FP) dan False Negative (FN). Untuk mengetahui kinerja model dapat dilakukan dengan menggunakan beberapa metrik seperti accuracy, precision, recall, F1-score [18].

### 1. Accuracy

Accuracy adalah metrik evaluasi yang paling umum digunakan untuk klasifikasi

$$Accuracy = \frac{(TP + TN)}{Total\ sample} \times 100$$

### 2. Precision

Precision mengindikasikan sejauh mana hasil prediksi model sesuai dengan data yang diminta. Rumus untuk menghitung Precision:

$$Precision = \frac{TP}{(TP + FP)} \times 100\%$$

### 3. Recall digunakan untuk mengukur sejauh mana sistem dapat menghasilkan nilai yang benar dengan target positif. Rumus untuk menghitung Recall:

$$Precision = \frac{TP}{(TP + FN)} \times 100\%$$

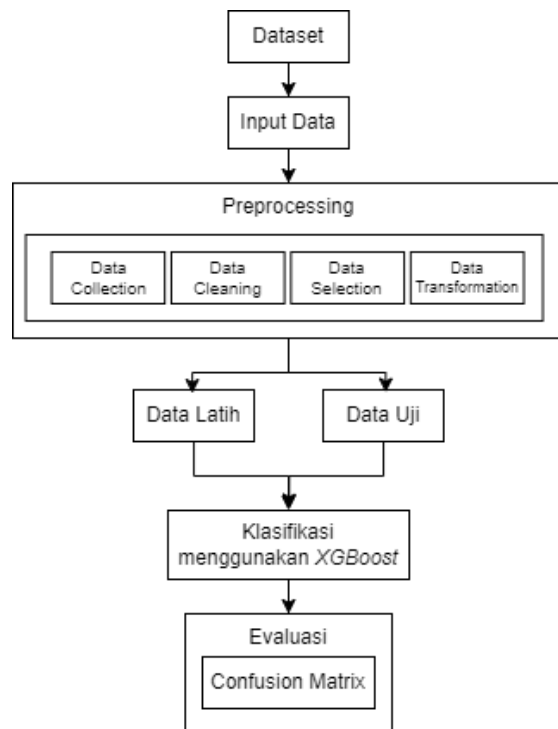
### 4. F1-Score

F1-Score dapat digunakan sebagai matrix evaluasi yang menggabungkan kedua nilai Precision dan Recall. F1-Score merupakan alternatif yang digunakan untuk mengukur akurasi sistem.

$$F - 1 = 2 \frac{(Precision \cdot Recall)}{(Precision + Recall)}$$

## 3. METODE PENELITIAN

Berikut gambaran tahapan penelitian yang dilakukan



Gambar 1. Tahapan Penelitian

### 3.1. Dataset

Pengambilan data Hiposentrum Gempa Bumi diambil dari Repositori BMKG yang dapat diakses di <https://repogempa.bmkg.go.id>, dengan 12 atribut yaitu Event ID, Origin Time, Latitude, Longitude, Magnitude, Pusat Gempa, Depth, MMI, Kota, Provinsi, dan Hiposentrum. Berikut Tabel 1 menunjukkan Atribut Data Gempa Bumi.

Tabel 1. Atribut Data Gempa Bumi

Atribut	Keterangan
No	Nomor urutan kejadian gempa bumi dari setiap awal bulan.
Event ID	Memuat ID untuk setiap event gempa yang terjadi
Origin Time	Memuat informasi mengenai kapan terjadinya gempa bumi dalam format waktu.
Latitude	Memuat informasi lokasi terjadinya gempa.
Longitude	Memuat informasi lokasi terjadinya gempa.
Magnitude	Memuat informasi mengenai besaran gempa
Pusat Gempa	Memuat informasi tentang dimana pusat terjadinya gempa dan jarak lokasi terkena gempa dengan pusat gempa.
Depth	Memuat informasi mengenai kedalaman gempa yang terjadi
MMI	Memuat informasi mengenai intensitas gempa yang terjadi dalam suatu daerah.

Atribut	Keterangan
Kota	Memuat informasi mengenai lokasi kota terjadinya gempa bumi
Provinsi	Memuat informasi mengenai lokasi provinsi terjadinya gempa bumi
Hiposentrum	Memuat informasi mengenai jarak hiposentrum.

### 3.2. Preprocessing

Proses *preprocessing* data melibatkan beberapa tahap penting untuk mempersiapkan data sebelum digunakan dalam analisis atau pemodelan. *Preprocessing* data membantu mengatasi masalah seperti nilai yang hilang, ketidakseimbangan skala fitur, atau variasi dalam tipe data. Dengan melalui langkah-langkah ini, data menjadi lebih siap untuk digunakan dalam berbagai proyek analisis data atau machine learning.

### 3.3. Pembagian Data

Pembagian data adalah salah satu langkah penting dalam pemodelan klasifikasi (dan juga dalam tugas analisis data). Tujuannya adalah untuk memisahkan data menjadi subset yang berbeda untuk melatih model, menguji model, dan mengukur kinerja model. Dengan cara ini, Anda dapat mengukur seberapa baik model Anda dapat melakukan prediksi pada data yang belum pernah dilihat sebelumnya, yang juga disebut sebagai kemampuan generalisasi model.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Preprocessing

Preprocessing yang dilakukan dalam penelitian ada beberapa tahapan diantaranya data collection, data cleaning, data selection dan data transformation.

### 4.2. Data Collection

Data yang diperoleh pada penelitian ini memiliki batasan jumlah kasus gempa yang hanya dapat diambil dalam kurun waktu per 1 bulan. Dalam kurun waktu tersebut data yang terkumpul kurang dari 1000 data, oleh karena itu dilakukan pengumpulan data tiap bulan dari tahun 2020-2022 dan menggabungkannya sehingga memperoleh data dengan total 1604 record data. Data yang terdiri dari 1604 data memiliki tiga label kelas Hiposentrum yaitu Dangkal, Sedang dan Dalam. Untuk setiap kelas terdapat kelas Dangkal dengan jumlah data 1437, kelas Sedang dengan Jumlah data 154 dan Kelas Dalam dengan jumlah data 12 dapat dilihat pada table.

Tabel 2. Distribusi Kelas Dataset Gempa Bumi

Nama Kelas	Jumlah Data
Dangkal	1437
Sedang	154
Dalam	12

### 4.3. Data Cleaning

Proses ini adalah menghapus data yang tidak valid, duplikat atau tidak relevan. Dari data yang telah

diperoleh tidak mengandung data duplikat ataupun tidak relevan dengan data lainnya.

### 4.4. Data Selection

Dalam tahapan data selection, seleksi data yang dilakukan adalah penghapusan atribut-atribut yang tidak digunakan dalam proses klasifikasi seperti No, Event ID, Origin Time, Kota dan Provinsi. Atribut-atribut tersebut tidak ada kaitannya dengan klasifikasi Hiposentrum, atribut yang nilainya hanya satu atau tidak beragam, dan atribut yang kosong sehingga atribut tersebut tidak diperlukan pada penelitian ini.

### 4.5. Data Transformation

Dalam data ini terdapat atribut yang perlu diubah tipe data dan salah satu atribut juga harus melakukan proses pemisahan atribut karena memiliki informasi yang dapat dipisahkan. Beberapa atribut yang akan ditransformasikan pada Tabel 3.

Tabel 3. Atribut Pusat Gempa Bumi

Pusat Gempa
Pusat gempa berada di darat 8 km tenggara Kairatu
Pusat gempa berada di laut 9 km Tenggara Kairatu-SBB
Pusat gempa berada di laut 68 km BaratDaya Pangandaran
Pusat gempa berada di Laut 165 km Tenggara Wakatobi
Pusat gempa berada di laut 85 km TimurLaut Daruba

Atribut ini akan dipisahkan menjadi 2 atribut pada Tabel 4.

Tabel 4. Pemisahan Atribut Pusat Gempa

Pemisahan Atribut	Deskripsi	Tipe Data
Pusat Gempa	Memiliki dua informasi yang akan menunjukkan dimana pusat gempa berada.	Kategori: Laut atau Darat
Jarak dengan pusat gempa	Memiliki informasi mengenai jarak antara lokasi gempa dengan pusat gempa.	Numerik

Setelah perubahan diatas, maka terdapat dataset yang digunakan ditujukan pada Tabel 5.

Tabel 5. Informasi Dataset Gempa Bumi

Data	Column	Dtype
0	Latitude	Float
1	Longitude	Float
2	Magnitude	Float
3	Pusat Gempa	Objek
4	Jarak dengan pusat	Float
5	Kedalaman	Int
6	MMI	Int
7	Hiposentrum	Object

Namun masih terdapat kesalahan tipe data untuk atribut Pusat Gempa dan Hiposentrum, oleh karena itu

akan dilakukan perubahan data dari tipe data object menjadi numerik ditunjukkan pada Tabel 6 dan Tabel 7.

Tabel 6. Perubahan Tipe Data Pusat Gempa

Atribut	Value
Laut	0
Darat	1

Tabel 7. Perubahan Tipe Data Hiposentrum

Atribut	Value
Dalam	0
Dangkal	1
Sedang	2

#### 4.6. Pembagian Data

Pada tahap ini, data penelitian akan dibagi menjadi dua bagian, yaitu data latih dan data uji. Presentase untuk pembagian data latih 70% dan data uji 30% membagi dataset dengan mempertahankan proporsi kelas yang sama. Fungsi dari data latih ini akan di proses untuk model klasifikasi dan data tes akan digunakan untuk menguji model [19]. Berikut adalah pembagian data latih pada Gambar 2 dan pembagian data uji pada Gambar 3.

	Latitude	Longitude	Magnitude	Pusat Gempa	Jarak dengan pusat	Kedalaman	PMI
61	-1.39	134.02	5.3	1	52.0	10	3
431	-3.25	125.97	4.9	0	51.0	10	3
1524	-1.46	127.35	4.8	0	35.0	10	3
665	2.77	128.48	4.9	0	82.0	8	3
92	-8.31	115.99	3.3	0	19.0	10	2
...	...	...	...	...	...	...	...
763	0.60	120.45	4.5	1	39.0	10	4
835	-5.54	105.13	3.8	1	17.0	10	3
1216	-6.92	107.02	2.4	1	9.0	4	2
559	-1.87	136.20	4.2	1	14.0	5	3
684	-2.53	121.35	2.8	1	20.0	10	2

1122 rows x 7 columns

Gambar 2. Data Latih

	Latitude	Longitude	Magnitude	Pusat Gempa	Jarak dengan pusat	Kedalaman	PMI
1074	5.88	95.24	3.0	1	8.0	10	2
326	1.69	99.15	3.4	1	17.0	15	2
1560	-6.96	105.26	4.7	0	48.0	21	3
1368	-2.45	120.92	3.4	1	31.0	8	3
1240	-5.58	121.78	3.6	0	29.0	10	3
...	...	...	...	...	...	...	...
1039	-1.60	99.29	5.1	1	55.0	17	3
1673	4.20	126.94	4.8	0	36.0	53	0
795	-2.88	119.44	3.8	1	31.0	10	3
434	0.21	124.56	4.5	0	61.0	10	2
1213	-3.29	100.25	5.4	0	124.0	10	3

481 rows x 7 columns

Gambar 3. Data Uji

#### 4.7. Model XGBoost

Pembuatan model XGBoost digunakan data training yang telah diubah ke dalam bentuk matriks, maka perlu dimasukkan data trained. Dalam model XGBoost memerlukan beberapa parameter pengukuran yang telah ditentukan setelah melakukan beberapa kali percobaan, sehingga diperoleh hasil prediksi klasifikasi yang baik.

#### 4.8. Parameter XGBoost

Parameter yang digunakan dalam model XGBoost ditunjukkan pada tabel berikut.

Tabel 8. Parameter yang digunakan XGBoost

Parameter	Kegunaan
n_estimators	Jumlah proses yang XGBoost akan coba pelajari
Min_child_weight	Jumlah bobot minimum dari semua pengamatan yang diperlukan dalam simpul anak
Max_Depth	Kedalaman maksimum pohon. Ketika meningkatkan nilai ini akan membuat model lebih kompleks dan lebih cenderung overfit.
Learning_rate	Ini mempengaruhi seberapa cepat model belajar dengan kesalahan residual
gamma	Gamma untuk mengontrol langsung fitur-fitur yang mendominasi dalam kumpulan data.
Colsample_bytree	Membatasi jumlah fitur untuk membangun setiap pohon. ini mempelajari cara mengoptimalkan variabel target menggunakan kumpulan fitur yang berbeda.

Dari tabel diatas menunjukkan beberapa parameter yang akan digunakan dalam model. Setelah menentukan parameter apa saja yang digunakan, selanjutnya diperlukan penyetelan parameter. Dalam penyetelan parameter ini diperlukan penentuan nilai parameter yang tepat.

#### 4.9. Hasil Eksperimen dan Analisis

Setelah dilakukan tahapan penelitian, Pada tahap eksperimen ini akan dilakukan percobaan-percobaan untuk menemukan hasil terbaik berdasarkan data yang dimiliki, data yang digunakan memiliki masalah mengenai kelas yang tidakimbang, perbedaan antar kelas dapat dilihat pada Tabel 9.

Tabel 9. Jumlah Kelas

Kelas	Jumlah
Dangkal	1437
Sedang	154
Dalam	12

Ketidakseimbangan data mengakibatkan situasi di mana jumlah sampel dalam setiap kelas tidak seimbang atau tidak proporsional. Ini dapat memiliki dampak signifikan pada performa model klasifikasi, Oleh karena itu akan dilakukan eksperimen dengan dua kondisi data berbeda yaitu Data Imbalanced dan Data yang sudah menggunakan Teknik oversampling untuk mengetahui perbedaan performa model XGBoost pada data imbalanced.

Pada tahap ini dilakukan beberapa eksperimen untuk mendapatkan hasil yang optimal berdasarkan data yang dimiliki, parameter yang digunakan serta Teknik data imbalanced. Hasil eksperimen

menunjukkan bahwa penanganan data yang tidak seimbang melalui teknik oversampling SMOTE dan penyetelan parameter mampu meningkatkan kinerja model XGBoost. Pada Table 10 berikut.

Tabel 10. Hasil Eksperimen

Parameter	HE1	HE2	HE3	HE4
N_estimators	0	400	0	400
Min_child_weight	0	1	0	1
Max_depth	0	12	0	12
Learning_rate	0.01	0.15	0.01	0.15
Gamma	0	0.0	0	0.0
Colsample_bytree	0	0.7	0	0.7
Oversampling	Tidak	Tidak	Ya	Ya
Akurasi	0.996	0.9958	0.1	0.1

Terdapat akurasi yang tidak jauh berbeda dari setiap eksperimen, tiap hasil eksperimen yang telah dilakukan didasarkan pada penelitian sebelumnya yang telah melakukan hal tersebut. Pada eksperimen setelah dilakukan oversampling didapat hasil akurasi yang sangat baik yaitu 100%. Namun, perlu diingat bahwa akurasi yang sangat tinggi juga perlu dikaitkan dengan kemungkinan overfitting [20].

Untuk hasil eksperimen yang akan digunakan pada penelitian ini adalah Hasil Eksperimen kedua, pemilihan hasil eksperimen ini karena penelitian yang dilakukan berfokus pada penerapan model XGBoost dan bagaimana akurasi yang diperoleh oleh model. Untuk menguji performa model, penelitian ini menggunakan evaluasi model klasifikasi dengan confusion matrix. Confusion matrix yang didapatkan ditujukan pada tabel 11.

Tabel 11. Hasil Confusion Matrix

Kelas	TP	FP	FN	TN
Dangkal	433	0	0	48
Sedang	42	2	0	437
Dalam	4	0	2	475

Skor akurasi yang didapat dari perhitungan confusion matrix adalah 0.9958 berarti model memiliki akurasi sekitar 99.58%, yang merupakan proporsi dari prediksi yang benar dari semua prediksi.

Selain akurasi model yang didapat dari confusion matrix maka dapat menghitung hasil evaluasi kinerja model dari beberapa evaluasi seperti Presisi, Recall dan F1-Score. Dengan matrix evaluasi yang didapatkan maka didapatkan hasil skor masing-masing matrix Precision, Recall, F1 Score pada tabel 12.

Tabel 12. Matrix Evaluasi

Matrix Evaluasi	Skor
Precision	0.9848
Recall	0.8889
F1-Score	0.9256

Precision mengukur sejauh mana model benar dalam memprediksi positif dari semua prediksi positif yang dilakukan. Dalam konteks tiga kelas, ini berarti sejauh mana prediksi model benar dalam

mengidentifikasi masing-masing kelas. Nilai 0.9848 berarti bahwa sekitar 98.48% dari prediksi yang dikategorikan sebagai positif oleh model adalah benar positif. Recall mengukur sejauh mana model benar dalam menemukan semua contoh positif yang ada. Dalam kasus tiga kelas, ini berarti sejauh mana model berhasil menemukan semua instance dari masing-masing kelas. Nilai 0.8889 berarti bahwa model berhasil menemukan sekitar 88.89% dari semua instance positif yang sebenarnya ada. F1-score memberikan gambaran holistik tentang seberapa baik model menggabungkan keduanya. Nilai 0.9256 menunjukkan rata-rata harmonis dari presisi dan recall.

## 5. KESIMPULAN DAN SARAN

Penelitian ini berhasil melakukan penerapan metode XGBoost dalam menganalisis data gempa bumi. Beberapa percobaan dilakukan untuk mengoptimalkan performa model. Percobaan pertama menggunakan teknik oversampling yang menghasilkan akurasi sempurna 100%. Meskipun angka ini menunjukkan performa yang luar biasa, namun harus dilakukan evaluasi lebih lanjut terhadap kemungkinan overfitting atau bias yang tidak terdeteksi. Percobaan selanjutnya melibatkan penyetelan parameter yang dimiliki model XGBoost, dan menghasilkan akurasi sebesar 99.58%. Meskipun tidak mencapai angka akurasi sempurna, hasil ini lebih realistis dan menunjukkan bahwa model memiliki kemampuan umum untuk mengklasifikasikan data gempa bumi dengan akurat. Selain itu, pentingnya akurasi 99.58% sebagai indikasi kehandalan model. Berdasarkan hasil percobaan dan pertimbangan yang matang, penelitian ini memilih eksperimen dengan penyetelan parameter sebagai model akhir. Keputusan ini diambil karena model yang disesuaikan dengan parameter memiliki kemampuan untuk menangani data yang tidak seimbang. Namun, perlu diingat bahwa hasil keseluruhan perlu diverifikasi dengan lebih lanjut melalui pengujian lebih lanjut untuk memastikan bahwa model ini mampu menggeneralisasi dengan baik di luar dataset pelatihan. Algoritma XGBoost memiliki parameter yang perlu dilakukan penyetelan. Oleh karena itu, perlu dilakukan pengecekan lebih lanjut terhadap pengaruh parameter terhadap data yang digunakan. Penggunaan parameter yang sesuai dapat memaksimalkan performa XGBoost. Selain itu, dapat juga dilakukan penelitian lebih lanjut dengan menggunakan algoritma lainnya untuk membandingkan performa klasifikasi yang dihasilkan.

## DAFTAR PUSTAKA

- [1] P. Studi Psikologi, "Meninjau Perilaku Terkait Bencana di Indonesia: Sebuah Kajian Literatur Nyoman Trisna Aryanata, Ni Made Sintya Noviana Utami," *Jurnal Psikologi MANDALA* 2019, vol. 3, no. 1, pp. 69–84.
- [2] D. Henley, H. Schulte Nordholt, and P. Boomgaard, *Environment, trade and society in*

- Southeast Asia : a longue durée perspective.* 2015.
- [3] D. Paton, "Living on the ring of fire: Perspectives on managing natural hazard risk in pacific rim countries," *Journal of Pacific Rim Psychology*, vol. 3, no. 1. Australian Academic Press, pp. 1–3, 2009. doi: 10.1375/prp.3.1.1.
  - [4] S. Galih Marlyono and G. Kamil Pasya, "PERANAN LITERASI INFORMASI BENCANA TERHADAP KESIAPSIAGAAN BENCANA MASYARAKAT JAWA BARAT," 2016.
  - [5] Sunarjo, M. Taufik Gunawan, and Sugeng Pribadi, *Gempa bumi : edisi populer*, vol. 228. Jakarta: Badan Meteorologi Klimatologi dan Geofisika, 2010, 2012.
  - [6] J. Lingkungan Dan Bencana Geologi *et al.*, "Analisis Kegempaan di Zona Sesar Palu Koro, Sulawesi Tengah Seismicity Analysis in Palu Koro Fault Zone, Central Sulawesi," *Jurnal Lingkungan dan Bencana Geologi*, vol. 6, no. 3, pp. 253–264, 2015.
  - [7] Lina Handayani, "Peningkatan aktivitas gempa bumi di Indonesia," 2010.
  - [8] E. H. Yulianti, O. Soesanto, and Y. Sukmawaty, "Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit," *JOMTA Journal of Mathematics: Theory and Applications*, vol. 4, no. 1, 2022.
  - [9] Nilam Novita Sari, "K-AFFINITY PROPAGATION (K-AP) CLUSTERING UNTUK KLASIFIKASI GEMPA BUMI," *DSPACE Universitas Islam Indonesia*, Apr. 2018.
  - [10] R. Ubaidillah, M. Muliadi, D. T. Nugraha, M. R. Faisal, and R. Herteno, "Implementasi XGBoost Pada Keseimbangan Liver Patient Dataset dengan SMOTE dan Hyperparameter Tuning Bayesian Search," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 3, p. 1723, Jul. 2022, doi: 10.30865/mib.v6i3.4146.
  - [11] R. Kumar and S. Geetha, "Malware classification using XGboost-Gradient boosted decision tree," *Advances in Science, Technology and Engineering Systems*, vol. 5, no. 5, pp. 536–549, 2020, doi: 10.25046/AJ050566.
  - [12] C. V. Priscilla and D. P. Prabha, "Influence of optimizing xgboost to handle class imbalance in credit card fraud detection," in *Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020*, Institute of Electrical and Electronics Engineers Inc., Aug. 2020, pp. 1309–1315. doi: 10.1109/ICSSIT48917.2020.9214206.
  - [13] J. Han, M. Kamber, and J. Pei, "Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)," 2011.
  - [14] B. Peterson and P. S. J. D. Baker, "Data Mining for Education."
  - [15] J. Lasama, A. Pradika, and A. Prasetyadi, *Klasifikasi Tsunami Pada Gempa Menggunakan Random Forest Classifier*.
  - [16] M. Bohnhoff, F. Bulut, G. Dresen, P. E. Malin, T. Eken, and M. Aktar, "An earthquake gap south of Istanbul," *Nat Commun*, vol. 4, 2013, doi: 10.1038/ncomms2999.
  - [17] H. Saputra *et al.*, "STUDI ANALISIS PARAMETER GEMPA DAA POLA SEBARAAYA BERDASARKAA DATA MULTIISTATIOO (STUDI KASUS KEJADIAA GEMPA PULAU SULAWESI TAHUU 200002014) Abstract: Study on Earthquake Parameters and Spreading Pattern Based on Multiistation Datas (Case Study on Earthquake Event in Sulawesi Island)," 2000.
  - [18] R. Yacoub Amazon Alexa and D. Axman Amazon Alexa, "Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models."
  - [19] M. Rizky Mubarak, R. Herteno, I. Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lambung Mangkurat Jalan Ahmad Yani Km, and K. Selatan, "HYPER-PARAMETER TUNING PADA XGBOOST UNTUK PREDIKSI KEBERLANGSUNGAN HIDUP PASIEN GAGAL JANTUNG."
  - [20] X. Ying, "An Overview of Overfitting and its Solutions," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Mar. 2019. doi: 10.1088/1742-6596/1168/2/022022.