

# Perbandingan Algoritma *Machine Learning* untuk Prediksi Potensi Tsunami di Pesisir Barat Lampung

DOI: <http://dx.doi.org/10.35889/jutisi.v14i3.3184>

Creative Commons License 4.0 (CC BY – NC)

Kevin Ramses Stifan<sup>1\*</sup>, Chairani Fauzi<sup>2</sup>

Teknik Informatika, Institut Informatika dan Bisnis Darmajaya, Bandar Lampung, Indonesia

\*e-mail Corresponding Author: kevinstif.2111010043@mail.darmajaya.ac.id

## Abstract

*Indonesia has a high level of vulnerability to tsunami disasters due to seismic activity in subduction zones. This condition requires a fast and accurate tsunami potential prediction system to support effective disaster mitigation. This study aims to compare the performance of five machine learning algorithms in predicting tsunami potential along the West Coast of Lampung, namely Random Forest, K-Nearest Neighbor, Naïve Bayes, Neural Network, and Decision Tree. The dataset consists of 352 earthquake records from 2023–2024 obtained from BMKG, using parameters such as magnitude, depth, and epicentral distance. Model evaluation was conducted using a Confusion Matrix with performance metrics including accuracy, precision, recall, and F1-score. The results indicate that the Random Forest algorithm achieved the best performance with an accuracy of 100% and balanced precision, recall, and F1-score values. These findings are expected to support the development of a machine learning-based tsunami early warning system that is adaptive to local geophysical characteristics.*

**Keywords:** Earthquake; Machine Learning; Random Forest; Tsunami prediction

## Abstrak

Indonesia memiliki tingkat kerentanan yang tinggi terhadap bencana tsunami akibat aktivitas gempa bumi di zona subduksi. Kondisi ini menuntut tersedianya sistem prediksi potensi tsunami yang cepat dan akurat untuk mendukung mitigasi bencana. Penelitian ini bertujuan membandingkan kinerja lima algoritma *machine learning* dalam memprediksi potensi tsunami di Pesisir Barat Lampung, yaitu *Random Forest*, *K-Nearest Neighbor*, *Naïve Bayes*, *Neural Network*, dan *Decision Tree*. Data yang digunakan berupa 352 catatan gempa bumi periode 2023–2024 yang bersumber dari BMKG dengan parameter magnitudo, kedalaman, dan jarak episentrum. Evaluasi model dilakukan menggunakan *Confusion Matrix* dengan metrik akurasi, *presisi*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* memberikan performa terbaik dengan tingkat akurasi 100% serta keseimbangan nilai *presisi*, *recall*, dan *F1-score*. Temuan ini diharapkan dapat mendukung pengembangan sistem peringatan dini tsunami berbasis *machine learning* yang adaptif terhadap karakteristik geofisika lokal.

**Kata kunci:** Gempa bumi; Machine Learning; Random Forest; Prediksi tsunami.

## 1. Pendahuluan

Indonesia termasuk negara dengan tingkat kerawanan bencana geologi yang sangat tinggi di dunia, karena posisinya berada berposisi di pertemuan tiga lempeng tektonik utama: lempeng Indo-Australia, Eurasia, dan Pasifik. [1]. Aktivitas pergerakan lempeng tersebut menyebabkan gempa bumi terjadi secara terus-menerus dan sebagian dapat memicu tsunami, khususnya bila terjadi pada zona subduksi di dasar laut [2]. Fenomena ini terbukti dari beberapa bencana besar seperti tsunami Aceh tahun 2004, tsunami Banten tahun 2018, serta gempa-tsunami Palu tahun yang sama. Akibat dari bencana ini tidak hanya menimbulkan korban jiwa, tetapi juga kerusakan infrastruktur dan kerugian ekonomi yang masif [3]. Risiko terulangnya bencana serupa tetap tinggi karena gempa bumi memiliki periode ulang dan dapat terjadi kembali di masa mendatang. Oleh sebab itu, penelitian mengenai prediksi potensi tsunami menjadi tema

yang sangat penting untuk dikaji, terutama dalam upaya mendukung sistem peringatan dini yang cepat, akurat, dan berbasis data guna meminimalkan risiko korban dan dampak kerugian di wilayah rawan bencana.

Pesisir Barat Lampung, khususnya wilayah Krui dan sekitarnya, memiliki posisi geografis yang menghadap langsung ke Samudera Hindia dan berada pada zona subduksi aktif, sehingga termasuk kategori daerah dengan risiko tinggi terhadap gempa bumi dan tsunami [4][5]. Kondisi geografis tersebut menjadikan wilayah ini rentan terhadap rambatan gelombang tsunami apabila terjadi gempa dengan magnitudo yang signifikan di sekitar zona subduksi. Data BMKG pada tahun 2023–2024 mencatat 352 kejadian gempa di Pesisir Barat Lampung dengan variasi magnitudo, kedalaman, dan jarak episentrum, namun sebagian besar gempa tersebut belum dapat diklasifikasikan secara otomatis apakah berpotensi tsunami atau tidak. Di sisi lain, kemampuan mitigasi wilayah ini masih belum optimal, baik dari sisi infrastruktur sistem peringatan dini maupun kesiapsiagaan masyarakat dalam menghadapi bencana pesisir [6]. Masalah utama yang muncul dan dapat diukur adalah ketiadaan model prediksi berbasis data yang mampu menentukan potensi tsunami secara cepat berdasarkan parameter seismik, sehingga keputusan evakuasi masih bergantung pada analisis manual dan berpotensi mengalami keterlambatan saat kondisi darurat.

Penelitian ini menawarkan solusi dengan menerapkan dan membandingkan beberapa algoritma *machine learning* sebagai model prediksi potensi tsunami yang mampu melakukan klasifikasi secara cepat dan otomatis berdasarkan parameter gempa bumi. Pendekatan *machine learning* dipilih karena mampu melakukan pembelajaran terhadap karakteristik data historis serta menghasilkan prediksi berbasis data tanpa memerlukan analisis manual. Penelitian sebelumnya menunjukkan bahwa algoritma *knn* dan *Naive Bayes* berkinerja baik selama melakukan klasifikasi potensi tsunami di Indonesia, dengan nilai *AUC* mencapai 94,4%) [7], sedangkan penelitian lain mengonfirmasi bahwa algoritma *Random Forest* mampu menghasilkan akurasi tinggi dalam klasifikasi gempa di Indonesia [8]. Pendekatan *ensemble learning* seperti *stacking* juga terbukti meningkatkan akurasi hingga 94%, karena menggabungkan beberapa model seperti *Decision Tree*, *Random Forest*, *Neural Network*, dan *kNN* [9]. Selain itu, pengembangan sistem menggunakan *Gaussian Naive Bayes* bahkan dapat diimplementasikan pada perangkat keras mikrokontroler untuk prediksi potensi tsunami berbasis variabel magnitudo, kedalaman, dan jarak episentrum [10]. Pada studi lain, *Random Forest* dan *LightGBM* menunjukkan performa stabil dalam pengolahan data seismik dengan nilai *MAE* dan *RMSE* rendah [11]. Temuan tersebut diperkuat oleh literatur bahwa *machine learning* memiliki kemampuan dalam mengenali pola, melakukan klasifikasi, serta menghasilkan prediksi berbasis data secara efisien [12], sehingga menjadi pendekatan yang tepat untuk meningkatkan kecepatan dan akurasi sistem prediksi tsunami di Pesisir Barat Lampung.

Penelitian ini diarahkan untuk menilai perbedaan performa lima algoritma *Machine Learning*, yaitu *Random Forest*, *K-nearest Neighbor*, *Naive Bayes*, *Neural Network*, dan *Decision Tree* dalam memprediksi potensi tsunami di Pesisir Barat Lampung berdasarkan parameter seismik seperti magnitudo, kedalaman, dan jarak episentrum. Melalui prosedur pelatihan dan evaluasi mengaplikasikan *Confusion Matrix* untuk penilaian. Dengan memanfaatkan metrik akurasi, presisi, recall, dan *F1-score*, penelitian ini berupaya menentukan algoritma yang memiliki performa paling akurat dan stabil sebagai dasar pengembangan sistem peringatan dini tsunami. Manfaat penelitian ini adalah memberikan kontribusi ilmiah untuk pengembangan model prediksi bencana berbasis data dan teknologi kecerdasan buatan, sekaligus manfaat praktis berupa rekomendasi model yang dapat diimplementasikan oleh pihak terkait seperti BMKG untuk mempercepat proses pengambilan keputusan dalam mitigasi bencana sehingga dapat meningkatkan kesiapsiagaan masyarakat di wilayah rawan tsunami.

## 2. Tinjauan Pustaka

### 1) “Classification of Potential Tsunami Disaster Due to Earthquakes in Indonesia Based on Machine Learning” (Mardiani et al., 2024).

Penelitian ini mengklasifikasikan potensi tsunami akibat gempa bumi di Indonesia menggunakan beberapa algoritma *machine learning*, yaitu *K-Nearest Neighbor (KNN)*, *Naive Bayes*, *Decision Tree*, dan *Random Forest*. Parameter data yang digunakan meliputi magnitudo, kedalaman, dan lokasi pusat gempa. Data diproses melalui tahap *preprocessing*, pembagian data latih dan uji, kemudian dievaluasi menggunakan metrik *accuracy*, *AUC*, *precision*, *recall*, dan *F1-score*. Hasil

menunjukkan bahwa algoritma *KNN* memiliki performa terbaik dengan nilai *AUC* 94,4%, sehingga dianggap paling optimal dalam klasifikasi potensi tsunami [7].

**2) “Implementation of Random Forest Algorithm to Classify Earthquake in Indonesia” (Pratiwi & Arum, 2025).**

Penelitian ini memanfaatkan *Algoritma Random Forest diterapkan dalam klasifikasi data* kejadian gempa di Indonesia. Dataset yang digunakan berisi data gempa dengan variabel magnitudo dan kedalaman. Proses penelitian meliputi tahapan preprocessing data dan evaluasi menggunakan *Confusion Matrix*. Hasil analisis memperlihatkan bahwa *Random Forest* mencapai akurasi hingga 97%, sehingga dikategorikan efektif pada kasus klasifikasi data seismic [8].

**3) “Klasifikasi Tsunami Gempa Bumi dengan Teknik Stacking Ensemble Machine Learning” (Sudarto & Kusri, 2024).**

Penelitian ini menggunakan metode *Stacking Ensemble*, yaitu menggabungkan beberapa algoritma seperti *Decision Tree*, *Random Forest*, *Neural Network*, *SVM*, dan *KNN* sebagai base model, dan *Logistic Regression* sebagai meta model. Parameter yang digunakan adalah magnitudo, kedalaman, dan jarak episentrum. Model dievaluasi melalui *Confusion Matrix*, temuan penelitian mengindikasikan bahwa model stacking meningkatkan akurasi menjadi 94%, mengatasi keterbatasan model tunggal dalam menghadapi data tsunami yang kompleks [9].

**4) “Implementasi Algoritma Gaussian Naïve Bayes Classifier Untuk Prediksi Potensi Tsunami Berbasis Mikrokontroler” (Saputra & Hakim, 2022).**

Penelitian ini mengembangkan sistem prediksi tsunami menggunakan *Gaussian Naïve Bayes* pada perangkat mikrokontroler (Arduino). Parameter input yang diproses meliputi magnitudo gempa, kedalaman hiposentrum, dan jarak ke pusat gempa. Model dievaluasi menggunakan perhitungan probabilitas dan pengujian langsung pada prototipe perangkat. Hasil penelitian menunjukkan akurasi 96%, membuktikan bahwa algoritma ini cukup efektif dan dapat diterapkan dalam sistem deteksi dini berbasis perangkat keras [10].

**5) “Prediksi Magnitudo Gempa Menggunakan Random Forest, Support Vector Regression, XGBoost, LightGBM, dan Multi-Layer Perceptron” (Maulita & Wahid, 2024).**

Penelitian ini membandingkan lima algoritma untuk memprediksi magnitude gempa, yaitu *Random Forest*, *Multi-Layer Perceptron (MLP)*, *SVR*, *LightGBM*, dan *XGBoost*. Parameter yang diproses adalah kedalaman gempa dan letak koordinat (*latitude–longitude*). Model dievaluasi menggunakan *MAE* dan *RMSE*. Hasil menunjukkan bahwa *Random Forest* dan *LightGBM* menjadi model paling stabil dengan nilai *MAE* = 0,47 dan *RMSE* = 0,63. [11].

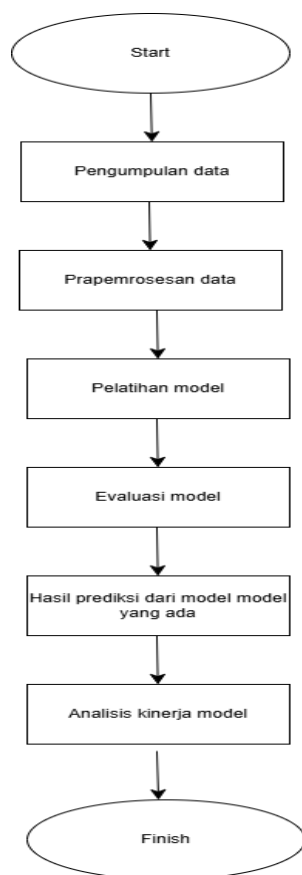
Berdasarkan telaah terhadap penelitian-penelitian sebelumnya, terlihat bahwa riset mengenai prediksi potensi tsunami dan aktivitas geologis dengan pendekatan *machine learning* telah dilakukan menggunakan berbagai metode, seperti *KNN* dan *Naïve Bayes*, *Random Forest*, *Stacking Ensemble*, *Gaussian Naïve Bayes* berbasis mikrokontroler, hingga algoritma regresi seperti *Random Forest*, *SVR*, *LightGBM*, *XGBoost*, dan *MLP*. Namun, penelitian-penelitian tersebut belum memfokuskan diri pada wilayah dengan karakteristik geologi spesifik, terutama daerah yang berada pada zona subduksi aktif seperti Pesisir Barat Lampung. Selain itu, sebagian penelitian hanya menguji satu atau dua algoritma, atau berorientasi pada prediksi parameter gempa (misalnya magnitudo), bukan pada klasifikasi potensi tsunami secara langsung. Berbeda dengan penelitian sebelumnya, penelitian ini menghadirkan kebaruan (*novelty*) dengan melakukan perbandingan lima algoritma sekaligus, yaitu *Random Forest*, *KNN*, *Naïve Bayes*, *Neural Network*, dan *Decision Tree*, menggunakan dataset lokal BMKG di Pesisir Barat Lampung periode 2023–2024 yang terdiri dari 352 kejadian gempa. Penelitian ini tidak terbatas pada pengukuran akurasi, namun juga memprioritaskan nilai *recall* sebagai indikator utama karena kesalahan *false negative* pada prediksi tsunami dapat menyebabkan risiko korban jiwa. Dengan demikian, penelitian ini menghasilkan menentukan model terbaik yang lebih dari sekadar menunjukkan keunggulan secara statistik, sekaligus relevan terhadap kebutuhan secara operasional untuk mendukung sistem peringatan dini tsunami pada wilayah pesisir yang memiliki kerentanan tinggi terhadap bencana.

### 3. Metodologi

Dalam studi ini, algoritma *machine learning* dimanfaatkan untuk digunakan untuk memprediksi gempa bumi yang berpotensi tsunami yang berada di Pesisir Barat Lampung dan melakukan perbandingan algoritma mana yang terbaik, data gempa yang digunakan yaitu pada tahun 2023 - 2024. Metode penelitian ini mencakup pengumpulan data, prapemrosesan data, pelatihan model, evaluasi model, hasil prediksi dari model yang ada, dan analisis kerja model.

#### 3.1. Alur penelitian

Tahapan pada Metode studi kasus ini menggunakan Algoritma *Machine Learning* Menggunakan 5 metode algoritma seperti *kNN*, *Naive Bayes*, *Decision Tree*, *Neural Network* dan *Random Forest* dengan kerangka kerja yang terdiri dari beberapa bagian pokok yang dapat mempermudah peneliti dalam melaksanakan penelitian sebab akan menjadi landasan dalam penelitian ini, sebagaimana ditunjukkan pada Gambar 1.



**Gambar 1.** Alur penelitian

#### 3.2. Pengumpulan data

Penelitian ini bermula dari pengumpulan data yang diperoleh secara resmi dari BMKG Lampung dan difokuskan pada kejadian gempa di wilayah Pesisir Barat Lampung. Data ini dipilih karena wilayah tersebut memiliki tingkat risiko tsunami yang tinggi akibat kedekatannya dengan zona subduksi aktif di Samudera Hindia. Total ada 352 data mengenai gempa yang terjadi pada pesisir barat Lampung pada tahun 2023 - 2024, dengan variabel data magnitudo(sr), kedalaman gempa(km), jarak episentrum(km) serta label klasifikasi potensi tsunami (ya=1 atau tidak=0), data tersebut diperoleh dari BMKG. Bisa dilihat pada tabel 1.

**Tabel 1.** Data Gempa Bumi Pesisir Barat Lampung

No	Magnitudo	Kedalaman	Jarak Episentrum	Potensi tsunami
1	2.2	57	20	0
2	3	1	86	0
3	4.7	34	57	1

No	Magnitudo	Kedalaman	Jarak Episentrum	Potensi tsunami
4	2.9	10	82	0
5	2.9	26	46	0
6	3.6	58	146	0
7	2.2	25	58	0
8	2.2	69	5	0
9	2.4	53	28	0

### 3.3. Prapemrosesan data

Sebelum data dimasukkan ke dalam model, dilakukan tahap prapemrosesan data untuk memastikan kualitas input data yaitu:

#### 1) Pembersihan data (*data cleaning*)

Prosedur pembersihan data dilaksanakan dengan tujuan menghapus data yang berpotensi menurunkan kualitas model. Prosedur pembersihan data diawali dengan melakukan pemeriksaan terhadap seluruh atribut pada dataset dalam rangka mendeteksi entri yang kosong (*missing value*) serta data yang tidak konsisten. Data yang memiliki nilai yang hilang pada salah satu variabel utama, yaitu magnitudo, kedalaman, atau jarak episentrum, dihapus dari dataset. Selain itu, dilakukan pemeriksaan logis terhadap nilai data untuk memastikan tidak terdapat nilai yang berada di luar rentang kewajaran parameter gempa bumi. Proses ini bertujuan menghasilkan dataset yang bersih dan bebas dari noise sebelum diproses lebih lanjut.

#### 2) Normalisasi

Setelah data bersih, dilakukan proses normalisasi untuk menyeragamkan skala antar variabel numerik. Normalisasi diterapkan pada variabel magnitudo, kedalaman, dan jarak episentrum, yang memiliki rentang nilai berbeda. Proses ini dilakukan dengan mentransformasikan nilai setiap variabel ke dalam rentang tertentu agar tidak terjadi dominasi variabel dengan skala lebih besar terhadap algoritma berbasis jarak maupun pembobotan. Dengan normalisasi, setiap faktor memiliki pengaruh yang seimbang dalam tahapan pembelajaran model.

#### 3) Pengkodean label

Tahap selanjutnya adalah pengkodean label pada variabel target, yaitu potensi tsunami. Variabel yang semula berbentuk kategorikal “ya” dan “tidak” dikonversi ke dalam bentuk numerik. Label “tidak” direpresentasikan dengan nilai 0, sedangkan label “ya” direpresentasikan dengan nilai 1. Pengkodean ini dilakukan agar variabel target dapat diinterpretasikan secara komputasional melalui algoritma *machine learning* yang memerlukan input numerik.

### 3.4. Pelatihan dan pengujian model

Setelah dilakukan prapemrosesan data, Kumpulan data dipisahkan menjadi dua subset: data latih (*training set*) 80% dan data uji (*testing set*) 20%. Model dilatih menggunakan *training set*, kemudian diuji pada *testing set* untuk mengevaluasi performa prediksi untuk data yang belum pernah dilihat oleh model.

### 3.5. Evaluasi kinerja model

Kinerja model dievaluasi menggunakan metrik *Confusion Matrix*, yang menghasilkan ukuran *accuracy*, *precision*, *recall*, dan *F1-score*. Metrik ini dipilih karena mampu menggambarkan keseimbangan antara kemampuan model dalam mendeteksi kejadian tsunami (sensitivitas) dan menghindari kesalahan prediksi positif (*false positive*).

### 3.6. Metode *Machine Learning*

Metode *machine learning* dimanfaatkan dalam penelitian ini untuk mengembangkan model klasifikasi potensi tsunami berdasarkan parameter gempa bumi. Istilah *machine learning* diciptakan oleh Arthur Samuel dan diperkenalkan pada 1959 dalam konteks pengembangan sistem komputer yang mampu mempelajari strategi permainan catur secara mandiri. Secara terminologis, *machine learning* merujuk pada pendekatan komputasi statistik yang memfokuskan diri pada proses prediksi berbasis data menggunakan algoritma tertentu [13]. Teknologi ini mencakup serangkaian algoritma yang dirancang untuk mengoptimalkan kinerja sistem melalui

proses pembelajaran dari data sampel yang tersedia. Merupakan bagian dari karakteristik utama dari *machine learning* fokus pada kemampuan dalam beradaptasi dan memodifikasi keputusan secara otomatis sebagai respons terhadap perubahan lingkungan atau data yang masuk [14]. Selain itu, *machine learning* memungkinkan sistem untuk mengekstraksi pola atau aturan keputusan, misalnya dalam pengenalan emosi, dengan mengandalkan data pelatihan yang telah dilabeli sebelumnya [15]. Pendekatan ini memperkuat peran *machine learning* dalam berbagai domain, terutama yang memerlukan analisis prediktif dan klasifikasi berbasis data.

### 3.7. Algoritma Random Forest

Random Forest adalah algoritma ensemble yang menggabungkan sejumlah pohon keputusan, dengan output akhir ditentukan melalui mekanisme voting dari semua pohon yang terbentuk [16] [17]. Berikut rumusnya:

$$H(x) = \text{mode}\{h_1(x), h_2(x), \dots, h_n(x)\} \quad (1)$$

Keterangan:

$h_i(x)$  = keluaran dari pohon keputusan ke- $i$

$H(x)$  = hasil klasifikasi akhir berdasarkan suara terbanyak

Pengacakan pada pemilihan fitur di tiap node menjadikan *Random Forest* lebih tahan terhadap *overfitting* [18].

### 3.8. Algoritma KNN

*KNN* mengklasifikasikan data baru dengan membandingkan jaraknya dengan sampel yang sudah ada dalam dataset, menggunakan ukuran jarak *Euclidean* [19][20]. Berikut rumusnya :

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Keterangan:

$X$  = data yang akan diprediksi

$Y$  = data pembandingan pada dataset

Data Baru akan diklasifikasikan berdasarkan kelas mayoritas dari tetangga terdekatnya.

### 3.9. Algoritma Naive Bayes

*Naive Bayes* memanfaatkan *teorema Bayes* untuk menghitung probabilitas kelas dari suatu data dengan anggapan bahwa setiap atribut tidak saling bergantung satu sama lain [21]. Berikut rumusnya:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (3)$$

Keterangan:

$P(C|X)$  = peluang data termasuk kelas  $C$

$P(X|C)$  = peluang fitur  $X$  muncul pada kelas  $C$

$P(C)$  = peluang awal dari kelas

Kelas dengan nilai probabilitas tertinggi menjadi hasil prediksi.

### 3.10. Algoritma Neural Network

*Artificial Neural Network* memodelkan hubungan antar variabel melalui *neuron* dengan bobot tertentu, dan pembelajaran dilakukan menggunakan algoritma *backpropagation* [21]. Berikut rumusnya:

$$y = f(\sum_{i=1}^n w_i x_i + b) \quad (4)$$

Dengan fungsi aktivasi umum:

$$f(x) = \frac{1}{1+e^{-x}} \quad (5)$$

Keterangan:

$x_i$  = input

$w_i$  = bobot

$b$  = bias

$f$  = fungsi aktivasi

### 3.11. Algoritma Decision Tree

*Decision Tree* membangun pohon keputusan dengan menentukan atribut paling optimal berdasarkan information gain, yang dihitung menggunakan entropy [22]. Berikut rumusnya:

$$\text{Entropy}(S) = -\sum p_i \log_2(p_i) \quad (6)$$

$$Gain(S, A) = Entropy(S) - \sum \frac{|S_u|}{|S|} Entropy(S_u) \quad (7)$$

Fitur dengan nilai gain paling tinggi dipilih menjadi keputusan pada node berikutnya.

### 3.12. Confussion matrix

*Confusion Matrix* merupakan tabel evaluasi berfungsi menilai performa model klasifikasi melalui perbandingan antara prediksi model dan data aktual, dengan tabel yang memuat empat komponen utama: TP, TN, FP, dan FN [23]. Nilai-nilai tersebut menjadi dasar dalam menentukan indikator kinerja termasuk akurasi, presisi, *recall*, dan *F1-score* [24]. Susunan *umum confusion matrix* ditampilkan pada Tabel 2 berikut:

**Tabel 2.** Struktur *Confussion matrix*.

	Prediksi Positif	Prediksi Negatif
aktual positif	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
Aktual negatif	<i>False Positive (FP)</i>	True Negative (TN)

Keterangan:

TP (*True Positive*) : kasus aktual positif dan berhasil diprediksi sebagai positif.

TN (*True Negative*) : kasus aktual negatif dan berhasil diprediksi sebagai negatif.

FP (*False Positive*) : kasus aktual negatif tetapi diprediksi positif (*false alarm*)

FN (*False Negative*) : kasus aktual positif tetapi diprediksi negatif (kejadian penting terlewat).

Dari hasil *confusion matrix*, beberapa metric evaluasi dapat dihitung sebagai berikut :

*Accuracy*: mengukur seberapa banyak prediksi model sesuai dengan data actual.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

*Precision*: mengukur tingkat ketepatan model dalam memberikan prediksi positif.

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

*Recall*: menilai kemampuan model dalam mengenali seluruh data dengan kelas positif.

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

*F1-Score*: merupakan harmonisasi dari nilai precision dan recall

$$Accuracy = 2 \times \frac{Precision \cdot Recall}{Precision + Recall} \quad (11)$$

## 4. Hasil dan Pembahasan

### 4.1. Pengumpulan data

Data gempa bumi yang dijadikan objek penelitian mencakup 352 kejadian di Pesisir Barat Lampung pada periode 2023–2024. Setiap data terdiri dari variabel magnitudo, kedalaman gempa, jarak episentrum, serta label potensi tsunami. Dataset ini menjadi dasar dalam seluruh tahapan pengujian model *machine learning*.

### 4.2. Prapemrosesan data

Hasil dari tahap prapemrosesan menunjukkan bahwa dataset akhir berada dalam kondisi siap digunakan untuk pemodelan. Seluruh data yang tidak lengkap telah tersaring, variabel numerik berada pada skala yang seragam, dan label potensi tsunami telah dikonversi ke dalam bentuk numerik. Tidak ditemukan data yang tidak konsisten pada dataset akhir yang digunakan untuk melatih serta menguji model

#### 4.2.1. Pembersihan data

Hasil pembersihan data menunjukkan bahwa dataset akhir yang diterapkan dalam penelitian ini tidak mengandung nilai kosong maupun data yang tidak konsisten. Seluruh data gempa yang tidak memenuhi kriteria kelengkapan telah tersaring, sehingga dataset yang digunakan sepenuhnya valid untuk tahap pemodelan.

#### 4.2.2. Normalisasi

Hasil normalisasi data menunjukkan bahwa seluruh variabel numerik, yaitu magnitudo, kedalaman gempa, dan jarak episentrum, telah berada pada skala yang seragam. Kondisi ini memastikan bahwa tidak terdapat variabel yang mendominasi proses pembelajaran model akibat perbedaan rentang nilai.

#### 4.2.3. Pengkodean label

Hasil pengkodean label menunjukkan bahwa variabel target potensi tsunami telah berhasil direpresentasikan dalam bentuk numerik. Kelas tidak berpotensi tsunami direpresentasikan dengan nilai 0, sedangkan kelas berpotensi tsunami direpresentasikan dengan nilai 1. Dengan demikian, seluruh data telah berada dalam format numerik dan siap dimanfaatkan selama tahap pelatihan dan pengujian model.

#### 4.2.4. Pembagian dataset

Luaran dari tahap pembagian dataset menghasilkan dua subset data, Dataset dibagi menjadi 80% untuk data latih dan 20% untuk data uji. Bersumber dari total 352 data, sebanyak 282 data digunakan untuk pelatihan, sedangkan 70 data digunakan untuk pengujian. Pembagian ini memungkinkan evaluasi kinerja model dilakukan pada data yang sebelumnya belum pernah digunakan dalam proses pelatihan.

### 4.3. Proses pelatihan dan pengujian model

Hasil pelatihan dan pengujian model ditampilkan dalam bentuk *Confusion Matrix* untuk masing-masing algoritma *machine learning*. *Confusion matrix* menggambarkan jumlah prediksi yang benar dan kesalahan klasifikasi pada kelas tsunami dan tidak tsunami. Perbedaan pola kesalahan antar algoritma menunjukkan variasi kemampuan model dalam mengklasifikasikan potensi tsunami.

#### 4.3.1. Hasil pelatihan *Confusion Matrix*

Hasil pelatihan dan pengujian model ditunjukkan melalui *Confusion Matrix* untuk masing-masing algoritma *machine learning*. *Confusion matrix* menggambarkan distribusi prediksi model terhadap data aktual, termasuk jumlah prediksi benar dan kesalahan klasifikasi pada kelas tsunami dan tidak tsunami.

Hasil dari 5 Algoritma menggunakan *Confusion Matrix* terlihat pada tabel berikut:

**Tabel 3.** Hasil *Confussion Matrix Random Forest*

<b>Actual Classification</b>	<b>Prediction Classification</b>		
	0	1	$\Sigma$
0	67	0	67
1	1	2	3
$\Sigma$	68	2	70

**Tabel 4.** Hasil *Confussion Matrix KNN*

<b>Actual Classification</b>	<b>Prediction Classificaton</b>		
	0	1	$\Sigma$
0	67	0	67
1	3	0	3
$\Sigma$	70	0	70

**Tabel 5.** Hasil *Confussion Matrix Naive Bayes*

<b>Actual Classification</b>	<b>Prediction Classification</b>		
	0	1	$\Sigma$
0	64	3	67
1	0	3	3
$\Sigma$	64	6	70



**Tabel 6.** Hasil *Confussion Matrix Neural Network*

<b>Actual Classification</b>	<b>Prediction Classification</b>		
	0	1	$\Sigma$
0	67	0	67
1	1	2	3
$\Sigma$	68	2	70

**Tabel 7.** Hasil *Confussion Matrix Decision Tree*

<b>Actual Classification</b>	<b>Prediction Classification</b>		
	0	1	$\Sigma$
0	67	0	67
1	3	0	3
$\Sigma$	70	0	70

#### 4.4. Hasil prediksi dari *confussion matrix*

Berdasarkan hasil *confussion matrix* dari 5 algoritma *Machine Learning*, dapat dilakukan perhitungan terhadap sejumlah metrik evaluasi performa model klasifikasi, antara lain akurasi, *precision*, *recall*, *F1-Score*. Berdasarkan *Confusion Matrix*, Berikut adalah tabel ringkasan matriknya menggunakan *tools orange*:

**Tabel 8.** Hasil prediksi matrik kelima algoritma

<i>Model</i>	<i>Akurasi</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 - Score</i>
<i>Random Forest</i>	1 (100%)	0.986 (98,6%)	0.986 (98,6%)	0.984 (98,4%)
<i>Knn</i>	0.940	0.979	0.957	0.964
<i>Naive Bayes</i>	0.985	0.979	0.957	0.964
<i>Neural Network</i>	0.985	0.986	0.986	0.984
<i>Decision Tree</i>	0.500	0.916	0.957	0.936

Hasil menunjukkan bahwa *Random Forest* unggul dibanding algoritma lain. *KNN*, *Naïve Bayes*, dan *Neural Network* cukup baik, tetapi masih terdapat kesalahan klasifikasi. *Decision Tree* menghasilkan akurasi terendah, hanya mencapai akurasi 50% karena cenderung *overfitting* pada dataset kecil. Perlu dicatat bahwa meskipun akurasi tinggi merupakan indikator performa yang baik, dalam konteks mitigasi bencana, *recall* memiliki peran yang jauh lebih penting. Hal ini karena *recall* menekankan pada kemampuan model mendeteksi semua kejadian tsunami yang benar-benar terjadi. Skenario false negative (tsunami terjadi namun tidak terdeteksi model) merupakan kondisi paling berbahaya karena dapat mengakibatkan korban jiwa akibat kelalaian sistem prediksi. Sebaliknya, skenario *false positive* (model memprediksi tsunami padahal tidak ada) meskipun dapat menimbulkan gangguan, relatif lebih dapat ditoleransi dibanding *false negative*. Misalnya, masyarakat mungkin dievakuasi secara sia-sia, namun tidak menimbulkan korban jiwa. Oleh karena itu, keseimbangan antara *precision* dan *recall* yang diwakili oleh *F1-score* menjadi acuan utama dalam menentukan model terbaik.

#### 4.5. Pembahasan

Hasil analisis memperlihatkan bahwa *Random Forest* menunjukkan performa terbaik dalam memprediksi potensi tsunami pada dataset BMKG Pesisir Barat Lampung dibandingkan algoritma *Naïve Bayes*, *KNN*, *Neural Network* dan *Decision Tree*. Temuan ini sejalan dengan penelitian Pratiwi & Arum (2025), yang menyatakan bahwa *Random Forest* unggul dalam penentuan kategori data gempa dengan akurasi tinggi [8], dan konsisten dengan penelitian Maulita & Wahid (2024) yang menegaskan stabilitas model *Random Forest* pada data seismik [11]. Penelitian ini memperkuat studi tersebut dengan konteks berbeda, yaitu prediksi potensi tsunami dan fokus pada variabel geofisika lokal. Berbeda dengan Mardiani et al. (2024) yang menyatakan bahwa *KNN* lebih unggul berdasarkan nilai *AUC* [7], hasil penelitian ini justru menunjukkan bahwa kinerja *KNN* masih lebih rendah daripada *Random Forest* pada dataset Lampung. Temuan ini mengindikasikan bahwa performa algoritma sangat dipengaruhi karakteristik data dan tidak dapat digeneralisasi. Dengan demikian, kontribusi signifikan penelitian ini berupa penyediaan bukti empiris bahwa *Random Forest* merupakan model paling optimal untuk prediksi potensi tsunami berbasis data lokal, serta memperkuat literatur bahwa pemilihan algoritma *machine learning* harus disesuaikan dengan karakteristik dataset dan kebutuhan mitigasi bencana.

## 5. Simpulan

Penelitian ini membandingkan lima algoritma *machine learning*, yaitu *Random Forest*, *k-NN*, *Naive Bayes*, *Neural Network*, dan *Decision Tree*, untuk memprediksi potensi tsunami di Pesisir Barat Lampung berdasarkan parameter magnitudo, kedalaman, dan jarak episentrum. Hasil evaluasi menggunakan *Confusion Matrix* mengindikasikan bahwa algoritma *Random Forest* mencapai kinerja terbaik dengan tingkat akurasi 100% serta indikator kinerja *precision*, *recall*, dan *F1-score* yang konsisten. Algoritma *Neural Network* juga menunjukkan hasil yang mendekati *Random Forest*, sedangkan *KNN* dan *Naive Bayes* berada pada tingkat performa menengah. *Decision Tree* menghasilkan kinerja terendah akibat kecenderungan *overfitting* pada dataset terbatas. Temuan ini menjawab tujuan penelitian, yaitu mengidentifikasi algoritma paling tepat untuk prediksi potensi tsunami di wilayah dengan karakteristik geologi kompleks. Berdasarkan temuan ini, *Random Forest* dianjurkan sebagai model utama untuk diimplementasikan dalam sistem prediksi dan peringatan dini tsunami. Selain itu, temuan ini mendorong kemungkinan pengembangan lebih lanjut melalui integrasi dengan data real-time BMKG, penggunaan dataset yang lebih besar, serta penerapan metode *ensemble* yang lebih kompleks untuk meningkatkan *robustnes* model. Dengan demikian, penelitian ini melampaui sekadar kontribusi dalam aspek akademik, tetapi juga memiliki nilai praktis dalam mendukung mitigasi bencana dan kesiapsiagaan masyarakat di wilayah rawan tsunami.

## Acknowledgements

Penulis mengucapkan terima kasih atas pelaksanaan program pendanaan riset inovatif produktif mandatory bertema Indonesia tahun 2023 – 2028 Batch II kepada Direktorat Jendral Pendidikan Tinggi, Riset, dan Teknologi, serta Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbud-Ristek DIKTI) dan Universitas Gadjah Mada yang telah memberikan dukungan dana penelitian ini berdasarkan Surat Keputusan Nomor : 2924/E3/AL.04/2024, 4809/UN1.P/HK.08.00/2024 Dukungan dana ini sangat penting untuk mempelancar berbagai tahapan penelitian, mulai dari pengumpulan data hingga analisis dan dokumentasi akhir. Selain itu, penulis mengucapkan terima kasih kepada Nanyang Technological University Singapore Institute of Research for Sustainability and Innovation dan Institut Informatika dan Bisnis Darmajaya beserta dengan LAB PUI-TI IIB Darmajaya, khususnya Fakultas Ilmu Komputer, atas dukungan dan dorongan yang terus menerus dalam penelitian ini.

## Daftar Referensi

- [1] A. Angraeni, L. H. Lubis, Sugeng, and M. H. Ginting, "Analisis Kerentanan Seismik Gempa Bumi Berdasarkan Nilai PGA Menggunakan Metode Esteve Pada Wilayah Kepulauan Nias," *J. Geo Image*, vol. 12, no. 120, pp. 141–147, 2023.
- [2] K. Sassa, K. Konagai, B. Tiwari, Z. Arbanas, and S. Sassa, *Progress in Landslide Research and Technology, Volume 1 Issue 1*, 2022. in *Progress in Landslide Research and Technology*. 2022. [Online]. Available: <https://books.google.co.id/books?id=p92mEAAAQBAJ>
- [3] J. Selva *et al.*, "Probabilistic tsunami forecasting for early warning," *Nat. Commun.*, vol. 12, no. 1, 2021, doi: 10.1038/s41467-021-25815-w.
- [4] E. I. Riyani and - Tamjuddin, "Tantangan Pengembangan Ekowisata Bahari Di Pulau Pisang Pesisir Barat Lampung," *EKOMBIS Rev. J. Ilm. Ekon. dan Bisnis*, vol. 5, no. 1, pp. 16–24, 2017, doi: 10.37676/ekombis.v5i1.326.
- [5] D. Pratiwi and A. Fitri, "Analisis Potensial Penjalaran Gelombang Tsunami di Pesisir Barat Lampung, Indonesia," *J. Tek. Sipil ITP*, vol. 8, no. 1, p. 5, 2021, doi: 10.21063/jts.2021.v8i1.05.
- [6] R. S. Oktari, Syamsidik, R. Idroes, H. Sofyan, and K. Munadi, "City resilience towards coastal hazards: An integrated bottom-up and top-down assessment," *Water (Switzerland)*, vol. 12, no. 10, 2020, doi: 10.3390/w12102823.
- [7] E. Mardiani *et al.*, "Classification of Potential Tsunami Disaster Due to Earthquakes in Indonesia Based on Machine Learning," *Int. J. Softw. Eng. Comput. Sci.*, vol. 4, no. 1, pp. 13–23, 2024, doi: 10.35870/ijsecs.v4i1.2084.
- [8] A. P. Pratiwi and P. R. Arum, "Implementation of Random Forest Algorithm to Classify Earthquake in Indonesia," *Eig. Math. J.*, vol. 8, no. 1, pp. 25–33, 2025.
- [9] S. Sudarto and K. Kusriani, "Klasifikasi Tsunami Gempa Bumi dengan Teknik Stacking Ensemble Machine Learning," *J. Inform. Polinema*, vol. 10, no. 4, pp. 511–520, 2024, doi:

- 10.33795/jip.v10i4.5655.
- [10] D. Irawan Saputra and D. Lukman Hakim, "Implementasi Algoritma Gaussian Naive Bayes Classifier Untuk Prediksi Potensi Tsunami Berbasis Mikrokontroler," *Epsil. J. Electr. Eng. Inf. Technol.*, vol. 2, pp. 122–138, 2022.
  - [11] I. Maulita and A. M. Wahid, "Prediksi Magnitudo Gempa Menggunakan Random Forest, Support Vector Regression, XGBoost, LightGBM, dan Multi-Layer Perceptron Berdasarkan Data Kedalaman dan Geolokasi," *J. Pendidik. dan Teknol. Indones.*, vol. 4, no. 5, pp. 221–232, 2024, doi: 10.52436/1.jpti.470.
  - [12] Muntiari Novita Ranti and Hanif Kharis Hudaiby, "Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning," *J. Ilmu Komput. dan Teknol.*, vol. 3, no. 1, pp. 1–6, 2022, [Online]. Available: <http://creativecommons.org/licenses/by/4.0/%0Ahttp://ejournal.uhb.ac.id/index.php/IKOMTI>
  - [13] R. R. Santoso, R. Megasari, and Y. A. Hambali, "Implementasi Metode Machine Learning Menggunakan Algoritma Evolving Artificial Neural Network Pada Kasus Prediksi Diagnosis Diabetes," *J. Apl. dan Teor. Ilmu Komput.*, vol. 3, no. 2, pp. 85–97, 2020, [Online]. Available: <https://ejournal.upi.edu/index.php/JATIKOM>
  - [14] H. K. Pambudi, P. G. A. Kusuma, F. Yulianti, and K. A. Julian, "Prediksi Status Pengiriman Barang Menggunakan Metode Machine Learning," *J. Ilm. Teknol. Infomasi Terap.*, vol. 6, no. 2, pp. 100–109, 2020, doi: 10.33197/jitter.vol6.iss2.2020.396.
  - [15] R. Nisa, S. Amriza, and D. Supriyadi, "Komparasi Metode Machine Learning Dan Deep Learning Untuk Deteksi Emosi Pada Text Di Sosial Media," *Penelit. Ilmu dan Teknol. Komput.*, vol. 13, no. 2, pp. 130–139, 2021.
  - [16] M. M. Mutoffar, M. Naseer, and A. Fadillah, "Klasifikasi Kualitas Air Sumur Menggunakan Algoritma Random Forest," vol. 04, no. 02, pp. 138–146, 2022.
  - [17] B. Prasajo and E. Haryatmi, "Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest," *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 2, pp. 79–89, 2021, doi: 10.25077/teknosi.v7i2.2021.79-89.
  - [18] Suci Amaliah, M. Nusrang, and A. Aswi, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 4, no. 3, pp. 121–127, 2022, doi: 10.35580/variensium31.
  - [19] A. Firdaus, "Aplikasi Algoritma K-Nearest Neighbor pada Analisis Sentimen Omicron Covid-19," *J. Ris. Stat.*, vol. 2, no. 2, pp. 85–92, 2022, doi: 10.29313/jrs.v2i2.1148.
  - [20] M. N. Maskuri, K. Sukerti, and R. M. Herdian Bhakti, "Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke Stroke Disease Predict Using KNN Algorithm," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 1, pp. 130–140, 2022.
  - [21] E. Nanda, Istikomah, N. A. Amari, and Y. Pristyanto, "Perbandingan Klasifikasi Algoritma K-Nn, Neural Network, Naïve Bayes, C 4.5 Untuk Mendeteksi Web Phising," *J. Inform. Komputer, Bisnis dan Manaj.*, vol. 16, no. 3, pp. 33–42, 2023, doi: 10.61805/fahma.v16i3.88.
  - [22] S. K. M. K. Mira, S. K. M. K. Azriel Christian Nurcahyo, S. K. M. K. Candra Gudiato, S. K. M. K. Noviyanti. P, and S. K. M. K. Listra Frigia Missianes Horhoruw, *Data Mining Mengeksplorasi Teknik-Teknik Data Mining dan Metode K-Means Teori, Konsep, Algoritma dan Studi Kasus*. Uwais Inspirasi Indonesia, 2025. [Online]. Available: <https://books.google.co.id/books?id=32BQEQAAQBAJ>
  - [23] E. Fauziningrum, M. Pd and M. P. Encis Indah Suryaningsih, S.T., "Evaluasi Dan Prediksi Penguasaan Bahasa Inggris Maritim Menggunakan Metode Decision Tree Dan Confusion Matrix (Studi Kasus Di Universitas Maritim Amni)," *Angew. Chemie Int. Ed.* 6(11), 951–952., pp. 5–24, 2021.
  - [24] A. F. Anjani, D. Anggraeni, and I. M. Tirta, "Implementasi Random Forest Menggunakan SMOTE untuk Analisis Sentimen Ulasan Aplikasi Sister for Students UNEJ," *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 2, pp. 163–172, 2023, doi: 10.25077/teknosi.v9i2.2023.163-172.