

KLASIFIKASI TSUNAMI GEMPA BUMI DENGAN TEKNIK STACKING ENSEMBLE MACHINE LEARNING

Sudarto¹, Kusri²

^{1,2}Magister Teknik Informatika, Universitas AMIKOM Yogyakarta
¹sudarto@students.amikom.ac.id, ²kusri@amikom.ac.id

Abstrak

Upaya untuk mendeteksi dan mengklasifikasikan tsunami secara dini sangat penting untuk mitigasi risiko dan pengurangan dampak bencana. Meskipun berbagai metode telah diusulkan untuk meningkatkan akurasi prediksi tsunami, tantangan utama masih terletak pada variasi akurasi antara model yang berbeda dan ketidakmampuan beberapa model untuk menangani data yang kompleks dan beragam. Teknik *stacking ensemble* menawarkan solusi dengan menggabungkan kekuatan beberapa model pembelajaran mesin untuk menghasilkan prediksi yang lebih akurat dan andal. Penelitian ini bertujuan untuk menguji efektivitas teknik *stacking ensemble* dalam klasifikasi tsunami yang disebabkan oleh gempa bumi. Teknik ini melibatkan penggabungan beberapa model pembelajaran mesin untuk meningkatkan akurasi prediksi. Dataset yang digunakan dalam penelitian ini adalah data tsunami pasca gempa bumi dari tahun 2001 hingga 2023 yang berasal dari Kaggle.com. Proses penelitian mencakup eksplorasi data awal (EDA), *preprocessing*, pembagian data menjadi set pelatihan dan pengujian, serta evaluasi model dasar Decision Tree, Random Forest, Neural Network, Support Vector Machine, Naive Bayes, dan K Nearest Neighbors. Hasil menunjukkan bahwa model *stacking ensemble* dengan Logistic Regression sebagai meta model memberikan kinerja terbaik dalam klasifikasi tsunami dengan akurasi sebesar 94%. Secara keseluruhan, *stacking ensemble* dapat meningkatkan akurasi prediksi dalam klasifikasi tsunami dan dapat digunakan sebagai dasar untuk pengembangan sistem peringatan dini tsunami yang lebih efektif.

Kata kunci: tsunami, gempa bumi, *stacking ensemble*, *machine learning*, klasifikasi.

1. Pendahuluan

Robke dan Volt menyatakan bahwa Tsunami merupakan salah satu dari bencana alam yang paling menakutkan yang dapat terjadi di daerah pesisir (Latue & Latue, 2023). Kejadian ini disebabkan oleh peristiwa alam seperti gempa bumi, letusan gunung berapi, atau longsor di bawah laut yang mengakibatkan pergerakan besar di dasar laut dan memicu gelombang laut yang sangat kuat (Kamigaichi dalam Latue & Latue, 2023).

Klasifikasi tsunami penting untuk pemantauan, deteksi dini, evaluasi dampak, penelitian, dan pengembangan sistem peringatan dini. Dengan mengidentifikasi daerah rentan, mengevaluasi kerusakan, dan menganalisis pola tsunami, klasifikasi tsunami membantu dalam mitigasi bencana, perlindungan masyarakat, dan pengembangan langkah-langkah preventif yang efektif (Mateen et al., 2024).

Stacking ensemble adalah salah satu metode ensemble machine yang melibatkan penggabungan hasil dari beberapa model pembelajaran mesin (learners) untuk meningkatkan kinerja prediksi. Konsep dasar dari *stacking* adalah menggabungkan prediksi dari model-model pembelajaran mesin yang berbeda untuk menciptakan model yang lebih kuat dan akurat (Zhou, 2012). Berikut adalah gambaran

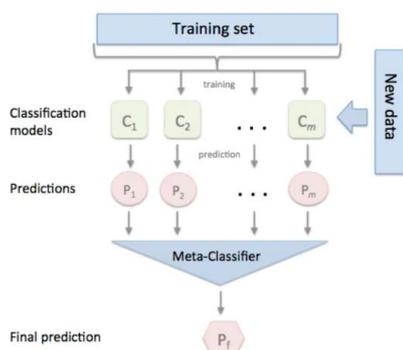
tentang bagaimana *stacking ensemble* bekerja (Zhou, 2012):

1. Pembelajaran Dasar (Pembelajar Tingkat Pertama)
 - a. Beberapa pembelajar dasar dilatih pada dataset yang sama menggunakan berbagai algoritma pembelajaran atau variasi dari algoritma yang sama.
 - b. Setiap pembelajar dasar menghasilkan prediksi untuk variabel target berdasarkan fitur input.
2. Meta Pembelajaran (Pembelajar Tingkat Kedua)
 - a. Sebuah meta pembelajar dilatih pada prediksi yang dibuat oleh pembelajar dasar.
 - b. Meta-pembelajar belajar bagaimana menggabungkan prediksi pembelajar dasar untuk membuat prediksi akhir.
 - c. Meta-pembelajar dapat berupa model sederhana seperti regresi linear, pohon keputusan, atau model yang lebih kompleks seperti jaringan saraf.
3. Proses Pelatihan

Proses pelatihan melibatkan dua tahap:

 - a. Tahap 1: Latih pembelajar dasar pada dataset asli.
 - b. Tahap 2: Gunakan prediksi pembelajar dasar sebagai fitur input untuk melatih meta-pembelajar.
4. Proses Prediksi

- a. Selama fase prediksi, data input dilewatkan melalui pembelajar dasar untuk menghasilkan prediksi.
 - b. Prediksi ini kemudian dimasukkan ke meta-pembelajar untuk menghasilkan prediksi akhir.
- Alur ilustrasi dari teknik *stacking ensemble* dapat dilihat pada Gambar 1.



Gambar 1. Ilustrasi teknik *stacking ensemble*

Beberapa peneliti sudah pernah melakukan penelitian terkait dengan klasifikasi Tsunami dan *stacking ensemble*. Sukmana et al., (2024), Wang et al., (2023), Saputra & Hakim, (2022). Mereka melakukan penelitian terkait dengan prediksi tsunami dengan menggunakan dataset, algoritma dan metode yang berbeda. Tingkat akurasi model yang terbentuk berbeda-beda dari terkecil 65.61%-penelitian oleh Sukmana et al. (2024) dan tertinggi 96%-penelitian oleh Saputra dan Hakim (2022). Namun penelitian oleh Saputra dan Hakim (2022) ini mempunyai kelemahan yaitu cakupan data latih yang terbatas hanya pada gempa bumi yang terjadi di Indonesia selama dua puluh tahun terakhir. Hal ini dapat mempengaruhi generalisasi algoritma dalam memprediksi potensi tsunami untuk wilayah lain yang memiliki karakteristik gempa bumi yang berbeda.

Syamsiah dan Purwandani, (2023), Adnan et al., (2024), Daza et al., (2024), mereka melakukan penelitian terkait dengan teknik *stacking ensemble* dengan tema yang berbeda-beda. Syamsiah dan Purwandani mengangkat tema peramalan pemakaian listrik, Adnan et al. mengangkat tema klasifikasi email spam dan Daza et al. membahas tema tentang diagnosis penyakit diabetes. Penelitian yang mereka lakukan ini berhasil meningkatkan akurasi model yang terbentuk. Sehingga dapat dikatakan bahwa teknik *stacking ensemble* efektif untuk meningkatkan akurasi klasifikasi model.

Dari beberapa penelitian tersebut di ataslah yang menjadi alasan kuat dan memotivasi kami untuk mengajukan penelitian ini, terutama penelitian yang dilakukan oleh Sukmana et al. (2024) dimana mereka

melakukan penelitian prediksi tsunami dengan menggunakan pendekatan perbandingan algoritma SVM dan RF, dan dataset diambil dari kaggle.com dengan tingkat akurasi yang dihasilkan 65.61% untuk model SVM dan 61.15% untuk model RF, dan juga penelitian yang dilakukan oleh Daza et al. (2024) dimana mereka melakukan penelitian klasifikasi penyakit diabetes dengan metode *stacking ensemble* yang menghasilkan akurasi tertinggi pada skenario 1A yakni 91,5%. Disamping itu, kami melihat belum ada penelitian tentang klasifikasi tsunami dengan teknik *stacking ensemble* ini. Oleh karenanya dengan dataset yang sama yang digunakan peneliti Sukmana et al. (2024), kami akan melakukan penelitian ini untuk meningkatkan akurasi klasifikasi tsunami tersebut dengan metode *stacking ensemble* yang sudah terbukti dapat meningkatkan akurasi.

Selain itu juga kami melihat bahwa dari beberapa penelitian klasifikasi tsunami diatas terjadi kesenjangan yang mencolok dimana akurasi penelitian yang dilakukan Sukmana dan kawan-kawan (2024) sangat rendah sekali yakni hanya 65,16%, sedangkan penelitian yang dilakukan oleh Saputra & Hakim (2022) mencapai akurasi 96 %. Hal ini jugalah yang menjadi problematika kami untuk membuktikan dan menganalisis tentang penelitian klasifikasi tsunami ini.

2. Metode Penelitian

2.1. Jenis, Sifat dan Pendekatan Penelitian

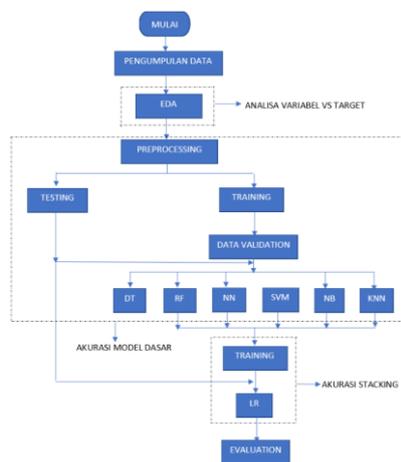
Penelitian ini merupakan jenis penelitian eksperimen yang memiliki sifat deskriptif. Pendekatan yang digunakan dalam penelitian ini adalah pendekatan kuantitatif, yang akan melibatkan proses analisis, implementasi, dan evaluasi model klasifikasi menggunakan algoritma *stacking ensemble* machine learning. Jenis penelitian eksperimen ini dipilih untuk memungkinkan analisis yang mendalam terkait dengan kinerja algoritma-algoritma tersebut dalam konteks klasifikasi tsunami pasca gempa bumi.

2.2. Pengumpulan Data

Data gempa bumi ini berasal dari *open dataset* yang disediakan di *platform* kaggle.com dalam format file csv dengan cara mendownloadnya via internet yang merupakan data tsunami pasca gempa bumi yang terjadi diseluruh dunia dari tahun 2001 sampai tahun 2023.

2.3. Alur Penelitian

Pada tahapan ini dijabarkan langkah-langkah yang dilakukan dalam menyusun dan mengerjakan penelitian. Beberapa tahapan yang dilakukan antara lain pengumpulan data, EDA, preprocessing data, pembagian dataset, membuat model dasar DT, RF, NN, SVM, NB, KNN, membuat meta model, dan evaluasi model. Tahapan tersebut dapat dilihat pada Gambar 2.



Gambar 2. Diagram alur penelitian.

2.4. Exploratory Data Analysis (EDA)

Exploratory Data Analysis adalah langkah awal dalam proses analisis data yang bertujuan untuk memahami struktur, pola, dan hubungan dalam dataset sebelum melakukan analisis lebih lanjut atau membangun model machine learning (Bell, 2014). EDA membantu dalam membangun pemahaman mendalam tentang data, mengidentifikasi potensi masalah, dan menentukan langkah-langkah berikutnya dalam analisis (Coelho & Richert, 2015).

Setelah Dataset diunduh dari Kaggle.com yaitu data tsunami pasca gempa bumi yang terjadi dari tahun 2001 sampai dengan tahun 2023 maka dilakukan Analisa data dengan menggunakan bahasa program python.

Adapun informasi dari dataset yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Informasi dataset

#	Column	Non-Null Count	Dtype
0	title	782 non-null	object
1	magnitude	782 non-null	float64
2	date_time	782 non-null	object
3	cdi	782 non-null	int64
4	mmi	782 non-null	int64
5	alert	415 non-null	object
6	tsunami	782 non-null	int64
7	sig	782 non-null	int64
8	net	782 non-null	object
9	nst	782 non-null	int64
10	dmin	782 non-null	float64
11	gap	782 non-null	float64
12	magType	782 non-null	object
13	depth	782 non-null	float64
14	latitude	782 non-null	float64
15	longitude	782 non-null	float64

16	location	777 non-null	object
17	continent	206 non-null	object
18	country	484 non-null	object

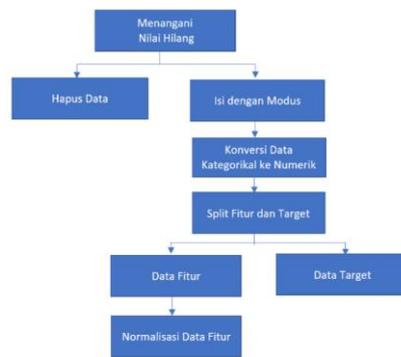
Dari Tabel 1 dapat lihat bahwa dataset terdiri dari 782 baris dan 19 kolom. Mempunyai tipe data yang berbeda yakni tipe data kategorikal (object) dan tipe data numerik (int64, float64).

Pada dataset terdapat data yang hilang, yakni terdapat 367 baris data pada fitur "alert", 5 pada fitur "location", 576 pada fitur "continent" dan 298 pada fitur "country" yang hilang. Kita dapat melihat sebagian data dari dataset yang ditampilkan pada halaman lampiran Tabel 2. Untuk mengetahui hubungan antara fitur dengan target maka digunakan matrik korelasi yang dapat dilihat pada halaman lampiran Tabel 3.

Visualisasi data dalam EDA sangat penting karena memungkinkan kita melihat pola, tren, dan hubungan antar variabel dengan jelas. Ini membantu mengidentifikasi distribusi data, outliers, dan missing data, serta mempermudah komunikasi hasil analisis (Bonaccorso, 2017). Penyebaran lokasi tsunami berdasarkan latitude dan longitude dapat dilihat pada Gambar 3 halaman Lampiran. Visualisasi hubungan antara data fitur numerik dan target dapat dilihat pada Gambar 4 halaman Lampiran. Jumlah terjadi tsunami dan tidak terjadi tsunami dari seluruh data dapat ditentukan bahwa terdapat 304 terjadi tsunami dan 478 tidak terjadi tsunami.

2.5. Preprocessing

Diagram alur preprocessing yang dilakukan dapat dilihat pada Gambar 5.



Gambar 5. Diagram alur preprocessing

Preprocessing data adalah langkah penting dalam setiap proyek analisis data dan machine learning. Tujuannya adalah untuk mempersiapkan data mentah menjadi bentuk yang sesuai dan siap digunakan oleh algoritma machine learning. Preprocessing memastikan bahwa data bersih, berkualitas tinggi, dan relevan untuk model yang akan dibangun (Cholissodin & Soebroto, 2019).

Commented [IF1]: Menurut saya, bagian ini masih menjadi bagian dari bab Metode, karena menjelaskan tahapan dari proses EDA yg dilakukan. Bagian ini juga mendukung penjelasan dari Gambar 2. Sehingga rekomendasi susunan bab 2 sbb: 2.1-2.3 sama seperti saat ini ditambah 2.4. EDA, 2.5. Preprocessing, 2.6 Klasifikasi dan 2.7. Evaluasi

Commented [ss2R1]: Sudah di sesuaikan dengan arahan diatas

Commented [IF3]: Dari penjelasan ini menurut saya sudah cukup untuk menginformasikan jumlah data dari tiap kelas, sehingga tidak perlu disajikan kode program python dan grafiknya Gambar 5 di bawah ini.

Commented [ss4R3]: Sudah dihilangkan kode python dn grafik gambar 5 tersebut

Commented [IF5]: Menurut saya, bagian ini masih menjadi bagian dari bab Metode

Commented [ss6R5]: Sudah dimasukan ke bab Metode

Dari informasi yang disajikan pada EDA, diketahui bahwa ada data yang hilang. Untuk menangani data yang hilang ini maka dapat dilakukan dengan beberapa cara yakni menghapus baris data atau mengisi data yang hilang dengan rata-rata, median atau modus. Tentu jika menghapus baris data yang hilang maka dapat menghilangkan informasi yang mungkin penting (Bell, 2014).

Oleh karenanya peneliti menangani data yang hilang tersebut dengan mengisi data dengan modus, yakni data yang paling sering muncul. Hal ini dikarenakan data yang hilang bertipe kategorikal.

Algoritma *machine learning*, memerlukan data numerik untuk melakukan proses pelatihan dan prediksi. Oleh karena itu, data kategorikal perlu diubah menjadi format numerik agar bisa digunakan dalam model klasifikasi. Dengan bantuan program python hal ini mudah dilakukan.

Dalam proses pengembangan model *machine learning*, memisahkan fitur dan target serta melakukan normalisasi adalah dua langkah krusial yang memastikan data dipersiapkan dengan baik untuk analisis dan pelatihan model (Coelho & Richert, 2015). Pemisahan fitur dan target serta normalisasi data dalam *preprocessing* klasifikasi bertujuan untuk memastikan model pembelajaran mesin dilatih secara efektif dan menghasilkan prediksi yang akurat. Pemisahan fitur (input) dari target (output) penting untuk mencegah kebocoran data, di mana model bisa secara tidak sengaja mempelajari informasi dari target selama pelatihan, yang dapat menyebabkan hasil yang menyesatkan. Normalisasi data, di sisi lain, bertujuan untuk mengatur skala fitur agar memiliki rentang nilai yang serupa, sehingga model dapat lebih mudah mengenali pola tanpa dipengaruhi oleh perbedaan skala antar fitur. Langkah-langkah ini memastikan bahwa model berfokus pada hubungan yang tepat antara input dan output serta meningkatkan stabilitas dan kinerja model dalam prediksi data baru. Dengan bantuan program python maka pemisahan fitur dan target serta normalisasi fitur dapat dengan mudah dilakukan. Hasilnya dapat dilihat pada halaman lampiran Tabel 3.

2.6. Klasifikasi

Memisahkan data menjadi set pelatihan (*training set*) dan set pengujian (*testing set*) sangat penting dalam proses *machine learning* untuk memastikan bahwa model yang kita bangun dapat memberikan prediksi yang akurat dan dapat diandalkan pada data yang belum pernah dilihat sebelumnya (Permana et al., 2023).

Pada penelitian ini, dataset dipisah menjadi dua bagian dengan perbandingan 80% untuk data training dan 20% untuk data testing. Tujuan dari pemisahan ini adalah untuk memastikan bahwa model yang dibangun dapat mempelajari pola dari data training dan kemudian diuji pada data testing yang belum pernah dilihat sebelumnya. Hal ini memungkinkan

evaluasi yang objektif terhadap kinerja model dalam kondisi yang menyerupai dunia nyata, memastikan bahwa model tidak hanya bekerja baik pada data yang sudah dikenalnya (*overfitting*) tetapi juga mampu generalisasi dan memberikan prediksi yang akurat pada data baru.

Langkah selanjutnya adalah menentukan model dasar dan model meta. Model dasar yang digunakan adalah Decision Tree (DT), Random Forest (RF), Neural Network (NN), Support Vektor Machine (SVM), Naïve Bayes (NB), K Nearest Neighbors (KNN). Sedangkan model meta yang digunakan adalah Logistic Regression (LR). Setiap model dasar ditraining dengan menggunakan data training dan ditesing dengan menggunakan data testing. Hasil prediksi dari model dasar kemudian dijadikan sebagai variabel masukan model meta untuk menghasilkan prediksi akhir.

2.7. Evaluation

Peneliti menggunakan K-Fold Cross Validation dengan $k=5$ untuk mengevaluasi kinerja model secara lebih akurat dan reliabel. Teknik ini membagi dataset menjadi 5 subset atau fold, lalu melatih model 5 kali, setiap iterasi menggunakan 4-fold sebagai data pelatihan dan 1-fold sebagai data pengujian. Dengan melakukan ini, peneliti dapat memastikan bahwa hasil evaluasi model tidak tergantung pada pembagian data tertentu, sehingga mengurangi risiko *overfitting* dan memberikan estimasi kinerja model yang lebih stabil dan dapat dipercaya.

Pemilihan $k=5$ didasarkan pada pertimbangan keseimbangan antara bias dan varians dalam evaluasi model. Dalam praktiknya, $k=5$ sering digunakan karena memberikan hasil yang cukup akurat tanpa memerlukan terlalu banyak komputasi, serta secara empiris telah terbukti memberikan keseimbangan yang baik antara stabilitas dan efisiensi dalam berbagai studi sebelumnya. Menurut Kuhn & Johnson (2013), nilai $k=5$ atau $k=10$ adalah pilihan umum yang memberikan estimasi yang baik untuk performa model tanpa meningkatkan kompleksitas secara signifikan.

Setelah menyelesaikan tahap *cross-validation* dan *training* serta testing model, langkah berikutnya melibatkan analisis lebih mendalam terhadap hasil evaluasi untuk memastikan model berkinerja optimal dan siap digunakan. Analisis hasil evaluasi dilakukan dengan memeriksa metrik kinerja seperti akurasi, presisi, recall, dan F1-score baik pada data testing maupun hasil *cross-validation* untuk memastikan konsistensi kinerja model dan mengidentifikasi adanya *overfitting* atau *underfitting*.

Evaluasi metrik kinerja dilakukan pada setiap model dasar yakni Decision Tree (DT), Random Forest (RF), Neural Network (NN), Support Vektor Machine (SVM), Naïve Bayes (NB), K Nearest Neighbors (KNN). Lalu dilakukan evaluasi metrik kinerja dari model *stacking ensemble* dengan

Commented [IF9]: Menurut kami screenshot proram tidak perlu ditampilkan disini.

Commented [ss10R9]: Sudah dihilangkan screenshot program

Commented [IF7]: Menurut saya, bagian ini masih menjadi bagian dari bab Metode

Commented [ss8R7]: Sudah dimasukan ke bagian bab Metode

menggunakan model Logistic Regression (LR) sebagai meta model.

3. Hasil dan Pembahasan

[Perbandingan metrik kinerja model-model tersebut dapat dilihat pada Tabel 4.

Tabel 4. [Perbandingan metrik kinerja model]

No	Model	Accuracy	Precision	Recall	F1-Score
1	DT	0.90	0.91	0.92	0.92
2	RF	0.90	0.93	0.93	0.92
3	NN	0.84	0.88	0.94	0.86
4	SVM	0.83	0.91	0.95	0.84
5	NB	0.78	0.85	0.96	0.80
6	KNN	0.84	0.88	0.97	0.86
7	Stacking	0.94	0.96	0.93	0.94

Pada Tabel 4 tersebut menunjukkan perbandingan hasil antara berbagai metode klasifikasi tsunami. Model dasar seperti Decision Tree (DT), Random Forest (RF), Neural Network (NN), Support Vector Machine (SVM), Naive Bayes (NB), dan K-Nearest Neighbors (KNN) dievaluasi terlebih dahulu. Sebagai contoh diperoleh hasil RF mencapai akurasi sebesar 90%, SVM mencapai akurasi 83%, namun saat teknik *stacking ensemble* diterapkan dengan Logistic Regression sebagai meta-model, akurasi meningkat menjadi 94%. Peningkatan ini menunjukkan bahwa kombinasi berbagai model dapat mengatasi kelemahan masing-masing model dasar, sehingga menghasilkan prediksi yang lebih akurat.

Peningkatan kinerja ini dicapai karena teknik *stacking ensemble* mampu mengintegrasikan keunggulan dari setiap model dasar. Sebagai contoh, Random Forest unggul dalam menangani data dengan banyak variabel, sedangkan Support Vector Machine lebih efektif untuk data dengan kelas-kelas yang terpisah dengan jelas. Kombinasi ini menghasilkan model yang lebih robust dan mampu memberikan hasil prediksi yang lebih akurat (RE & Valentini, 2012).

Evaluasi kinerja model dilakukan menggunakan teknik K-Fold Cross Validation dengan k=5, yang membantu mendapatkan hasil yang lebih akurat dan reliabel, serta mengurangi risiko overfitting. Teknik ini memungkinkan estimasi performa model yang lebih stabil, mendukung penelitian-penelitian terbaru yang menunjukkan bahwa K-Fold Cross Validation dengan nilai k antara 5 dan 10 adalah optimal untuk banyak kasus klasifikasi (Kuhn & Johnson, 2013).

Implikasi dari hasil penelitian ini sangat penting dalam konteks mitigasi bencana. Akurasi prediksi yang lebih tinggi dapat meningkatkan keandalan sistem peringatan dini tsunami, yang pada gilirannya dapat membantu mengurangi risiko dan dampak bencana. Penelitian ini memberikan kontribusi

penting dalam pengembangan sistem prediksi yang lebih akurat dan efisien, serta memperkaya literatur terkait teknik ensemble dalam machine learning. Penelitian lebih lanjut dapat difokuskan pada pengujian teknik ini pada dataset yang lebih besar atau berbeda untuk menguji generalisasi model.

4. Kesimpulan

Penelitian ini menunjukkan bahwa teknik *stacking ensemble* machine learning mampu memberikan peningkatan akurasi dalam klasifikasi kejadian tsunami yang disebabkan oleh gempa bumi. Metode *stacking ensemble* yang menggabungkan beberapa model pembelajaran mesin terbukti efektif dalam menghasilkan prediksi yang lebih akurat dibandingkan dengan model pembelajaran mesin individual. Model *stacking ensemble* ini menggunakan Logistic Regression sebagai meta model yang mengintegrasikan output dari model dasar seperti Decision Tree, Random Forest, Neural Network, Support Vector Machine, Naive Bayes, dan K Nearest Neighbors.

Dari hasil evaluasi, model *stacking ensemble* menunjukkan kinerja yang lebih unggul dalam metrik akurasi, presisi, recall, dan F1-score dengan masing-masing nilai sebesar 94%, 96%, 93% dan 94%. Hal ini menunjukkan bahwa metode ini mampu menangani kompleksitas dan variasi data tsunami pasca gempa bumi dengan lebih baik.

Penelitian ini memberikan kontribusi penting dalam pengembangan sistem peringatan dini tsunami yang lebih akurat dan dapat diandalkan. Implementasi teknik *stacking ensemble* dapat membantu dalam mitigasi risiko dan perlindungan masyarakat dari ancaman tsunami. Untuk penelitian selanjutnya, disarankan untuk mengintegrasikan data real-time dan mengembangkan model yang dapat melakukan prediksi secara dinamis dengan memperhitungkan faktor-faktor lingkungan lainnya.

Daftar Pustaka:

- Adnan, M., Imam, M. O., Javed, M. F., & Murtza, I. (2024). Improving spam email classification accuracy using ensemble techniques: a stacking approach. *International Journal of Information Security*, 23(1), 505–517. <https://doi.org/10.1007/s10207-023-00756-1>
- Bell, J. (2014). *Machine Learning: Hands-On for Developers and Technical Professionals*. John Wiley & Sons, Inc.
- Bonaccorso, G. (2017). *Machine Learning Algorithms*. Packt Publishing Ltd.
- Cholissodin, I., & Soebroto, A. A. (2019). *Ai , Machine Learning & Deep Learning (Teori & Implementasi)* (Issue July 2019).
- Coelho, L., & Richert, W. (2015). *Building Machine Learning Systems with Python* (Second). Packt Publishing.
- Daza, A., Sancez, Caorlos Fidel Ponce, Apaza-Perez,

Commented [IF11]: Apakah tampilan di bawah ini sudah bisa terwakili di Tabel 4, menurut hemat kami jika memang intinya sudah terwakili di Tabel 4 screenshot di bawah ini juga bisa untuk tidak dimasukkan

Commented [ss12R11]: Tabel 4. Sudah mewakili screenshot tsb dan sudah dihilangkan

Commented [IF13]: Perlu diformat dalam bentuk tabel, dengan font isi tabel 8pt. Jad kami sarankan bukan berupa tabel yang discreenshot. Diperlukan juga pembahasan dari hasil ini. Bisa dibandingkan hasilnya dari tiap metode, dan bagaimana hal itu bisa didapatkan, bisa disertakan juga sitasi dari referensi jika ada yang mendukung statement yang digunakan. Bagian ini yang menurut kami lebih tepat untuk dimasukkan ke dalam bab Hasil dan Pembahasan.

Commented [ss14R13]: Tabel sudah diganti tidak lagi berupa screenshot. Sudah dibuat narasi hasil pembahasan sesuai adviceny

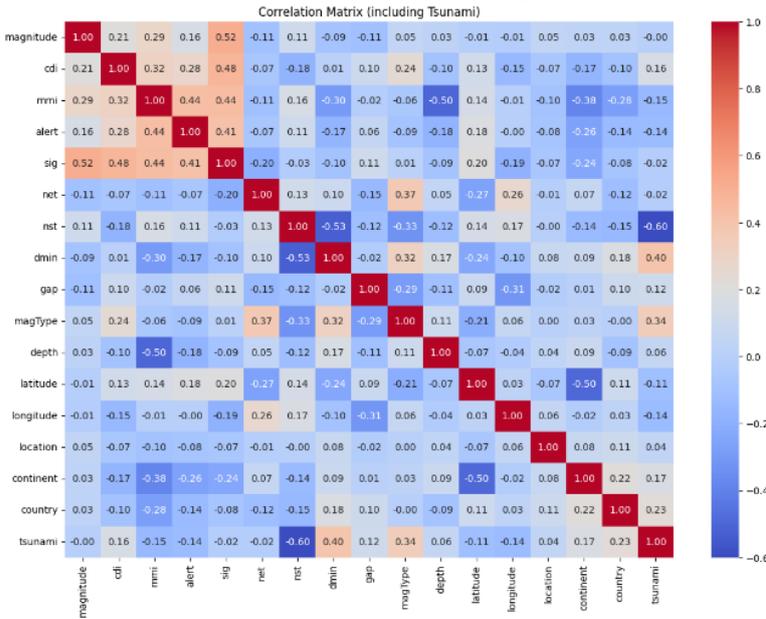
- G., Pinto, J., & Ramos, K. Z. (2024). *Stacking ensemble approach to diagnosing the disease of diabetes*. 44(December 2023). <https://doi.org/10.1016/j.imu.2023.101427>
- Irawan Saputra, D., & Hakim, D. L. (2022). Implementasi Algoritma Gaussian Naive Bayes Classifier Untuk Prediksi Potensi Tsunami Berbasis Mikrokontroler. *EPSILON: Journal of Electrical Engineering and Information Technology*, 20(2), 122–138. <https://doi.org/10.55893/epsilon.v20i2.94>
- Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. In *Applied Predictive Modeling*. <https://doi.org/10.1007/978-1-4614-6849-3>
- Latue, T., & Latue, P. C. (2023). Analisis Spasial Prediksi Bahaya Tsunami di Kecamatan Salahutu Pulau Ambon. *Buana Jurnal Geografi, Ekologi Dan Kebencanaan*, 1(1), 21–30. <https://doi.org/10.56211/buana.v1i1.342>
- Mateen, S., Nuthammachot, N., & Techato, K. (2024). Random forest and artificial neural network-based tsunami forests classification using data fusion of Sentinel-2 and Airbus Vision-1 satellites: A case study of Garhi Chandan, Pakistan. *Open Geosciences*, 16(1), 1–25. <https://doi.org/10.1515/geo-2022-0595>
- Permana, A., S, W., Santoso, W., & Wibowo, G. (2023). MACHINE LEARNING. In A. Y. M.Pd (Ed.), *Machine Learning* (Vol. 45, Issue 13). PT GLOBAL EKSEKUTIF TEKNOLOGI.
- RE, M., & Valentini, G. (2012). *Ensemble Methods*. <https://doi.org/10.1201/b11822-34>
- Sukmana, H. T., Durachman, Y., Amri, & Supardi. (2024). Comparative Analysis of SVM and RF Algorithms for Tsunami Prediction: A Performance Evaluation Study. *Journal of Applied Data Sciences*, 5(1), 84–99. <https://doi.org/10.47738/jads.v5i1.159>
- Syamsiah, N. O., & Purwandani, I. (2023). *Penerapan Metode Stacking Untuk Meningkatkan Akurasi Hasil Peramalan Konsumsi Listrik Abstrak*. 4(1), 15–25.
- Wang, Y., Imai, K., Miyashita, T., Ariyoshi, K., Takahashi, N., & Satake, K. (2023). Coastal tsunami prediction in Tohoku region, Japan, based on S-net observations using artificial neural network. *Earth, Planets and Space*. <https://doi.org/10.1186/s40623-023-01912-6>
- Zhou, Z.-H. (2012). *Ensemble Methods Foundations and Algorithms* (Second Edn). Taylor & Francis Group. <https://doi.org/10.1201/b11822-34>

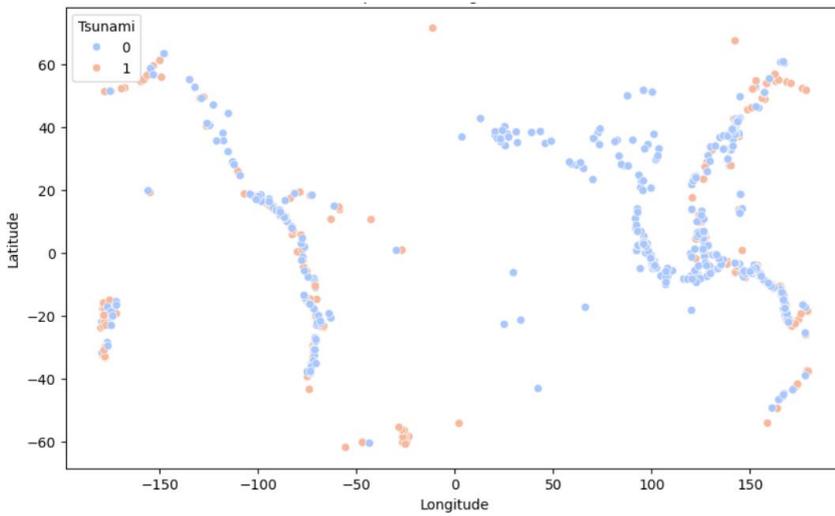
Lampiran

Tabel 2. Dataset Tsunami Pasca gempa bumi

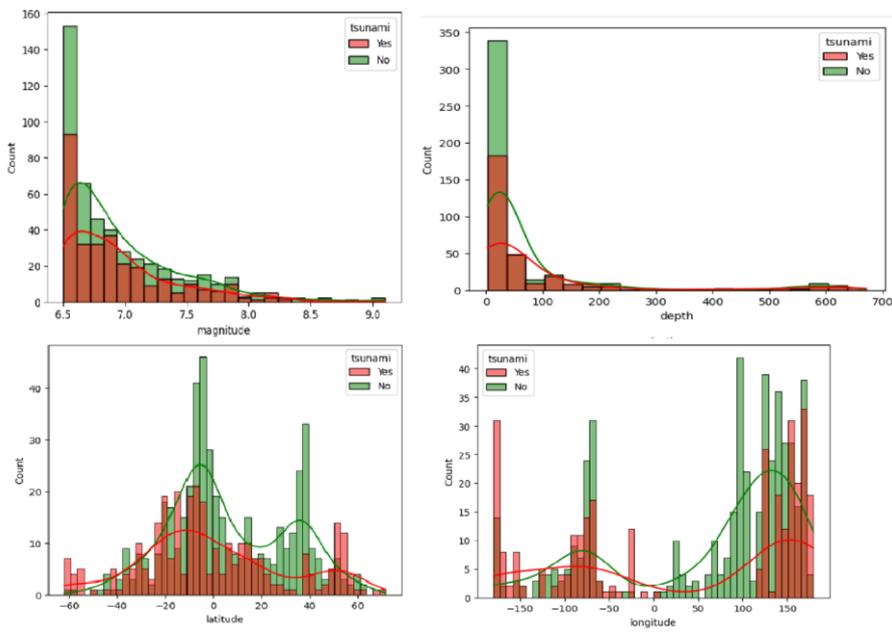
Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Column13	Column14	Column15	Column16	Column17	Column18	Column19
time	magnitude	date_time	lat	lon	alert	tsunami	sig	net	rst	dmin	gap	magType	depth	latitude	longitude	location	continent	country
M7.0 - 18 km SW of Mangrove, Solomon Islands	7	01-11-2022 010	8	7	green	1	768	us	127	6509	17	mw	14	-9.790	155.956	Makungu, Solomon Islands	Oceania	Solomon Islands
M6.9 - 204 km SE of Bengkulu, Indonesia	6.9	8-11-2022 113	4	4	green	0	735	us	99	2329	34	mw	25	-4.959	100.738	Bengkulu, Indonesia	Asia	Indonesia
M7.0 - 7	7	2-11-2022 010	3	3	green	1	750	us	167	1125	18	mw	139	-02.058	-119.385		Oceania	Fiji
M7.3 - 205 km ESE of Naitua, Tonga	7.3	1-11-2022 104	5	5	green	1	833	us	149	1465	21	mw	37	-15.2618	-172.129	Naitua, Tonga	Oceania	Tonga
M6.6 - 66	6.6	8-11-2022 101	0	2	green	1	670	us	131	4398	27	mw	634.604	-25.9468	178.728		Oceania	Fiji Islands
M7.0 - south of the Fiji Islands	7	8-11-2022 095	4	3	green	1	755	us	162	4578	26	mw	660	-26.0462	178.816		Oceania	Fiji Islands
M6.8 - south of the Fiji Islands	6.8	8-11-2022 093	1	3	green	1	711	us	136	4678	22	mw	630.379	-25.9078	178.162		Oceania	Fiji Islands
M6.7 - 60 km SW of Boca Chica, Panama	6.7	15-10-2022 115	7	6	green	1	787	us	145	1151	37	mw	20	7.612	-83.3368	Boca Chica, Panama	North America	Panama
M6.8 - 10 km SW of Aguajalita, Mexico	6.8	0-09-2022 061	8	7	yellow	1	1179	us	175	2137	92	mw	20	18.133	-102.913	Aguajalita, Mexico	North America	Mexico
M7.0 - 30 km SW of Aguajalita, Mexico	7.0	0-09-2022 180	9	8	yellow	1	1799	us	271	2133	68	mw	26.943	18.2667	-103.202	Aguajalita, Mexico	North America	Mexico
M6.9 - 90 km E of Taipei, Taiwan	6.9	8-09-2022 064	9	9	yellow	1	887	us	152	6463	34	mw	10	23.144	121.307	Yuliang, Taiwan	Asia	Taiwan
M6.5 - 96 km SE of Luga, Taiwan	6.5	7-09-2022 114	7	7	green	1	756	us	178	643	54	mw	10	23.629	121.148	Luga, Taiwan	Asia	Taiwan
M7.0 - 200 km SE of Kunging, Vanuatu	7	4-09-2022 110	7	5	green	1	761	us	162	3377	45	mw	137	-21.2077	-152.918	Kunging, Vanuatu	Oceania	Vanuatu
M7.6 - 70 km E of Karamu, Papua New Guinea	7.6	0-09-2022 214	8	8	yellow	1	945	us	272	3158	12	mw	116	-6.2237	164.671	Karamu, Papua New Guinea	Oceania	Papua New Guinea
M6.6 - 48 km SE of Kunging, China	6.6	0-09-2022 065	9	8	orange	0	5653	us	142	4454	34	mw	12	29.230	122.778	Kunging, China	Asia	People's Republic of China
M6.6 - south of the Kermadec Islands	6.6	4-08-2022 134	7	6	green	1	672	us	68	5263	34	mw	30	52.6922	-178.959		Oceania	Kermadec Islands
M7.0 - 11 km NE of Bantay, Philippines	7	07-07-2022 004	9	8	yellow	1	1351	us	102	5276	22	mw	33.778	17.5878	120.809	Bantay, Philippines	Asia	Philippines
M6.5 - 107 km SW of Arequipa, Brazil	6.5	8-06-2022 005	3	2	green	0	933	us	236	1998	31	mw	628.371	-46.618	-73.1667	Tanauca, Brazil	South America	Brazil
M7.2 - 10 km W of Arequipa, Peru	7.2	0-06-2022 120	7	5	green	1	876	us	146	2494	40	mw	236	-14.8628	-70.3381	Arequipa, Peru	South America	Peru
M6.8 - 68	6.8	9-09-2022 011	2	5	green	1	733	us	137	6371	45	mw	10	54.3335	239.637		Oceania	Kermadec Islands
M6.8 - 86 km NNW of San Antonio de los Cabes	6.8	0-06-2022 210	6	5	green	1	742	us	162	1505	30	mw	220	-23.6141	-64.7285	San Antonio de los Cabes, South America	South America	Argentina
M6.6 - 55 km SW of Masapaga, Nicaragua	6.6	15-08-2022 074	6	5	green	1	742	us	0	6314	36	mw	27	11.5138	-86.9919	Masapaga, Nicaragua	North America	Nicaragua
M7.0 - southeast of the Loyalty Islands	7	11-01-2022 014	6	4	green	1	763	us	0	2705	26	mw	10	-22.1712	171.345		Oceania	Loyalty Islands
M6.9 - 284 km ESE of Taitene, New Caledonia	6.9	0-03-2022 205	6	4	green	1	738	us	0	2497	42	mw	10	-22.72	170.727	Taitene, New Caledonia	Oceania	New Caledonia
M6.7 - 65 km E of Hualien City, Taiwan	6.7	0-01-2022 114	8	7	green	1	806	us	0	6399	32	mw	24	23.1421	121.676	Hualien City, Taiwan	Asia	Taiwan
M7.3 - 17 km ENE of Niumic, Japan	7.3	6-03-2022 143	9	8	red	1	2397	us	0	2936	29	mw	41	37.7015	141.587	Niumic, Japan	Asia	Japan
M6.7 - 104 km W of Pacuman, Indonesia	6.7	3-10-2022 110	9	6	green	0	708	us	0	2188	43	mw	28	-0.6811	98.3618	Pacuman, Indonesia	Asia	Indonesia
M6.8 - south of the Fiji Islands	6.8	0-02-2022 202	2	3	green	1	712	us	0	578	12	mw	535	-23.7862	-179.968		Oceania	Fiji Islands
M6.5 - 68 km NW of Arequipa, Peru	6.5	0-06-2022 155	8	6	green	0	496	us	0	1036	22	mw	130	-44.91	-70.9395	Arequipa, Peru	South America	Peru
M6.5 - Kermadec Islands region	6.5	9-01-2022 024	7	4	green	1	651	us	0	1088	17	mw	8	29.535	-176.729	Kermadec Islands region	Oceania	Kermadec Islands
M6.6 - 80 km SW of Labuan, Indonesia	6.6	4-02-2022 090	8	6	green	0	787	us	0	2438	22	mw	33	-6.8261	103.221	Labuan, Indonesia	Asia	Indonesia
M6.5 - 71 km SE of Nihsiki, Alaska	6.5	1-01-2022 123	0	3	green	1	650	us	97	161607	108	M	37	52.5102	-148.08	Nihsiki, Alaska	North America	Alaska
M6.5 - 71 km SE of Nihsiki, Alaska	6.5	1-01-2022 123	0	3	green	1	650	us	97	161607	108	M	37	52.5102	-148.08	Nihsiki, Alaska	North America	Alaska
M6.6 - 31 km SE of Nihsiki, Alaska	6.6	1-01-2022 121	8	5	green	1	674	us	0	6107	49	mw	19	52.6282	-148.12	Nihsiki, Alaska	North America	Alaska
M6.7 - 91 km SE of Nihsiki, Alaska	6.7	1-01-2022 113	0	4	green	1	691	us	30	931943	136	M	38	52.648	-147.736	Nihsiki, Alaska	North America	Alaska
M6.7 - 51 km SE of Nihsiki, Alaska	6.7	1-01-2022 113	0	4	green	1	691	us	30	931943	136	M	38	52.648	-147.736	Nihsiki, Alaska	North America	Alaska
M6.8 - 100 km SE of Nihsiki, Alaska	6.8	1-01-2022 113	7	6	green	1	730	us	0	692	61	mw	20	52.6563	-147.917	Nihsiki, Alaska	North America	Alaska
M6.6 - 48 km WNW of Pitsi, Cyprus	6.6	1-01-2022 010	6	6	green	0	976	us	0	1168	21	mw	21	35.1456	31.9659	Pitsi, Cyprus	Asia	Cyprus

Tabel 3. Matrik korelasi antar variabel dan target





Gambar 3. Penyebaran tsunami pasca gempa bumi berdasarkan latitude dan longitude



Gambar 4. Hubungan Data Numerik dengan Target Tsunami

Tabel 3. Normalisasi Fitur

	magnitude	cdi	mmi	alert	sig	net	nst	dmin	gap	magType	depth	latitude	longitude	location
0	7.0	8	7	0	768	9	117	0.509	17.0	8	14.000	-9.7963	159.596	212
1	6.9	4	4	0	735	9	99	2.229	34.0	8	25.000	-4.9559	100.738	48
2	7.0	3	3	0	755	9	147	3.125	18.0	8	579.000	-20.0508	-178.346	163
3	7.3	5	5	0	833	9	149	1.865	21.0	8	37.000	-19.2918	-172.129	253
4	6.6	0	2	0	670	9	131	4.998	27.0	8	624.464	-25.5948	178.278	163
...
777	7.7	0	8	0	912	9	427	0.000	0.0	7	60.000	13.0490	-88.660	307
778	6.9	5	7	0	745	0	0	0.000	0.0	5	36.400	56.7744	-153.281	270
779	7.1	0	7	0	776	9	372	0.000	0.0	6	103.000	-14.9280	167.170	299
780	6.8	0	5	0	711	9	64	0.000	0.0	7	33.000	6.6310	126.899	232
781	7.5	0	7	0	865	9	324	0.000	0.0	7	33.000	6.8980	126.579	202

Halaman ini sengaja dikosongkan