



Analisis prediksi gempa di pulau Sumatra menggunakan metode Random Forest

Raynaldo Syah Pratama¹⁾, Hendra Kurniawan²⁾

*¹⁾Faculty sains data, Institut Informatika dan Bisnis Darmajaya, bandar lampung, indonesia
Email First Author^{1*)}; Email Second Author²⁾; : raynaldosyahpratama@gmail.com,
hendra.kurniawan@darmajaya.ac.id*

Abstract

Penelitian ini bertujuan untuk memprediksi zona rawan gempa di Pulau Sumatera menggunakan algoritma Random Forest. Data yang digunakan mencakup catatan historis gempa bumi dari tahun 2010 hingga 2023, meliputi latitude, longitude, kedalaman, dan magnitudo. Algoritma Random Forest dipilih karena kemampuannya menangani data kompleks dengan akurasi tinggi dan fleksibilitas dalam interpretasi hasil. Model ini menunjukkan performa yang baik dalam mengklasifikasikan zona "Non-Risk Zone" dengan akurasi 74%, dan hasil ini mengindikasikan potensi besar algoritma dalam mendukung analisis data seismik yang kompleks. Penelitian ini memberikan dasar kuat untuk pengembangan lebih lanjut dengan meningkatkan kapasitas model untuk mendeteksi zona "Risk Zone" melalui optimalisasi data. Hasil penelitian diharapkan dapat menjadi acuan strategis dalam pengembangan sistem mitigasi bencana berbasis data yang efektif, membantu pemerintah dan pemangku kebijakan dalam meminimalkan dampak bencana secara signifikan.

Kata kunci : prediksi gempa, Random Forest, zona rawan, Sumatera, mitigasi bencana.

Abstrak

This study aims to predict earthquake-prone zones in Sumatra Island using the Random Forest algorithm. The dataset consists of historical earthquake records from 2010 to 2023, including latitude, longitude, depth, and magnitude. Random Forest was selected for its ability to handle complex datasets with high accuracy and flexibility in result interpretation. The model demonstrated strong performance in classifying "Non-Risk Zone" with an accuracy of 74%, highlighting its potential for supporting complex seismic data analysis. This research lays a solid foundation for further improvements, particularly in enhancing the model's capacity to identify "Risk Zone" through data optimization. The results are expected to serve as a strategic reference in developing effective data-driven disaster.mitigation systems, aiding the government and policymakers in significantly reducing disaster impacts.

Keywords : earthquake prediction, Random Forest, prone zones, Sumatra, disaster mitigation.



1 Pendahuluan

Indonesia merupakan salah satu negara dengan tingkat aktivitas seismik tertinggi di dunia karena berada di pertemuan tiga lempeng tektonik utama: Eurasia, Indo-Australia, dan Pasifik. Pulau Sumatera termasuk wilayah dengan risiko gempa tinggi akibat adanya zona subduksi Megathrust dan Sesar Sumatera. Bencana gempa bumi di wilayah ini tidak hanya menimbulkan kerusakan infrastruktur, tetapi juga dampak sosial dan ekonomi yang signifikan. Berbagai upaya mitigasi dilakukan, termasuk penerapan teknologi pembelajaran mesin (machine learning) untuk memetakan zona rawan gempa secara lebih cepat dan akurat. Studi sebelumnya telah menggunakan berbagai algoritma untuk menganalisis data seismik. Ika Maulita dan Arif Mu'amar Wahid (2024) menunjukkan bahwa algoritma seperti LightGBM unggul dalam memprediksi magnitudo gempa dengan akurasi tinggi berdasarkan kedalaman dan koordinat gempa [1].

Sementara itu, Agustian Noor (2018) membandingkan algoritma SVM konvensional dengan SVM berbasis Particle Swarm Optimization (PSO) menggunakan data gempa dari wilayah Sumatera. Hasil penelitian menunjukkan bahwa SVM konvensional lebih unggul dengan RMSE sebesar 9,720 dibandingkan PSO yang mencapai 37,685. Hal ini mengindikasikan potensi untuk meningkatkan kinerja SVM melalui optimasi algoritma, terutama dalam memodelkan data gempa yang kompleks [2].

Studi lain oleh Ipin Sugiyarto dkk. (2021) memanfaatkan teknik data mining dengan algoritma K-Means untuk mengelompokkan dampak gempa di Provinsi Sumatera Barat. Penelitian tersebut menganalisis data historis gempa dari tahun 2001 hingga 2018, dengan fokus pada tingkat kerusakan fasilitas umum, kerugian ekonomi, serta jumlah korban di berbagai wilayah terdampak. Hasil penelitian menunjukkan bahwa algoritma K-Means mampu menghasilkan kluster yang efektif dalam memetakan wilayah berdasarkan intensitas dampaknya, sehingga daerah dengan tingkat kerusakan paling tinggi, seperti Kota Padang pada gempa tahun 2009, dapat diidentifikasi secara lebih sistematis. Metode ini tidak hanya memberikan informasi yang lebih terorganisasi, tetapi juga mendukung pengambilan keputusan dalam mitigasi bencana, seperti penentuan prioritas penanganan pada daerah yang paling terdampak [3].

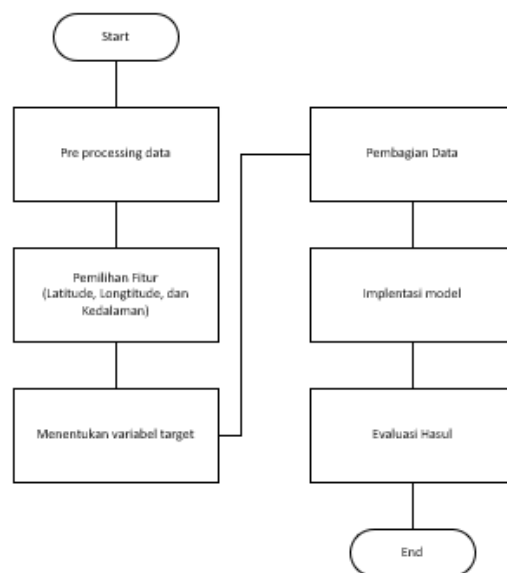
Dalam konteks tersebut, penelitian ini bertujuan mengembangkan model prediksi zona rawan gempa di wilayah Sumatera menggunakan algoritma Random Forest. Data historis gempa bumi, yang mencakup parameter seperti latitude, longitude, kedalaman, dan magnitudo, digunakan untuk membangun model prediktif. Random Forest dipilih karena kemampuannya yang unggul dalam menangani data dengan variasi tinggi dan memberikan hasil klasifikasi yang akurat melalui mekanisme ensemble learning [4]. Urgensi penelitian ini terletak pada kebutuhan akan metode prediksi yang tidak hanya cepat, tetapi juga andal dalam mengolah data seismik yang kompleks dan heterogen, terutama di wilayah dengan tingkat risiko gempa tinggi seperti Sumatera. Hasil



penelitian ini diharapkan dapat memberikan informasi yang akurat untuk mendukung mitigasi risiko bencana berbasis

2 Metode Penelitian

Metodologi penelitian ini dirancang secara sistematis untuk mengembangkan model prediksi zona rawan gempa di wilayah Sumatera menggunakan algoritma Random Forest. Setiap tahap penelitian, mulai dari pengumpulan data hingga pelatihan model, disusun untuk memastikan akurasi, validitas, dan relevansi hasil dalam konteks mitigasi risiko bencana gempa. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1, yang menggambarkan proses mulai dari pengumpulan data, prapemrosesan, hingga pelatihan dan pengujian model.



Gambar 1. Alur Penelitian

2.1 Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini terdiri atas catatan historis gempa bumi di wilayah Pulau Sumatera selama periode tahun 2010 hingga 2023. Dataset ini mencakup informasi penting, antara lain: latitude dan longitude, yang menunjukkan lokasi geografis episentrum gempa; kedalaman (depth), yang menggambarkan posisi vertikal episentrum dari permukaan bumi; serta magnitudo, yang menunjukkan besarnya kekuatan gempa. Informasi ini diambil dari sumber data tepercaya seperti Kaggle, dengan total ribuan entri yang mencakup rentang variasi spasial dan temporal aktivitas seismik. Distribusi data gempa dalam dataset ini sangat heterogen, baik dalam hal lokasi geografis maupun parameter gempa lainnya, sehingga memberikan tantangan tersendiri



dalam pemodelan. Data ini juga telah mengalami proses penyaringan awal untuk memastikan bahwa hanya catatan yang valid dan relevan dengan wilayah Sumatera yang digunakan dalam analisis.

2.2 Pemilihan Fitur

Pemilihan fitur merupakan tahap krusial dalam proses pemodelan untuk memastikan bahwa hanya atribut yang relevan digunakan. Pada penelitian ini, atribut yang dipilih meliputi latitude, longitude, kedalaman, dan magnitudo. Keempat atribut ini dianggap sangat signifikan karena latitude dan longitude membantu menentukan lokasi geografis dari episentrum gempa, yang menjadi dasar pemetaan zona rawan [5]. Kedalaman gempa memiliki pengaruh signifikan terhadap tingkat kerusakan di permukaan, di mana gempa dangkal cenderung memiliki dampak yang lebih besar dibandingkan gempa dalam. Magnitudo juga merupakan indikator utama dalam mengukur kekuatan gempa yang secara langsung berkaitan dengan potensi kerusakan. Pemilihan fitur dilakukan secara manual berdasarkan pengetahuan domain serta kajian terhadap penelitian terdahulu, untuk memastikan bahwa model dapat menangkap hubungan yang bermakna antara atribut dan klasifikasi zona rawan gempa [6].

2.3 Pemilihan Model

Algoritma Random Forest dipilih sebagai model utama untuk penelitian ini. Random Forest merupakan algoritma berbasis ensemble learning yang menggabungkan prediksi dari sejumlah decision trees untuk menghasilkan hasil akhir yang lebih andal [7]. Proses ini dilakukan dengan membuat setiap pohon keputusan (decision tree) menggunakan subset data yang diambil secara acak dengan teknik bootstrap sampling. Salah satu keunggulan Random Forest adalah kemampuannya dalam menangani data yang kompleks dan beragam tanpa memerlukan praproses data yang rumit, seperti normalisasi atau standardisasi [8].

Selain itu, algoritma ini memiliki mekanisme bawaan untuk menangani fitur yang tidak relevan, sehingga dapat meningkatkan akurasi tanpa risiko overfitting [9]. Dengan keunggulan ini, Random Forest sangat cocok untuk dataset gempa bumi yang biasanya memiliki karakteristik beragam dan tidak linear. Selain itu, Random Forest memiliki fleksibilitas dalam penggunaan metrik penting, seperti skor akurasi dan tingkat feature importance, yang dapat membantu dalam interpretasi hasil prediksi [10].

2.4 Pembagian data

Pembagian data dilakukan dengan memisahkan dataset menjadi dua bagian, yaitu 80% data pelatihan untuk membangun model dan 20% data pengujian untuk mengevaluasi performa model. Pemisahan ini bertujuan untuk memastikan bahwa model memiliki kemampuan generalisasi yang baik terhadap data yang belum pernah dilihat sebelumnya. Model kemudian dilatih menggunakan data latih dengan algoritma Random Forest, dan setelah selesai dilatih, data uji digunakan untuk



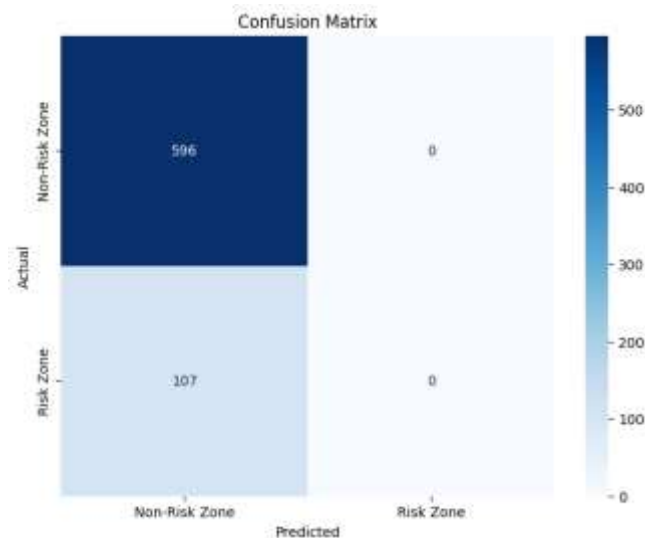
mengevaluasi kemampuan model dalam memprediksi zona rawan gempa serta memahami sejauh mana model dapat memetakan pola risiko berdasarkan karakteristik data.

3 Hasil dan Pembahasan

Penelitian ini menghasilkan model prediksi zona rawan gempa di wilayah Sumatera dengan memanfaatkan algoritma Random Forest sebagai metode utama. Algoritma ini dipilih karena kemampuannya yang unggul dalam menangani data kompleks dan memberikan hasil prediktif yang dapat diinterpretasikan dengan baik. Proses evaluasi model dilakukan menggunakan data uji yang telah dipisahkan dari data pelatihan, yaitu sebesar 20% dari total dataset. Hasil performa model kemudian dianalisis secara menyeluruh melalui penyajian matriks kebingungan (confusion matrix), yang menunjukkan distribusi prediksi dan label aktual, serta melalui metrik-metrik klasifikasi seperti akurasi, presisi, recall, dan f1-score. Evaluasi ini tidak hanya menyoroti kemampuan model secara umum, tetapi juga memberikan wawasan mengenai bagaimana model mampu membedakan zona rawan dan tidak rawan dengan tingkat akurasi yang berbeda. Hal ini menjadi langkah penting dalam mengukur efektivitas algoritma Random Forest untuk aplikasi mitigasi risiko bencana gempa bumi [11].

3.1 Matriks Kebingungan

Analisis performa model Random Forest pada data uji dilakukan untuk mengevaluasi kemampuan klasifikasi terhadap dua kategori utama, yaitu "Non-Risk Zone" dan "Risk Zone." Evaluasi ini bertujuan untuk mengidentifikasi kekuatan dan kelemahan model dalam memetakan zona rawan gempa serta memahami distribusi hasil prediksi terhadap label sebenarnya. Hasil evaluasi disajikan dalam bentuk matriks kebingungan yang dijelaskan pada Gambar 2.



Gambar 2.Matriks Kebingungan



Berdasarkan Gambar 2 matriks kebingungan menunjukkan bahwa model mampu mengklasifikasikan zona "Non-Risk Zone" (zona tidak rawan) dengan baik, dengan total 596 prediksi benar dari 703 data uji. Namun, performa model untuk kategori "Risk Zone" (zona rawan) masih kurang optimal, dengan semua data di kategori ini diklasifikasikan sebagai "Non-Risk Zone". Hal ini mengindikasikan bahwa model belum sepenuhnya mampu mendeteksi zona rawan gempa.

3.2 Evaluasi Metrik Klasifikasi

Selain matriks kebingungan, evaluasi performa model juga dilengkapi dengan metrik klasifikasi lainnya, yang meliputi akurasi, presisi, recall, dan f1-score untuk masing-masing kategori. Metrik-metrik ini bertujuan untuk memberikan gambaran yang lebih rinci mengenai sejauh mana model mampu melakukan klasifikasi secara akurat dan seimbang antara kategori "Non-Risk Zone" dan "Risk Zone". Informasi ini disajikan pada Tabel evaluasi yang terlihat pada tabel 1.

Accuracy	Precision	F-1 score	Recall
74%	87%	84%	82%

Tabel 1. Tabel kinerja model

Tabel 1 memberikan informasi lebih lanjut mengenai performa model. Akurasi keseluruhan model tercatat sebesar 74%, yang mencerminkan kemampuan model dalam melakukan prediksi secara umum. Namun, presisi untuk kategori "Risk Zone" hanya mencapai 0.24, dengan recall sebesar 0.31 dan f1-score sebesar 0.27. Nilai-nilai ini menunjukkan bahwa model cenderung bias terhadap kategori "Non-Risk Zone", sehingga kemampuannya untuk mendeteksi zona rawan masih rendah.

3.3 Pembahasan

Hasil ini mengungkapkan beberapa aspek penting dalam penggunaan algoritma Random Forest untuk memprediksi zona rawan gempa. Kemampuan model untuk mengenali zona "Non-Risk Zone" dengan baik menunjukkan bahwa fitur-fitur yang digunakan seperti latitude, longitude,



dan kedalaman cukup relevan. Namun, ketidakseimbangan dalam data, di mana jumlah data pada kategori "Risk Zone" lebih sedikit dibandingkan "Non-Risk Zone", kemungkinan besar menjadi salah satu penyebab rendahnya performa model pada kategori rawan.

Kinerja yang kurang optimal pada kategori "Risk Zone" mengindikasikan perlunya langkah lanjutan dalam penelitian, seperti meningkatkan kualitas data melalui teknik penyeimbangan data seperti SMOTE (Synthetic Minority Over-sampling Technique), pemilihan fitur tambahan, atau eksplorasi algoritma alternatif yang lebih sesuai untuk dataset tidak seimbang. Kombinasi dari strategi ini diharapkan dapat memperbaiki hasil model dan memberikan prediksi yang lebih akurat untuk mendukung mitigasi risiko gempa bumi secara efektif.

4 . Kesimpulan

Penelitian ini berhasil mengembangkan model prediksi zona rawan gempa di Pulau Sumatera dengan menggunakan algoritma Random Forest. Algoritma ini menunjukkan performa yang baik, dengan akurasi sebesar 74% dalam mengklasifikasikan zona "Non-Risk Zone". Hasil tersebut menggaris bawahi potensi Random Forest sebagai metode yang andal untuk mengolah data seismik yang kompleks dan beragam. Penelitian ini juga membuka peluang pengembangan lebih lanjut, khususnya dalam meningkatkan kemampuan model untuk mendeteksi zona "Risk Zone". Optimalisasi data melalui teknik seperti penyeimbangan data atau penambahan fitur baru dapat meningkatkan akurasi dan presisi model secara keseluruhan. Temuan ini tidak hanya relevan dalam konteks mitigasi risiko gempa bumi di Sumatera, tetapi juga berpotensi diterapkan pada wilayah lain dengan tingkat risiko yang serupa. Dengan demikian, penelitian ini memberikan kontribusi signifikan dalam pengembangan sistem mitigasi bencana berbasis data yang efektif. Hasil penelitian ini diharapkan dapat menjadi rujukan bagi pemerintah dan pemangku kebijakan dalam merancang strategi mitigasi yang lebih terukur dan berbasis bukti untuk meminimalkan dampak bencana gempa bumi terhadap masyarakat.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Institut Informatika dan Bisnis Darmajaya atas dukungan dan fasilitas yang diberikan selama proses penelitian ini berlangsung. Ucapan terima kasih khusus disampaikan kepada Bapak Hendra Kurniawan yang telah memberikan bimbingan, masukan, dan semangat yang sangat berarti dalam penyusunan penelitian ini. Penulis juga menyampaikan apresiasi kepada Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) serta platform Kaggle atas ketersediaan data gempa bumi yang menjadi dasar utama dalam analisis penelitian ini. Akhir kata, penulis menghargai segala bentuk bantuan dan dukungan dari berbagai pihak yang telah berkontribusi, baik secara langsung maupun tidak langsung, demi kelancaran dan keberhasilan penelitian ini.



Daftar Pustaka

- [1] I. Sugiyarto, R. Irawan, and D. Rosiyadi, “Pengelompokan Dampak Gempa Bumi Dan Kerusakan Pada Wilayah Berpotensi Gempa Di Provinsi Sumatra Barat,” *Journal of Students’ Research in Computer Science*, vol. 2, no. 2, pp. 211–222, 2021, doi: 10.31599/jsrscs.v2i2.850.
- [2] J. Humaniora, “Support Vector Machin E Biasa Dan Support Vector Machine Berbasis Particle,” vol. 4, pp. 31–37, 2018.
- [3] I. Maulita *et al.*, “Prediksi Magnitudo Gempa Menggunakan Random Forest , Support Vector Regression , XGBoost , LightGBM , dan Multi-Layer Perceptron Berdasarkan Data Kedalaman dan Geolokasi Predicting Earthquake Magnitude Using Random Forest , Support Vector Regression , XGB,” vol. 4, no. 5, pp. 221–232, 2024.
- [4] Y. A. Mustofa, “Pendekatan Ensemble pada Analisis Sentimen Ulasan Aplikasi Google Play Store Ensemble Approach to Sentiment Analysis of Google Play Store App Reviews,” vol. 6, pp. 181–188, 2024.
- [5] L. Irawan, L. H. Hasibuan, and F. Fauzi, “Analisa Prediksi Efek Kerusakan Gempa Dari Magnitudo (Skala Richter) Dengan Metode Algoritma Id3 Menggunakan Aplikasi Data Mining Orange,” *Jurnal Teknologi Informasi: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, vol. 14, no. 2, pp. 189–201, 2020, doi: 10.47111/jti.v14i2.1079.
- [6] S. Siswanto, Ngatono, and S. Febri Saputra, “Prototype Sistem Peringatan Dini Bencana Gempa Bumi Dan Tsunami Berbasis Internet of Things,” *PROSISKO: Jurnal Pengembangan Riset dan Observasi Sistem Komputer*, vol. 9, no. 1, pp. 60–66, 2022, doi: 10.30656/prosisko.v9i1.4743.
- [7] M. Huljanah, Z. Rustam, S. Utama, and T. Siswantining, “Feature Selection using Random Forest Classifier for Predicting Prostate Cancer,” *IOP Conf Ser Mater Sci Eng*, vol. 546, no. 5, 2019, doi: 10.1088/1757-899X/546/5/052031.
- [8] M. Ardiansyah, “Model Ensemble Algoritma Naive Bayes Dan Random Forest Dalam Klasifikasi Penyakit Paru-paru Untuk Meningkatkan Akurasi,” *SMARTLOCK: Jurnal Sains dan Teknologi*, vol. 2, no. 2, pp. 32–38, 2023, [Online]. Available: <https://e-jurnal.nobel.ac.id/index.php/smartlock>
- [9] R. Supriyadi, W. Gata, N. Maulidah, and A. Fauzi, “Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah,” *E-Bisnis : Jurnal Ilmiah Ekonomi dan Bisnis*, vol. 13, no. 2, pp. 67–75, 2020, doi: 10.51903/e-bisnis.v13i2.247.
- [10] F. Fakhurrozi *et al.*, “Prediksi Kerusakan Bangunan Pasca Gempa Bumi Menggunakan Metode Deep Neural Network,” *Jurnal Teknologi Sistem Informasi dan Aplikasi*, vol. 7, no. 1, pp. 131–142, 2024, doi: 10.32493/jtsi.v7i1.37181.
- [11] H. Tantyoko, D. K. Sari, and A. R. Wijaya, “Prediksi Potensial Gempa Bumi Indonesia Menggunakan Metode Random Forest Dan Feature Selection,” *IDEALIS : InDonEsiA journal Information System*, vol. 6, no. 2, pp. 83–89, 2023, doi: 10.36080/idealis.v6i2.3036.