

Spatial Analysis of Earthquake Risk on Sumatra Island Using the Neyman–Scott Cox Process (NSCP) and *Random Forest*

Raisqa Faadillah Jatmiko, Ichsan Dwi Risky Syahputra, Fajar Rahmat, dan Tessy Octavia Mukhti*

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: tessyoctaviam@fmipa.unp.ac.id

Submitted : 14 Mei 2026

Revised : 19 Agustus 2026

Accepted : 19 Agustus 2026

ABSTRACT

Indonesia located at the intersection of the Indo-Australian, Pacific, and Eurasian tectonic plates, is highly vulnerable to geological disasters, particularly earthquakes. Sumatra Island, a densely populated region with high agricultural productivity, is strongly affected by seismic activity. This study aims to analyze earthquake risk on Sumatra Island by integrating the Neyman–Scott Cox Process (NSCP) spatial model and the Random Forest algorithm. Secondary data on earthquakes in Sumatra from 2020 to 2025 with magnitudes of ≥ 5 were obtained from the United States Geological Survey (USGS). Geological covariates included active fault zones, subduction zones, and volcanic locations obtained from the Indonesian Earthquake Source and Hazard Map and the Center for Volcanology and Geological Hazard Mitigation. The NSCP model was used to identify spatial clustering patterns of earthquake events, while Random Forest was applied to classify earthquake risk into high- and moderate-risk categories based on geospatial variables. Among the four NSCP cluster models, the Thomas Cluster model provided the best spatial representation based on the lowest Akaike Information Criterion (AIC) value. Distance to the subduction zone was the most influential variable in earthquake risk classification. Most western areas of Sumatra were identified as high-risk zones. The resulting risk-zoning map can support disaster mitigation policies, regional planning, and agricultural infrastructure protection. This study is expected to strengthen early warning systems, maintain food distribution and production, and contribute to food security and sustainable development in disaster-prone areas. Furthermore, this approach can be applied to other regions in Indonesia with similar geological characteristics and support more effective disaster risk management and spatial planning.

Keywords: Earthquake Risk, Sumatra Island, Neyman-Scott Cox Process, Random Forest, Spatial



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Indonesia merupakan negara yang sangat akrab dengan berbagai bencana alam seperti gempa bumi, tanah longsor, dan tsunami. Posisi geografisnya yang unik, diapit oleh tiga lempeng tektonik utama - Lempeng Indo-Australia, Lempeng Pasifik, dan Lempeng Eurasia - membuatnya menjadi salah satu wilayah yang paling rentan bencana alam di dunia (Susanto et al., 2023). Letaknya yang strategis di tengah-tengah Cincin Api Pasifik dan sabuk gempa Alpide Belt membuat sekitar 81% wilayah Indonesia (Yonantan, 2024) selalu berada dalam ancaman terkena bencana. Setiap tahun, Indonesia seringkali menghadapi serangkaian fenomena alam, mulai dari angin kencang dan banjir, hingga tanah longsor, gempa bumi, dan tsunami yang merusak. Berdasarkan data dari *Centre for Research on the Epidemiology of Disasters* (CRED) tahun 2023, Indonesia tercatat sebagai salah satu negara dengan frekuensi bencana alam tertinggi di dunia, terutama banjir dan gempa bumi.

Salah satu wilayah dengan aktivitas gempa bumi tertinggi di Indonesia adalah Pulau Sumatera. Pulau Sumatera, yang terletak di antara lempeng tektonik Indo-Australia, Pasifik, dan Eurasia, serta di Cincin Api Pasifik, Patahan Semangko sangat rentan terhadap berbagai bencana, terutama gempa bumi (Susanto et al., 2023). Posisi geografis ini menjadikannya salah satu wilayah paling rawan bencana di dunia (Yonantan, 2024). Aktivitas lempeng tektonik, keberadaan Sesar Sumatera, dan gunung berapi aktif di Pegunungan Bukit Barisan (Susanto et al., 2023) menjadikan pulau ini titik gempa paling aktif. Bukti nyata kerentanan ini adalah Gempa Bumi dan Tsunami Aceh pada tahun 2004.

Gempa bumi berskala besar merusak ketahanan pangan, khususnya di daerah terdampak langsung (Launuru, 2022). Sebagai contoh, gempa di Pasaman dan Pasaman Barat pada Februari 2022 mengakibatkan kerusakan serius

pada 1.056 hektare sawah, 18,2 hektare jalan usaha tani, dan 17 traktor tangan (Maulana, 2022). Dampak ini tidak hanya mengganggu produksi dan distribusi pangan, tetapi juga mengikis mata pencarian, meningkatkan risiko kemiskinan, dan menghambat pencapaian Tujuan Pembangunan Berkelanjutan (SDGs), khususnya Tujuan 2 (Tanpa Kelaparan) dan Tujuan 11 (Kota dan Komunitas Berkelanjutan). Oleh karena itu, diperlukan analisis risiko gempa bumi untuk mengidentifikasi wilayah yang memiliki tingkat risiko tinggi sebagai dasar dalam mendukung upaya mitigasi bencana.

Pada penelitian ini akan dilakukan analisis risiko gempa bumi di Pulau Sumatera menggunakan *Neyman-Scott Cox Process* dan *Random Forest*. *Neyman-Scott Cox Process* (NSCP) merupakan salah satu model *Cox Process* dalam analisis pola titik spasial yang digunakan untuk memodelkan pola kejadian yang membentuk kelompok (*cluster*). Penerapan metode ini pada kejadian gempa telah dilakukan oleh Trisnisa et al. (2019) untuk menganalisis kejadian gempa bumi di Pulau Jawa dengan mempertimbangkan faktor geologi berupa gunung berapi dan zona subduksi. Hasil penelitian menunjukkan bahwa kejadian gempa bumi di Pulau Jawa memiliki pola pengelompokan (*cluster*). *Neyman-Scott Cox Process* dikonstruksikan berdasarkan proses titik induk (*parent process*), yang dalam konteks gempa bumi dapat merepresentasikan kejadian gempa utama (*mainshock*) dan diasumsikan mengikuti proses Poisson stasioner (Choiruddin et al., 2021) yang memiliki empat model *cluster* yaitu *Matern Cluster Process*, *Thomas Cluster Process*, *Cauchy Cluster Process*, dan *Variance-Gamma Cluster process* yang masing-masingnya memiliki nilai *fpair correlation function* dan *K-function* yang berbeda-beda (Erlinda et al. 2025). Parameter interaksi setiap model diestimasi menggunakan metode *second-order composite likelihood*. Selanjutnya, kinerja keempat model dibandingkan menggunakan *Akaike Information Criterion* (AIC), dan model dengan nilai AIC terendah dipilih sebagai model terbaik untuk menggambarkan pola pengelompokan serta memetakan risiko gempa bumi secara spasial.

Berdasarkan penelitian terdahulu tersebut, pendekatan NSCP mampu menggambarkan pola pengelompokan kejadian gempa serta keterkaitannya dengan faktor geologi. Namun, analisis tersebut belum secara langsung mengklasifikasikan wilayah berdasarkan tingkat risiko gempa. Oleh karena itu, penelitian ini mengintegrasikan algoritma *machine learning Random Forest* pada tahap analisis selanjutnya. *Random Forest* merupakan salah satu algoritma dalam kelompok *ensemble learning* yang digunakan untuk klasifikasi atau regresi. *Random Forest* ini bekerja dengan membentuk sejumlah model dari data pelatihan yang di ambil secara acak melalui metode *bagging*, kemudian menggabungkan hasil prediksi dari seluruh model tersebut untuk memperoleh hasil akhir yang stabil dan akurat (Irawan et al., 2021).

Penerapan *Random Forest* pada analisis gempa bumi telah dilakukan oleh Puspita et al. (2025) untuk memprediksi probabilitas kejadian gempa bumi di Indonesia berdasarkan data historis dan koordinat geografis. Hasil penelitian menunjukkan bahwa *Random Forest* mampu menghasilkan tingkat akurasi sebesar 90,78%. Selain itu, Pratiwi and Arum (2025) menerapkan *Random Forest* untuk mengklasifikasikan kejadian gempa bumi di Indonesia yang berpotensi menimbulkan tsunami dan memperoleh tingkat akurasi sebesar 97%.

Berdasarkan hasil penelitian sebelumnya, penelitian ini bertujuan untuk menganalisis risiko gempa bumi di Pulau Sumatera melalui penerapan *Neyman-Scott Cox Process* dan algoritma *Random Forest*. Pengembangan penelitian difokuskan pada pemodelan pola spasial kejadian gempa menggunakan beberapa model *cluster* pada NSCP, dilanjutkan dengan pemilihan model terbaik berdasarkan nilai *Akaike Information Criterion* (AIC). Selanjutnya, algoritma *Random Forest* diterapkan untuk mengklasifikasikan tingkat risiko gempa bumi berdasarkan faktor-faktor geospasial dan geologi serta mengevaluasi kinerja model melalui *confusion matrix*. Selain itu, penelitian ini mengidentifikasi faktor geologi yang paling berpengaruh terhadap klasifikasi risiko gempa berdasarkan nilai *Mean Decrease Accuracy*, sehingga dapat memberikan gambaran mengenai pola dan tingkat risiko gempa bumi di Pulau Sumatera.

II. METODE PENELITIAN

A. Sumber data dan Variabel Penelitian

Penelitian ini merupakan penelitian terapan dengan menggunakan data sekunder yang mencakup kejadian gempa bumi di wilayah Sumatera pada periode 2020–2025 dengan magnitudo ≥ 5 Skala Richter (SR), diperoleh dari database online USGS (<https://earthquake.usgs.gov>) yang memuat koordinat latitude dan longitude dalam rentang 95° BT – 105° BT dan 6° LU – 6° LS. Selain itu, data kovariat geologis yang digunakan meliputi zona sesar aktif dan zona subduksi yang diambil dari Peta Sumber dan Bahaya Gempa Bumi Indonesia terbitan Kementerian PUPR tahun 2017, serta data lokasi gunung api yang diperoleh dari situs resmi Pusat Vulkanologi dan Mitigasi Bencana Geologi (<https://vsi.esdm.go.id>). Variabel yang digunakan dalam penelitian ini meliputi titik koordinat episentrum gempa bumi, jarak episentrum gempa bumi terhadap sesar aktif terdekat, jarak episentrum gempa bumi terhadap zona subduksi terdekat, dan jarak episentrum gempa bumi terhadap gunung api terdekat (Mukhti et al., 2021).

B. Teknik Analisis Data

Pada penelitian ini proses pengolahan data dilakukan dengan menggunakan *Neyman-Scott Cox Process* dan *Random Forest*. dengan bantuan *software* RStudio. Adapun tahap analisis yang dilakukan adalah sebagai berikut:

1. Menyiapkan data yang akan diteliti yaitu data kejadian gempa bumi, data sesar aktif, zona subduksi, dan gunung api di wilayah Pulau Sumatera.
2. Melakukan *preprocessing data* berupa seleksi data gempa bumi dengan magnitudo ≥ 5 SR, pembersihan data, dan transformasi data spasial.
3. Menghitung jarak terdekat antara titik episentrum gempa bumi dengan sesar aktif, zona subduksi, dan gunung api menggunakan packages di Rstudio
4. Melakukan pengujian homogenitas atau stasioneritas persebaran gempa bumi menggunakan uji Chi-Squared dengan rumus:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

Statistik uji χ^2 digunakan untuk membandingkan frekuensi observasi (O_i) dengan frekuensi harapan (E_i) pada setiap kelompok pengamatan sebanyak k kelompok.

5. Melakukan pengujian independensi titik persebaran gempa bumi menggunakan pendekatan grafik Ripley's *K-Function* untuk mengetahui pola pengelompokan titik gempa bumi. Fungsi intensitas proses titik dinyatakan sebagai:

$$\mu(B) = E[N(B)] = \int_B \lambda(u; \theta, z(u)) du \quad (2)$$

Pada rumus tersebut, $\mu(B)$ merupakan nilai harapan dari banyaknya titik pada wilayah B , sedangkan $\lambda(u; \theta, z(u))$ merupakan fungsi intensitas yang dipengaruhi parameter θ dan kovariat spasial $z(u)$.

6. Memodelkan persebaran titik gempa bumi menggunakan metode cluster point process *Cauchy*, *VarGamma*, *Thomas*, dan *MatClust* menggunakan package spatstat pada *software* R. Fungsi intensitas *Neyman-Scott Cox Process* dinyatakan sebagai:

$$\lambda_c(u; \Psi) = \exp(\gamma + \theta^T z(u)) k(u - c; \omega) \quad (3)$$

Fungsi intensitas tersebut menunjukkan bahwa distribusi titik gempa bumi dipengaruhi oleh kovariat spasial $z(u)$, parameter intensitas γ , serta fungsi kernel $k(u - c; \omega)$ yang menggambarkan pola pengelompokan titik di sekitar pusat kluster c .

Model umum *Neyman-Scott Cox Process* dinyatakan sebagai:

$$\lambda(u; \Psi) = \exp(\gamma + \theta^T z(u)) \sum_{c \in C} k(u - c; \omega) \quad (4)$$

7. Menentukan model terbaik berdasarkan nilai *Akaike Information Criterion* (AIC) terkecil dengan rumus:

$$AIC = -2\ln(L) + 2p \quad (5)$$

Dimana L merupakan nilai *likelihood* model dan p merupakan jumlah parameter model. Model terbaik dipilih berdasarkan nilai AIC terkecil.

Dimana $d_{i,C}$ adalah jarak antara *non-medoid* ke- i dengan *medoid* ke- C , x_{ik} adalah nilai objek *non-medoid* ke- i pada variabel ke- k , C_{jk} adalah nilai objek *medoid* ke- j pada variabel ke- k , p adalah banyaknya variabel yang diamati, n adalah banyaknya objek yang diamati, m adalah banyaknya *medoid* yang terbentuk.

8. Melakukan klasifikasi wilayah risiko gempa bumi menggunakan algoritma *Random Forest* berdasarkan variabel geospasial yang digunakan. Prediksi *Random Forest* diperoleh berdasarkan voting mayoritas dengan rumus:

$$\hat{f}(x) = \text{majority vote}\{T_1(x), T_2(x), \dots, T_B(x)\} \quad (6)$$

Pada rumus tersebut, $T_b(x)$ merupakan hasil prediksi dari pohon keputusan ke- b , sedangkan B merupakan jumlah keseluruhan pohon keputusan yang dibentuk.

9. Mengevaluasi model menggunakan *Out-of-Bag* (OOB error) dan tingkat kepentingan variabel menggunakan indeks Gini dengan rumus:

$$Gini = 1 - \sum_{i=1}^c p_i^2 \quad (7)$$

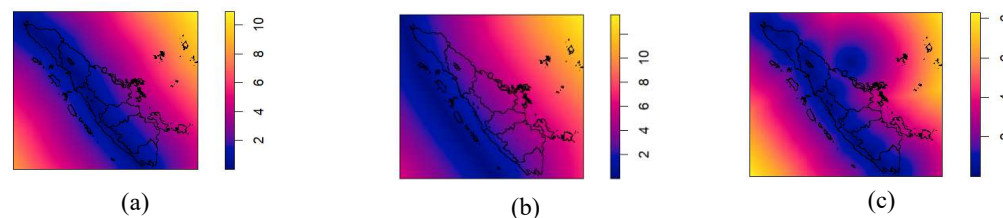
Nilai p_i menunjukkan proporsi kelas ke- i pada suatu node dan c menyatakan jumlah kelas yang digunakan dalam klasifikasi.

10. Membuat peta klasifikasi risiko gempa bumi di Pulau Sumatera berdasarkan hasil model *Neyman-Scott Cox Process* dan *Random Forest*.

III. HASIL DAN PEMBAHASAN

A. Karakteristik Sebaran Gempa Bumi dengan Metode Spasial.

Karakteristik Sebaran Gempa Bumi di pulau Sumatera dapat dilihat melalui analisis memahami sebaran spasial gempa bumi berdasarkan sumber tektoniknya menggunakan metode spasial yang dapat dilihat melalui visualisasi kerapatan gempa pada Gambar 1.

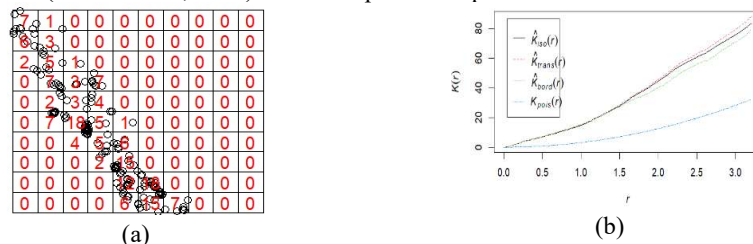


Gambar 1. (a) Sesar Aktif, (b) Subduksi (c) Gunung Api

Hasil yang diberikan oleh Gambar 1 adalah rentang jarak episentrum gempabumi ke sesar aktif terdekat, subduksi terdekat, dan gunung api terdekat. Jarak tersebut dapat dilihat berdasarkan warna pada Gambar yang menunjukkan jika warna semakin gelap maka jarak titik gempa bumi ke masing-masing kovariat semakin dekat, dimana satuan pada Gambar 1 adalah unit 100km²

B. Hasil Uji Homogenitas dan Uji Independensi

Analisis spasial gempa bumi di Pulau Sumatera dilakukan melalui uji homogenitas dan independensi. Hasil uji Chi-square menunjukkan p-value < 0,05 sehingga distribusi gempa bumi tidak homogen dan memiliki spasial trend pada jalur tertentu yang diduga dipengaruhi oleh sesar aktif dan zona subduksi (Izati and Choiruddin, 2022). Selanjutnya, uji independensi menggunakan inhomogeneous K-function menunjukkan seluruh kurva berada di atas kurva Poisson, yang menandakan pola persebaran gempa bumi cenderung mengelompok (clustering). Hasil kedua uji tersebut menunjukkan bahwa persebaran gempa bumi di Pulau Sumatera tidak acak dan dipengaruhi oleh karakteristik geologi aktif (Mukhti et al., 2021). Hal ini dapat dilihat pada Gambar 2.



Gambar 2. (a) Grid Spasial dan (b) Independensi

C. Estimasi Parameter dan Pemilihan Model Terbaik

Estimasi parameter dilakukan untuk melihat pengaruh masing-masing variabel terhadap model yang dibentuk. Pada penelitian ini dilakukan pemodelan spasial terhadap sebaran kejadian gempa bumi di wilayah Sumatera menggunakan beberapa pendekatan model *point process* yaitu *Cauchy*, *Variance Gamma* (VarGamma), *Thomas Cluster*, dan *Matérn Cluster* (MatClust) (Choiruddin et al., 2021). Evaluasi terhadap kinerja masing-masing

model dilakukan dengan menggunakan *Akaike Information Criterion* (AIC) (Portet, 2020). Nilai AIC digunakan untuk menilai keseimbangan antara akurasi dan kompleksitas model, di mana nilai AIC yang lebih rendah menunjukkan model yang lebih baik yang dapat dilihat pada Tabel 1.

Tabel 1. Nilai AIC

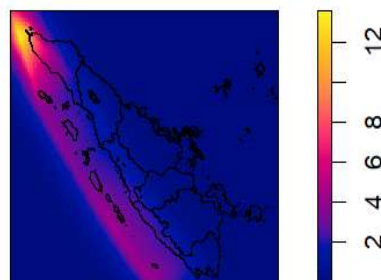
METODE	NILAI AIC
Cauchy	-19022.55
VarGamma	-19035.18
Thomas	-19052.27
MatClust	-19043.22

Berdasarkan Tabel 1, model *Thomas Cluster* memiliki nilai AIC terendah (-19052.27), menjadikannya model terbaik dalam menggambarkan pola spasial kejadian gempa di Sumatera. Sehingga nilai estimasi parameter untuk masing-masing kovariat diperoleh pada Tabel 2,

Tabel 2. Nilai Estimasi Parameter

Indikator	Estimate	Ztest
Intercept	3.0007044	5.4821489
Sesar	-0.5683377	-0.6592922
Subduksi	-1.0170439	-4.9089462
Gunung api	-0.1304896	-0.1422196

Berdasarkan Tabel 2 hasil estimasi parameter menunjukkan bahwa hasil nilai *Z-test* Subduksi sebesar -4.9089462 merupakan nilai absolut paling besar dibandingkan variabel sesar dan gunung api. Koefisien Subduksi sebesar -1.0170439 menunjukkan adanya hubungan negatif, yaitu semakin jauh jarak suatu lokasi dari zona subduksi, maka intensitas gempa yang diprediksi cenderung semakin rendah. Sementara itu, variabel sesar memiliki nilai *Z-test* sebesar -0.6592922 dan gunung api sebesar -0.1304896 yang menunjukkan bahwa kedua variabel tersebut tidak memberikan pengaruh yang signifikan secara statistik terhadap intensitas gempa. Adapun Intercept memiliki nilai *Z-test* sebesar 5,4821489, yang menunjukkan bahwa komponen dasar model berpengaruh signifikan. Dengan demikian, Subduksi merupakan faktor geologi yang paling dominan dan signifikan dalam model, sedangkan Sesar dan Gunung api tidak menunjukkan pengaruh yang signifikan secara statistik. Model yang terbentuk selanjutnya digunakan untuk memprediksi intensitas gempa dan divisualisasikan dalam peta, yang dapat dilihat pada Gambar 3.



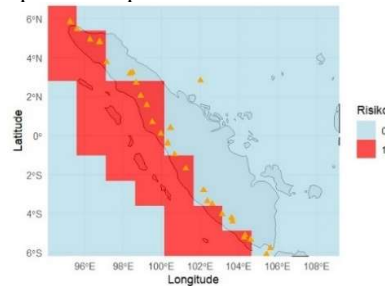
Gambar 3. Peta Prediksi

Berdasarkan Gambar 3, terlihat bahwa hasil pemodelan spasial kejadian gempa bumi di Pulau Sumatera menunjukkan pola yang tidak merata. Nilai yang tinggi ditunjukkan oleh warna kuning hingga merah muda, sedangkan nilai yang rendah ditunjukkan oleh warna ungu hingga biru tua. Wilayah bagian barat laut Pulau Sumatera, terutama di sekitar Aceh dan Sumatera bagian utara, memiliki nilai intensitas atau risiko yang relatif lebih tinggi. Sementara itu, sebagian besar wilayah bagian tengah hingga selatan Pulau Sumatera didominasi oleh warna biru tua, yang menunjukkan nilai hasil pemodelan relatif lebih rendah. Dengan demikian, hasil pemodelan menunjukkan bahwa pola kejadian atau tingkat risiko gempa bumi di Pulau Sumatera tidak tersebar secara homogen, melainkan memiliki konsentrasi yang lebih tinggi pada wilayah tertentu.

D. Pemodelan Menggunakan *Random Forest*

Pemodelan risiko gempa bumi dilakukan dengan pendekatan supervised learning menggunakan algoritma *Random Forest*, untuk mengklasifikasikan wilayah di Pulau Sumatera ke dalam kategori risiko gempa berdasarkan

variabel- variabel geospasial yang relevan membagi zona risiko dengan kategori 0 (Sedang) dan 1 (Tinggi). *Random Forest* digunakan dalam penelitian ini karena kemampuannya dalam menangani data yang kompleks, tidak linear, serta menghasilkan prediksi yang akurat dapat dilihat pada Gambar 4.



Gambar 4. Pemodelan dengan *Random Forest*

Hasil pemodelan *Random Forest* menunjukkan wilayah barat Pulau Sumatera dari Aceh hingga Lampung didominasi zona risiko gempa tinggi. Dalam penelitian ini, zona risiko tinggi merujuk pada wilayah yang berdasarkan hasil klasifikasi model memiliki tingkat intensitas gempa yang lebih tinggi dibandingkan zona risiko sedang. Pola zona risiko tinggi terutama terlihat pada wilayah yang berdekatan dengan zona subduksi dan sesar aktif. Sementara itu, wilayah dengan zona risiko sedang cenderung berada di sekitar area tersebut. Pola spasial yang dihasilkan sejalan dengan karakteristik sebaran sumber gempa yang diperoleh pada analisis sebelumnya. Selanjutnya, evaluasi menggunakan *confusion matrix* menunjukkan kemampuan model dalam mengklasifikasikan kedua kelas risiko tersebut. Hasil evaluasi *confusion matrix* disajikan pada Tabel 3.

Tabel 3. *Confusion Matrix*

Aktual	Sedang	Tinggi	Class.error
Sedang	0	3	1.00000000
Tinggi	1	23	0.04166667

Berdasarkan Tabel 3, model *Random Forest* mampu mengklasifikasikan kelas risiko “Tinggi” dengan baik, yaitu sebanyak 23 dari 24 grid spasial berhasil diklasifikasikan dengan benar, dengan class error sebesar 4,17%. Dalam penelitian ini, kelas risiko “Tinggi” menunjukkan grid spasial yang memiliki tingkat risiko gempa lebih besar berdasarkan hasil klasifikasi model, sedangkan kelas risiko “Sedang” menunjukkan tingkat risiko yang lebih rendah dibandingkan kelas “Tinggi”. Sebaliknya, seluruh 3 grid pada kelas risiko “Sedang” diklasifikasikan sebagai “Tinggi”, sehingga menghasilkan class error sebesar 100%. Hal ini menunjukkan bahwa model belum mampu mengenali kelas risiko “Sedang” dengan baik dan cenderung mengklasifikasikan grid ke dalam kelas “Tinggi”. Dengan demikian, model memiliki kemampuan yang baik dalam mengidentifikasi wilayah dengan risiko tinggi, tetapi masih kurang optimal dalam membedakan wilayah dengan risiko sedang. Selanjutnya, faktor yang paling dominan dalam menentukan risiko gempa bumi di Pulau Sumatera dapat diketahui berdasarkan nilai *Mean Decrease Accuracy* yang disajikan pada Tabel 4.

Tabel 4. *Mean Decrease Accuracy*

KATEGORI	MEAN DECREASE ACCURACY
Jarak Sesar aktif	1.8665772
Subduksi	3.5337239
Gunung Api	0.7173973

Berdasarkan Tabel 4, variabel jarak terhadap zona subduksi memiliki nilai *Mean Decrease Accuracy* tertinggi yaitu sebesar 3.5337239, diikuti dengan variabel jarak terhadap sesar aktif berada pada urutan kedua dengan nilai sebesar 1.8665772, sedangkan variabel jarak terhadap gunung api memiliki nilai terendah, yaitu sebesar 0.7173973. Nilai *Mean Decrease Accuracy* yang lebih tinggi menunjukkan bahwa variabel tersebut memiliki kontribusi yang lebih besar terhadap kinerja model *Random Forest*. Dengan demikian, jarak terhadap zona subduksi merupakan variabel yang paling penting dalam klasifikasi risiko gempa bumi, diikuti oleh jarak terhadap sesar aktif dan jarak terhadap gunung api.

IV. KESIMPULAN

Penelitian ini berhasil memetakan zona risiko gempa bumi di Pulau Sumatera secara akurat. Analisis *Neyman-Scott Cox Process* menunjukkan pola gempa yang mengelompok di zona subduksi dan sesar aktif, dengan

model *Thomas* sebagai yang terbaik ($AIC = -19052,27$). Jarak ke zona subduksi berpengaruh signifikan terhadap intensitas gempa. Model *Random Forest* mengklasifikasikan risiko menjadi sedang dan tinggi dengan akurasi baik, meskipun terdapat 1 kesalahan klasifikasi pada kategori sedang. Variabel jarak ke subduksi menjadi yang paling dominan (Mean Decrease Accuracy = 3,53), diikuti jarak ke sesar aktif. Wilayah barat Sumatera teridentifikasi sebagai zona paling rentan. Hasil ini mendukung pemerintah dalam merumuskan kebijakan mitigasi bencana di daerah dengan prediksi gempa bumi yang tinggi di Pulau Sumatera.

DAFTAR PUSTKA

- Choiruddin, A., Aisah, T., Trisnisa, F., & Iriawan, N. (2021). Quantifying effect of geological factors on distribution of earthquake occurrences by inhomogeneous Cox processes. *Pure and Applied Geophysics*, 178, 1579–1592. <https://doi.org/10.1007/s00024-021-02713-2>
- Damiana. (n.d.). BMKG mulai siap siaga megathrust-tsunami Selat Sunda–Mentawai, kenapa? CNBC Indonesia.
- Erlinda, E., Choiruddin, A., & Widhianingsih, T. D. A. (2025). Convolutional neural networks for inhomogeneous Neyman–Scott Cox process with application in earthquake risk prediction. In 2025 4th International Conference on Electronics Representation and Algorithm (ICERA) (pp. 168–173). IEEE. <https://doi.org/10.1109/ICERA66156.2025.11087270>
- Febtiawan, E. P., Akbar, L. A. S. I., & Rachman, A. S. (2024). Forecasting produksi energi photovoltaic menggunakan algoritma *Random Forest* classification. *Journal of Science and Technology*, 5, 1053–1062. <https://doi.org/10.47065/josh.v5i4.5514>
- Irawan, D., Perkasa, E. B., Yurindra, Y., Wahyuningsih, D., & Helmud, E. (2021). Perbandingan klasifikasi SMS berbasis Support Vector Machine, Naive Bayes Classifier, *Random Forest*, dan Bagging Classifier. *Jurnal Sisfokom*, 10(3), 432–437. <https://doi.org/10.32736/sisfokom.v10i3.1302>
- Pemodelan risiko penyebaran COVID-19 di Surabaya Raya menggunakan model Cauchy Cluster Process. *Inferensi*, 5(1), 1–8. <https://doi.org/10.12962/j27213862.v5i1.12345>
- Launuru, M. I. (2022). Dampak gempa bumi bagi kehidupan masyarakat Desa Liang Kecamatan Salahutu Kabupaten Maluku Tengah. *Jurnal Pendidikan Tambusai*, 6(3), 14165–14174. <https://doi.org/10.31004/jptam.v6i3.6124>
- Maulana, A. (2022, March 10). 1.056 hektare sawah irigasi di Pasaman Barat terancam kekeringan. ANTARA News Sumatera Selatan. Artikel ANTARA News Sumatera Selatan
- Mukhti, T. O., Choiruddin, A., & Purihadi. (2021). Generalized additive Poisson models for quantifying geological factors effect on earthquake risk mapping. *Journal of Physics: Conference Series*, 1863, Article 012030. <https://doi.org/10.1088/1742-6596/1863/1/012030>
- Portet, S. (2020). A primer on model selection using the Akaike Information Criterion. *Infectious Disease Modelling*, 5, 111–128. <https://doi.org/10.1016/j.idm.2019.12.010>
- Pratiwi, A. P., & Arum, P. R. (2025). Implementation of random forest algorithm to classify earthquake in Indonesia. *EIGEN Mathematics Journal*, 8(1), 25–33. <https://doi.org/10.29303/emj.v8i1.185>
- Puspita, D. D., Steffi, S., Hoendarto, G., & Tjen, J. (2025). Random forest analysis for predicting the probability of earthquake in Indonesia. *Social Science and Humanities Journal*, 9(1), 6295–6304. <https://doi.org/10.18535/sshj.v9i01.1574>
- Susanto, T. Y., Choiruddin, A., & Purnomo, J. D. T. (2023). On the earthquake distribution modeling in Sumatra by Cauchy Cluster Process: Comparing log-linear and log-additive intensity models. *Jurnal Sains Matematik*, 52(2), 655–667. <https://doi.org/10.17576/jsm-2023-5202-25>
- Trisnisa, F., Metrikasari, R., Rabbanie, R., Sakdiyah, K., & Choiruddin, A. (2019). Model *inhomogeneous spatial Cox processes* untuk pemetaan risiko gempabumi di Pulau Jawa. *Inferensi*, 2(2), 107–116. <https://doi.org/10.12962/j27213862.v2i2.6825>
- Yonatan, A. Z. (2024, May 4). 10 negara dengan bencana alam terbanyak 2023, ada Indonesia. GoodStats. Artikel GoodStats