

Klasifikasi Risiko Gempa Bumi menggunakan metode Decision Tree

Niko Akbar^{1)*}, Hamzah Alghifari²⁾, Nurul Abdillah³⁾, Oki Dahwanu⁴⁾

^{1),2),3),4)}Prodi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Jambi, Indonesia

*Coresponding Email: nikoakbar@unja.ac.id

Abstrak

Gempa bumi merupakan salah satu bencana alam yang sulit diprediksi namun memiliki dampak besar terhadap kehidupan manusia. Oleh karena itu, diperlukan suatu pendekatan analitis yang mampu mengidentifikasi pola dan hubungan antarparameter gempa untuk mendukung sistem peringatan dini. Penelitian ini bertujuan untuk melakukan Klasifikasi Dampak akan kejadian gempa bumi menggunakan metode Data Mining berbasis Decision Tree. Data yang digunakan berasal dari katalog gempa yang memuat atribut seperti Tanggal Kejadian (*timestamp*), magnitudo, kedalaman, serta koordinat lokasi (*latitude* dan *longitude*). Proses analisis meliputi tahap pembersihan data (*data cleaning*), transformasi, dan pembuatan model klasifikasi Decision Tree untuk menentukan tingkat potensi dan dampak gempa serta mengetahui keakuratan gempa bumi berdasarkan data dari tahun ketahun. Hasil penelitian menunjukkan bahwa atribut magnitudo memiliki pengaruh signifikan terhadap tingkat risiko gempa. Model Decision Tree yang dibangun mampu menghasilkan aturan klasifikasi seperti "Jika magnitudo <5 maka berpotensi Risiko gempa bumi rendah, sedangkan magnitudo antara 5-7 berisiko gempa bumi sedang, dan magnitudo ≥ 7 maka berpotensi Risiko Gempa Bumi Tinggi", yang dapat digunakan untuk mendukung pengambilan keputusan dalam mitigasi bencana. Dengan demikian, metode Decision Tree terbukti efektif dalam mengungkap pola tersembunyi dari data gempa bumi dan dapat menjadi dasar bagi sistem prediksi serta peringatan dini gempa di masa mendatang. Disimpulkan bahwa akurasi masing - masing sebesar 100 %, sedangkan recall sebesar 100 % tapi hasil precision menunjukkan statistik yang berbeda yakni Prediksi Tidak Berpotensi Gempa Bumi Besar sebesar 100%, sedangkan Prediksi Tidak berpotensi Gempa Bumi sedang sebesar 99,67%, dan terakhir prediksi Tidak berpotensi Gempa Bumi Kecil sebesar 96 %. Hasil pengujian ternyata menghasilkan magnitude rendah dengan ukuran $>5,350$ tergolong rendah, sedangkan magnitude rendah dengan ukuran $<5,350$ mempunyai frekuensi yang banyak. Dan Beberapa data berdasarkan statistik Magnitudo >5.350 ukuran sedang dari data sebanyak 16 data. Sedangkan magnitudo ≤ 5.350 Ukuran Rendah sebanyak 6716 Data yang ditemukan dan sudah dianalisis.

Kata Kunci : Klasifikasi, Data Mining, Decision Tree, Gempa Bumi, Risiko Gempa Bumi.

Abstract

*Earthquakes are one of the natural disasters that are difficult to predict but have a major impact on human life. Therefore, an analytical approach is needed that is capable of identifying patterns and relationships between earthquake parameters to support early warning systems. This study aims to classify the impact of earthquakes using a Decision Tree-based Data Mining method. The data used comes from an earthquake catalog containing attributes such as the date of occurrence (*timestamp*), magnitude, depth, and location coordinates (*latitude* and *longitude*). The analysis process includes data cleaning, transformation, and the creation of a Decision Tree classification model to determine the potential and impact of earthquakes and to determine the accuracy of earthquakes based on data*

from year to year. The results of the study show that the magnitude attribute has a significant influence on the level of earthquake risk. The Decision Tree model that was built was able to produce classification rules such as "If the magnitude is <5, then the earthquake risk potential is low, while a magnitude between 5-7 indicates a moderate earthquake risk, and a magnitude of ≥ 7 indicates a high earthquake risk potential," which can be used to support decision-making in disaster mitigation. Thus, the Decision Tree method has proven to be effective in uncovering hidden patterns in earthquake data and can serve as the basis for earthquake prediction and early warning systems in the future. It was concluded that the accuracy of each was 100%, while the recall was 100%, but the precision results showed different statistics, namely the prediction of no potential for a major earthquake was 100%, while the prediction of no potential for a moderate earthquake was 99.67%, and finally the prediction of no potential for a minor earthquake was 96%.

Keyword : *Clasification, Data Mining, Decision Tree, Earthquake, Earthquake Risk.*

PENDAHULUAN

Indonesia terletak di atas lempengan tektonik dari pasifik, Eurasia, dan Indo Australia Dimana dilanjutkan bergerak aktif, sehingga rentan terhadap gempa bumi akibat pelepasan gelombang seismic pada batuan di kerak bumi[1]. Pemicu lain gempa bumi adalah aktivitas vulkanik gunung berapi aktif yang mengelilingi kepulauan indonesia[2]. Badan meteorologi, Klimatologi, dan Geofisika mencatat gempa bumi mencapai 400 kali setiap bulan hingga pada tahun 2019 terdapat 11.660 gempa bumi di situs repogempa.bmkg.id[1].

Kuantitas Gempa Bumi berdasarkan data setiap tahun dapat digunakan sebagai dasar untuk menganalisis distribusi gempa bumi dengan metode data mining, yang merupakan metode pengolahan data berskala besar untuk memperoleh informasi baru yang mudah dipahami[3], [4].

Beberapa analisis gempa bumi telah dilakukan menggunakan berbagai metode untuk menentukan distribusi gempa bumi, daerah rawan gempa, dan dampak gempa bumi berdasarkan data tersebut. Analisis gempa bumi dapat dilakukan dengan pendekatan klasifikasi wilayah. Berdasarkan kutipan Ismail[3] menyatakan bahwa klasifikasi wilayah gempa bumi dapat dilakukan menggunakan algoritma random forest berdasarkan peristiwa gempa bumi dalam bentuk koordinat (latitude, longitude), kedalaman (depth), dan magnitudo (kekuatan energi seismik gempa bumi), dengan akurasi 99,97%.

Metode lain yang dapat digunakan dalam analisis gempa bumi adalah K-means Clustering[2], [4].

Beberapa Penelitian yang telah dilakukan berupa Sistem Algoritma Naive Bayes menurut damuri[5], memiliki fungsi untuk menemukan pengetahuan atau pola-pola kesamaan karakteristik dalam suatu kelompok atau kelas tertentu. Prediksi tingkat penerimaan bantuan sembako yang digunakan terdapat dua kelas, yaitu layak dan tidak layak. Data yang digunakan untuk prediksi, yaitu data yang diambil dari sampel data warga di desa XYZ. Dari hasil evaluasi menggunakan confusion matrix didapatkan akurasi yang dihasilkan untuk 135 data training dengan 40 data testing dan tujuh atribut yang digunakan menghasilkan akurasi sebesar 86%, recall 85%, dan presisi 88%. Kemudian penelitian lain menyebutkan[6] metode algoritma K-Means Clustering Data Mining didapatkan daerah penjualan produk yang tinggi, sedang, dan rendah pada Tingkat Penjualan Paket Data Telkomsel. Daerah dengan penjualan produk rendah akan dilakukan promosi penjualan produk, serta penjualan yang tinggi tidak perlu diadakan promosi, jadinya Cluster tinggi (C1) nilainya 16, Cluster Sedang (C2) bernilai 7, dan Cluster terendah (C3) bernilai 29.

Penelitian ini bertujuan untuk Klasifikasi dampak risiko gempa berdasarkan data - data dengan metode Decision tree.

Metode Decision Tree atau pohon keputusan merupakan suatu struktur pohon yang digunakan sebagai metode penalaran untuk mencari solusi suatu permasalahan inputan[7]. Bergantung pada jenis nilai dalam dataset, berbentuk pohon. Metode ini merupakan teknik klasifikasi dalam data mining yang mengubah data dalam jumlah besar menjadi pohon keputusan yang mewakili aturan. Digunakan juga untuk menguji data dan menemukan korelasi antara variabel input dan variabel target. Adapun kelebihan menggunakan metode Decision tree adalah mudah dipahami dan di interpretasi cocok untuk pengambilan keputusan yang transparan, Serta dapat mengatasi kumpulan data yang kompleks dan beragam

tanpa normalisasi atau transformasi data yang rumit. Serta membantu mengidentifikasi faktor-faktor yang paling penting dalam pengambilan keputusan.

METODE PENELITIAN

A. Alur Penelitian

1. Identifikasi Masalah

Dalam penelitian ini diperlukan Identifikasi, Gambaran, dan menganalisis permasalahan yang berkaitan dengan Kejadian Gempa Bumi menggunakan metode Data Mining berbasis Decision Tree

2. Menentukan Tujuan

Tujuan yang dicapai dalam Penelitian ini adalah untuk menentukan Klasifikasi (misalnya, gempa berpotensi tsunami atau tidak, gempa ringan/sedang/berat, daerah rawan/tidak rawan) atau Klasifikasi kategori risiko. Dan beberapa kemungkinan lain yang bisa ditemukan dari penelitian ini.

3. Studi Literatur

Mencari berbagai sumber dan berbagai referensi untuk menemukan Kegunaan lain bahwa Decision Tree sebagai metode tidak hanya digunakan untuk alat analisis untuk mendukung pengambilan keputusan tetapi berguna untuk mengidentifikasi kriteria – kriteria yang signifikan serta faktor-faktor yang mempengaruhi akan terjadinya prediksi gempa bumi dan kekuatan gempa bumi berdasarkan data-data lampau.

4. Mengumpulkan Data

Data yang akan digunakan adalah data yang berasal dari kaggle yang di rilis oleh BMKG (<https://www.kaggle.com/datasets/kekavigi/earthquakes-in-indonesia>) dan Katalog Gempa Bumi dari Survei Geologis Amerika Serikat.

5. Pengolahan Data

Karena data yang akan dipakai berupa data yang belum diolah, maka dilakukan Pembersihan Data.

Pembersihan data merupakan langkah penting untuk Persiapan data yang akan diperiksa untuk mengidentifikasi dan menangani nilai yang hilang atau data tidak konsisten. Proses ini penting untuk memastikan keandalan dan akurasi data dengan memperbaiki ketidak sesuaian dan meningkatkan kualitas secara keseluruhan. Proses ini melibatkan verifikasi menyeluruh untuk mendeteksi anomali atau ketidakteraturan.

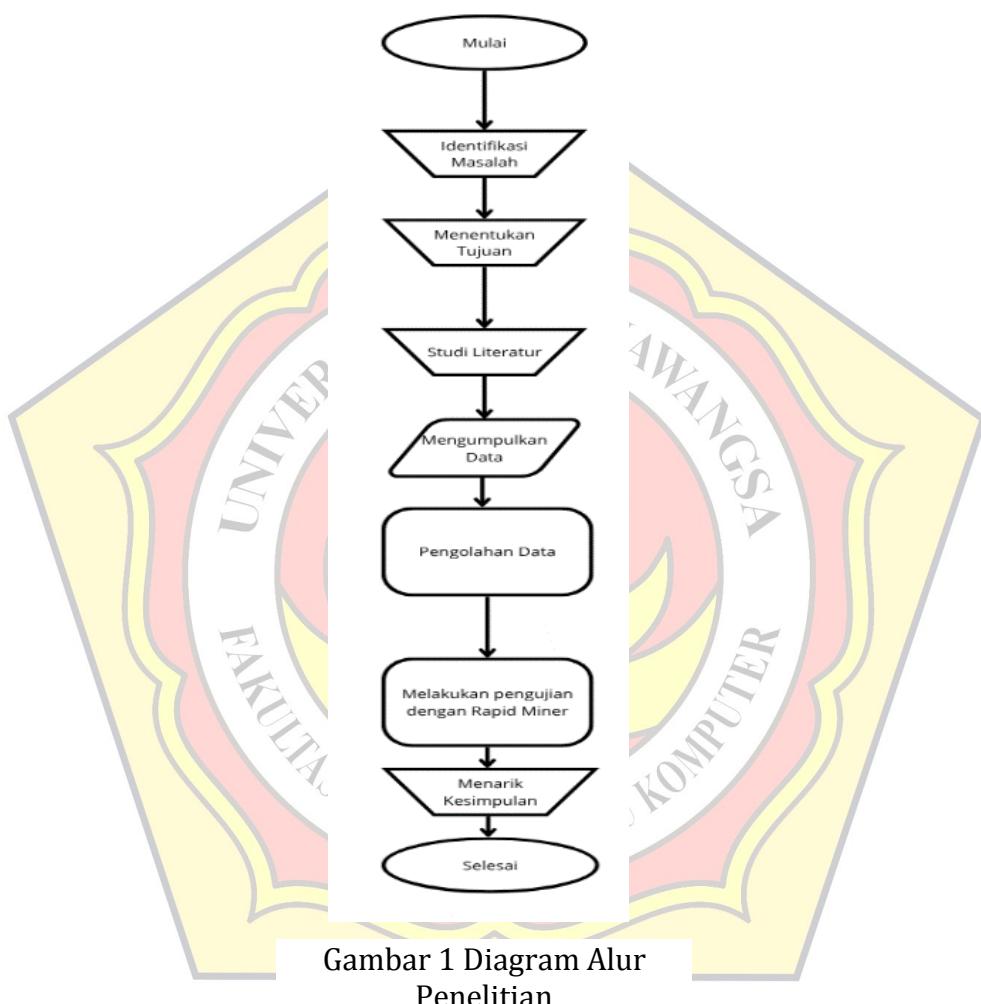
6. Melakukan Pengujian

Menguji data dilakukan dengan Aplikasi Rapid Miner. Rapid Miner adalah Software Penambangan Data. Menggunakan aturan dan algoritma pemrosesan data, memiliki akurasi dan metode tinggi yang mudah digunakan. Aplikasi ini juga menganalisis data secara mandiri atau alat pemrosesan data ini diintegrasikan ke dalam produk. Menggunakan teknologi deskriptif dan prediktif untuk memberikan pengguna pengetahuan yang dibutuhkan untuk mencapai hasil yang optimal[8], [9].

Beberapa fitur Rapid Miner antara lain

- Bentuk Grafis Canggih, seperti Bagan Histogram, Bagan Pohon, dan Plot Sebar 3D.
- Banyak Variasi Plugin.
- Menyediakan Teknik Penambangan Data dan Pembelajaran Mesin seperti ETL (Ekstrak, Transformasi, Muat), Pemrosesan Awal Data, Visualisasi, Pemodelan, dan Evaluasi.
- Proses Penambangan Data, terdiri dari Operator Bersarang yang ditulis dalam XML dan dibuat menggunakan GUI.
- Mengintegrasikan Proyek dengan Statistik R

Pada tahap ini, peneliti melakukan penarikan kesimpulan yang telah dilaksanakan berdasarkan hasil analisis yang telah dilakukan. Tahap ini menjadikan penulis untuk menjawab hasil dari sebuah penelitian. Selain itu, memiliki peran penting dalam proses analisis data, karena melibatkan evaluasi mendalam terhadap temuan – temuan , dan pengambilan keputusan berdasarkan bukti – bukti.



Metode Analisis Data

metode Analisis Data yang digunakan adalah Metode Decision Tree, sedangkan metode pemilihan atribut menggunakan Gain Ratio, Memilih atribut berdasarkan nilai "Gain" tertinggi dari atribut-atribut yang ada, rumusnya sebagai berikut:

Perhitungan Gain

Gain adalah ukuran efektivitas suatu atribut dalam mengklasifikasi data.

$$\text{Gain } (S, A) = \text{Entropy} \sum_{i=0}^n [n_i * \text{Entropy } (S_i)] \quad (1)$$

Keterangan :

S : Himpunan.

A : Atribut.

n : Jumlah Partisi Atribut Aa.

S_i : Jumlah Kasus pada partisi ke-i.

S : Jumlah kasus dalam S[10].

Cabang pohon menggunakan multiway, kemudian terdapat pruning (pemangkasan), dan output yang diharapkan adalah klasifikasi.

Aturan Decision Tree

Untuk mengetahui hasil Kategori Decision Tree dan menghasilkan hasil pertimbangan, maka dibuatlah aturan.

Kategori Tingkat Bahaya Gempa, Data dikategorikan Tingkat bahaya gempa dari Nilai Magnitudo (M), Tergolong Rendah ($M < 5.5$); Tergolong Sedang (5.5 – 7), dan Tergolong Tinggi ($M > 7$).

Confusion Matrix

Kinerja Sistem Klasifikasi menggambarkan seberapa baik sistem klasifikasi data tersebut berfungsi. Confusion Matrix merupakan metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Secara umum, Confusion Matrix berisi informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang diperoleh. Dari Confusion Matrix, kita bisa menentukan Accuracy, Precision, dan Recall, Beberapa Rumus dari Accuracy, Precision, dan Recall sebagai berikut :

Accuracy

$$\frac{(TP+TN)}{TP+TN+FP+FN} \times 100\% \quad (1)$$

Precision

$$\frac{TP}{TP+FP} \times 100\% \quad (2)$$

Recall

$$\frac{TP}{TP+FN} \times 100\% \quad (3)$$

HASIL DAN PEMBAHASAN

Data yang digunakan untuk klasifikasi dan Prediksi Gempa Bumi diperoleh dari situs Kaggle yang data-datanya berasal dari BMKG dari tahun 2008 -2023. Beberapa data yang akan diangkat dan diteliti berupa No, Tanggal, Jam Gempa, Latitude, Longitude, Depth, Magnitude, Remark yang merupakan tempat terjadinya gempa (Penanda tempat), Kemudian menambahkan baris baru, yaitu Potensi tsunami dan Risiko Gempa.

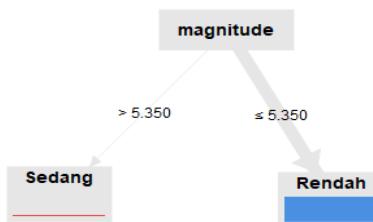
Dari seluruh data sebanyak 92887 buah yang sudah di processing, beberapa buah data akan dijadikan data training dan data testing. Pada data training akan diambil sebanyak 30 data di tahun 2019 – 2023 sehingga masing-masing data pada 4 tahun tersebut jumlahnya sebanyak 150 data training.

Tabel 1. Data Gempa Bumi

No	Tanggal	Waktu	latitude	longitude	Risiko Gempa
1	1/26/2023	58:35.6	0.88	123.72	Rendah
2	1/26/2023	30:23.4	3.06	127.24	Rendah
3	1/26/2023	20:32.5	2.67	127.13	Rendah
4	1/26/2023	11:48.6	3.16	127.17	Rendah
5	1/26/2023	00:31.0	1.92	127.04	Rendah
6	1/26/2023	54:20.1	-3.52	140.65	Rendah
7	1/26/2023	42:29.9	2.7	126.06	Rendah
8	1/26/2023	22:54.8	3.21	127.21	Rendah
9	1/26/2023	22:01.0	-2.24	140.28	Rendah
10	1/26/2023	12:19.4	3	127.14	Rendah
....
92887	11/1/2008	12:31:2	-0.6	98.9	Rendah

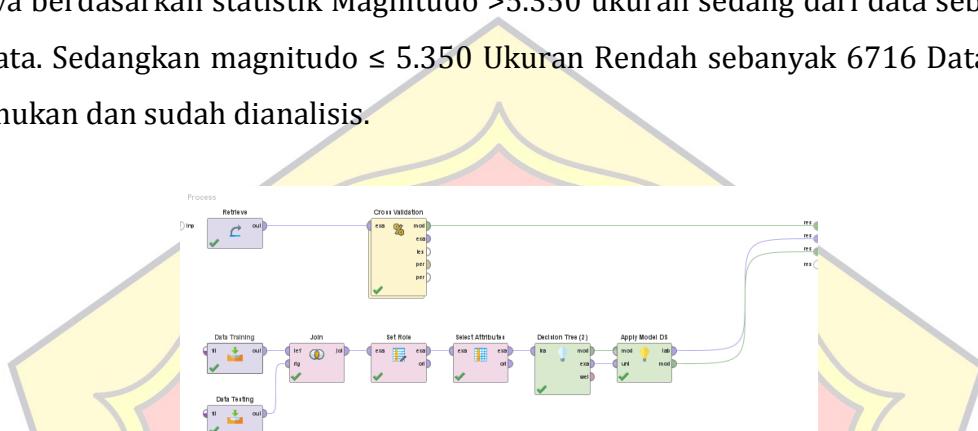
Sedangkan Data Testing diambil sebanyak 10 buah data di tahun yang sama sehingga jumlah data yang akan ditesting sebanyak 50 buah data.

Selanjutnya pengujian dari dataset ini dilakukan dengan menggunakan aplikasi rapid miner. Proses yang dilakukan pertama kali adalah menambahkan dataset ke dalam aplikasi. Kemudian menggunakan operator Join -> Set Role -> Select Attribute -> Decision Tree -> Apply Model. Operator Join digunakan untuk menggabungkan Dataset lebih dari satu, yakni data training dan data testing, kemudian diberikan label untuk "Risiko Gempa", kemudian Select Attribut digunakan untuk mengambil kolom-kolom penting yang akan dipakai dalam mengolah data dan digunakan untuk mendapatkan hasil dari decision tree. Langkah selanjutnya dengan menambahkan operator Decision tree maka beberapa data diolah menggunakan algoritma decision tree. Dan terakhir diberikan operator Apply Model untuk memberikan hasil dalam bentuk visual decision tree.



Gambar 2. Hasil Decision Tree

Berdasarkan Decision Tree dalam bentuk grafik kebanyakan Data Gempa Bumi yang sudah di analisis menghasilkan Dampak Resiko Sedang dengan Ukuran Magnitudo / Ukuran Energi Gempa Bumi Jumlahnya sedikit, sedangkan Dampak Risiko Gempa Bumi Rendah berdasarkan data dari tahun ke tahun jumlahnya sedikit, sedangkan Ukuran Dampak Risiko Gempa bumi Tinggi tidak ada yang berarti Risiko Gempa Bumi di tahun berikutnya tidak akan terjadi. Hal ini dijelaskan bahwa berdasarkan statistik Magnitudo >5.350 ukuran sedang dari data sebanyak 16 data. Sedangkan magnitudo ≤ 5.350 Ukuran Rendah sebanyak 6716 Data yang ditemukan dan sudah dianalisis.

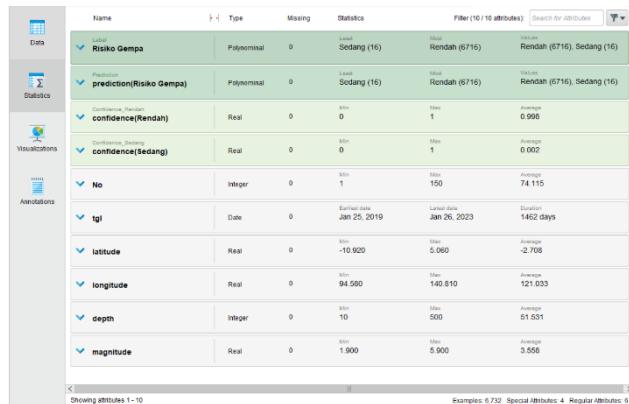


Gambar 3. Proses Data Mining dengan Metode Decision Tree dan Validasi Silang

Berdasarkan Proses dilakukan Validasi Silang berdasarkan data yang sudah disiapkan, kemudian dilakukan persiapan data training dan data testing yang terdiri dari 20 data dari tahun 2019 -2023 dan dilakukan metode decision tree untuk mengetahui prediksi gempa di tahun depan apakah terjadi peluang gempa bumi, dan apa hasil kekuatan gempa tersebut.

Row No.	Risiko Gempa	prediction(Re)	confidence(Re)	confidence(No)	No	tgl	latitude	longitude	dept
1	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
2	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
3	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
4	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
5	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
6	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
7	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
8	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
9	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
10	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
11	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
12	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
13	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
14	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
15	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
16	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
17	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
18	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
19	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125
20	Rendah	Rendah	1	0	1	Jan 26, 2023	0.880	123.720	125

Gambar 4. Data Risiko Gempa Bumi setelah dilakukan Decision Tree



Name	Type	Missing	Statistics	Filter (10 / 10 attributes)	Search for Attributes
Risiko Gempa	Polynomial	0	Least: Sedang (16) Mean: Rendah (6716), Sedang (16)		
Prediction(Risiko Gempa)	Polynomial	0	Least: Sedang (16) Mean: Rendah (6716), Sedang (16)		
confidence(Rendah)	Real	0	Min: 0 Max: 1 Average: 0.998		
confidence(Sedang)	Real	0	Min: 0 Max: 1 Average: 0.002		
No	Integer	0	Min: 1 Max: 150 Average: 74.115		
tgj	Date	0	Earliest date: Jan 25, 2019 Latest date: Jan 26, 2023 Duration: 1462 days		
latitude	Real	0	Min: -10.920 Max: 5.060 Average: -2.708		
longitude	Real	0	Min: 94.580 Max: 140.810 Average: 121.033		
depth	Integer	0	Min: 10 Max: 500 Average: 51.531		
magnitude	Real	0	Min: 1.900 Max: 5.900 Average: 3.556		

Gambar 5. Hasil Validasi Silang Dalam Bentuk Statistik

Setelah dilakukan Validasi Silang maka diketahui Akurasi masing – masing sebesar 100 %, sedangkan recall sebesar 100 % tapi hasil precision menunjukkan statistik yang berbeda yakni Prediksi Tidak Berpotensi Gempa Bumi Besar sebesar 100%, sedangkan Prediksi Tidak berpotensi Gempa Bumi sedang sebesar 99,67%, dan terakhir prediksi Tidak berpotensi Gempa Bumi Kecil sebesar 96 %.

SIMPULAN

Bisa disimpulkan bahwa akurasi masing – masing sebesar 100 %, sedangkan recall sebesar 100 % tapi hasil precision menunjukkan statistik yang berbeda yakni Prediksi Tidak Berpotensi Gempa Bumi Besar sebesar 100%, sedangkan Prediksi Tidak berpotensi Gempa Bumi sedang sebesar 99,67%, dan terakhir prediksi Tidak berpotensi Gempa Bumi Kecil sebesar 96 %.

Untuk Decision tree, ditentukan beberapa data yang diambil dan sudah dilakukan percobaan. berdasarkan Hasil pengujian ternyata menghasilkan magnitude rendah dengan ukuran $>5,350$ tergolong rendah, sedangkan magnitude rendah dengan ukuran $<5,350$ mempunyai frekuensi yang banyak.

Dan Beberapa data berdasarkan statistik Magnitudo >5.350 ukuran sedang dari data sebanyak 16 data. Sedangkan magnitudo ≤ 5.350 Ukuran Rendah sebanyak 6716 Data yang ditemukan dan sudah dianalisis.

DAFTAR PUSTAKA

- [1] J. Homepage, D. Kurmiati, M. Zakiy Fauzi, and A. Falegas, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Clustering of Earthquake Prone Areas in Indonesia Using K-Medoids Algorithm Klasterisasi Daerah Rawan Gempa Bumi di Indonesia Menggunakan Algoritma K-Medoids," vol. 1, pp. 47–57, 2021.
- [2] M. Murdiaty, A. Angela, and C. Sylvia, "Pengelompokan Data Bencana Alam Berdasarkan Wilayah, Waktu, Jumlah Korban dan Kerusakan Fasilitas Dengan Algoritma K-Means," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 3, p. 744, Jul. 2020, doi: 10.30865/mib.v4i3.2213.
- [3] Ismail, "KLASIFIKASI AREA GEMPA BUMI MENGGUNAKAN ALGORITMA RANDOM FOREST," *Jurnal Ilmiah Informatika Komputer*, vol. 26, no. 1, pp. 56–64, 2021, doi: 10.35760/ik.2021.v26i1.3853.
- [4] HARAHP CN, REVANTHIKA F, and YUFIS AZHAR, "Analisis Gempa Bumi Pada Pulau Jawa Menggunakan Clustering Algoritma K-Means," *Jurnal Dinamika Informatika*, vol. 9, no. 1, pp. 51–60, 2020.
- [5] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Riset Komputer)*, vol. 8, no. 6, p. 219, Dec. 2021, doi: 10.30865/jurikom.v8i6.3655.
- [6] S. Handoko, F. Fauziah, and E. T. E. Handayani, "IMPLEMENTASI DATA MINING UNTUK MENENTUKAN TINGKAT PENJUALAN PAKET DATA TELKOMSEL MENGGUNAKAN METODE K-MEANS CLUSTERING," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- [7] Z. Nurizati, A. Hidayat, D. Vernanda, and T. Hendriawan, "Analisis Kelayakan Penurunan UKT Pada Mahasiswa dengan Menggunakan Metode Decision Tree," vol. 18, no. 1.
- [8] H. H. Nisa, Ari Wahyono, and Donna Setiawati, "IMPLEMENTASI ALGORIMA K-MEANS DALAM PENGELOMPOKAN ARMADA BUS LAYAK JALAN DI TERMINAL KLATEN," *JITU : Journal Informatic Technology And Communication*, vol. 8, no. 2, pp. 91–99, Nov. 2024, doi: 10.36596/jitu.v8i2.1603.
- [9] N. K. W. Patrianingsih and I. K. A. Sugianta, "Analisis Kelayakan Kredit Koperasi Mitra Tani Mandiri Dengan Algoritma Naïve Bayes," *ZONAsi: Jurnal Sistem Informasi*, vol. 6, no. 2, pp. 298–307, 2024.
- [10] S. T. M. K. Yahya, *Data Mining*. CV Jejak (Jejak Publisher), 2022. [Online]. Available: <https://books.google.co.id/books?id=0J2mEAAAQBAJ>