



Penerapan Machine Learning dalam Prediksi Klasifikasi Big Data Kedalaman Gempa Bumi di Indonesia Tahun 2015-2024

Ruth Maharani Aquilia Hutagaol¹, Vilanata Tesalonika Lana², Zena Azzahra
Dzunnurain³, Robert Kurniawan⁴

^{1, 2, 3}Program Studi Statistika, Politeknik Statistika STIS, Jakarta, Indonesia

⁴Program Studi Komputasi Statistik, Politeknik Statistika STIS, Jakarta, Indonesia

¹212112343@stis.ac.id

³212112431@stis.ac.id

⁴robertk@stis.ac.id

Corresponding author email: ²212112412@stis.ac.id

Abstract: Predicting the characteristics of an earthquake, especially its depth, is an important starting point for almost all quantitative seismological analysis. This research was conducted to evaluate predictions by classifying earthquakes based on its depth using the variables used. The method used is machine learning with 3 classification prediction models named Naïve Bayes, KNN, and Random Forest. From the performance measurements, the Random Forest model was selected as the best prediction model, which will be used to predict earthquake depth classification. The Accuracy is quite high ($> 85\%$). It also has a small difference between train and test data (fitted model confirmed), and has high prediction quality (accuracy of test data and train data is in the range of $83.77\%-89.66\%$), so it can be concluded that the Random Forest model is the best model for predicting earthquake depth classification in Indonesia.

Keywords: earthquake depth, machine learning, classification, performance measurement, prediction

Abstrak: Mengetahui karakteristik gempa bumi, terutama kedalamannya, merupakan titik awal yang penting untuk sebagian besar analisis seismologi kuantitatif. penelitian ini dilakukan untuk mengevaluasi prediksi dengan mengklasifikasikan gempa bumi berdasarkan kedalamannya menggunakan variabel-variabel yang digunakan. Metode yang digunakan adalah machine learning dengan 3 model prediksi klasifikasi, yaitu Naïve Bayes, KNN, dan Random Forest. Dari performance measurements, terpilih model Random Forest sebagai model terbaik, untuk kemudian digunakan untuk memprediksi klasifikasi kedalaman gempa bumi. Akurasi yang dihasilkan cukup tinggi ($> 85\%$), memiliki selisih yang kecil antara data train dan test (membuktikan model fitted), dan memiliki kualitas prediksi yang tinggi (akurasi data test dan data percobaan berada pada range $83,77\%-89,66\%$), sehingga dapat disimpulkan bahwa model Random Forest adalah model terbaik untuk melakukan prediksi klasifikasi kedalaman gempa bumi di Indonesia.

Kata kunci: kedalaman gempa, machine learning, klasifikasi, performance measurement, prediksi

I. PENDAHULUAN

Gempa bumi adalah pergerakan kerak bumi atau mantel bagian atas secara tiba-tiba yang melepaskan tekanan tektonik dan menimbulkan gelombang seismik [1], pecahnya patahan yang tertanam dan memancarkan gelombang elastis ke permukaan bumi untuk menggetarkan tanah [1], pelepasan energi yang menimbulkan guncangan dan getaran pada tanah. Getaran tanah mencakup spektrum frekuensi yang luas dari puluhan hertz hingga seperseratus detik atau lebih rendah. Getaran tanah dicatat oleh pendulum yang disebut seismograf. Para ilmuwan menggunakan seismometer untuk mencatat dan mengukur kekuatan gempa. Faktanya, gempa bumi terjadi di mana-mana setiap hari, sedemikian rupa sehingga para ilmuwan memperkirakan bahwa ribuan gempa kecil yang hampir tidak terlihat terjadi setiap hari dan tidak menimbulkan kerusakan fisik. Gempa bumi sering terjadi di Indonesia karena terdapat banyak lempeng tektonik dan banyak gunung berapi. Gempa bumi disebabkan oleh tiga faktor utama: pergerakan lempeng, aktivitas manusia, dan letusan gunung berapi [2].

Mengetahui lokasi gempa bumi (lintang, bujur, kedalaman, dan waktu terjadinya) merupakan titik awal yang penting untuk sebagian besar analisis seismologi kuantitatif. Metode untuk menentukan lokasi gempa dengan menggunakan waktu tiba gelombang seismik [1]. Lokasi gempa yang digunakan



adalah kedalaman. Gempa bumi dapat terjadi dimana saja dimulai dengan kedalaman 0 sampai sekitar 700 kilometer di bawah permukaan bumi [3]. Kedalaman gempa tersebut dibagi menjadi tiga zona yaitu dangkal, sedang, dan dalam [4]. Berdasarkan Badan Nasional Penanggulangan Bencana (BNPB), gempa bumi dibagi menjadi tiga macam, yaitu gempa bumi dangkal, gempa bumi menengah, dan gempa bumi dalam. Gempa bumi dangkal merupakan gempa bumi yang terjadi pada kedalaman dengan rentang 0 sampai 60 km di bawah permukaan bumi. Gempa bumi yang terjadi pada kedalaman dengan rentang 60 sampai 300 km di bawah permukaan bumi disebut sebagai gempa bumi menengah, sedangkan gempa bumi yang terjadi pada kedalaman lebih dari 300 km di bawah permukaan bumi disebut sebagai gempa bumi dalam. Semua gempa bumi yang terjadi lebih dari 70 km di bawah permukaan bumi terlokalisasi di dalam lempengan besar litosfer yang tenggelam ke dalam mantel bumi.

Kerusakan yang disebabkan oleh gempa bumi dengan kedalaman yang berbahaya dapat terjadi dengan fatal. Oleh sebab itu, prediksi gempa bumi sangat penting untuk dilakukan. Namun dalam praktiknya, memprediksi gempa bumi bukanlah hal yang mudah karena memprediksi gempa bumi merupakan kegiatan yang sangat kompleks. Hal tersebut berkaitan dengan pola yang ditimbulkan oleh alam dan kontribusi variabel-variabel penjelas untuk memprediksi gempa bumi. Dengan tingkat kompleksitas tersebut, telah banyak dilakukan penelitian terkait prediksi gempa bumi dengan berbagai macam metode serta variabel penjelas. Penelitian Mondol menggunakan empat metode yaitu regresi linear, regresi polinomial, algoritma Random Forest, dan Long Short-Term Memory untuk memprediksi magnitude dan depth dari gempa bumi. Hasil dari penelitian ini menunjukkan bahwa terdapat perbedaan model terbaik antara untuk memprediksi magnitude, memprediksi depth, dan memprediksi secara keseluruhan. Model terbaik untuk memprediksi magnitude gempa bumi adalah regresi polinomial, model terbaik untuk memprediksi kedalaman gempa bumi adalah Random Forest, dan model terbaik secara keseluruhan untuk memprediksi adalah regresi polynomial [5]. Sementara Jain et al. dalam penelitiannya menggunakan algoritma Random Forest, Multi-Layer Perceptron (MLP) Regression, dan Support Vector Regression (SVR) untuk memprediksi magnitude gempa bumi. Metode ini diterapkan pada radius yang berbeda. Hasil menunjukkan bahwa MLP lebih baik daripada algoritma lain [6].

Tantyoko et al. menggunakan algoritma Random Forest serta Feature Selection, Support Vector Machine (SVM), Naïve Bayes, dan Decision Tree untuk memprediksi potensial gempa bumi. Random Forest digunakan dengan beberapa pohon keputusan. Penelitian ini juga melakukan pemilihan fitur untuk mendapatkan fitur-fitur dengan korelasi yang kuat sehingga fitur hasil seleksi adalah tanggal, latitude, longitude, magnitude, dan lokasi. Hasil dari penelitian ini menunjukkan bahwa prediksi Random Forest dengan pemilihan fitur lebih baik dibandingkan tanpa seleksi fitur serta lebih baik dari algoritma yang lain [7]. Macedo et al. dalam membuat prediksi perpindahan lereng bukit yang disebabkan oleh gempa kerak dangkal di seluruh dunia menggunakan beberapa algoritma regresi berbasis geographical engineering dan machine learning. Kedua metode direkomendasikan untuk digunakan, dengan metode machine learning terbaik yang digunakan adalah Random Forest dengan performance measurement berupa RSE yang terkecil [8].

Performa dari prediksi ditentukan oleh pemilihan variabel penjelas dan juga, prediksi kedalaman gempa bumi dilakukan dengan menggunakan beberapa variabel penjelas yang dianggap memiliki hubungan dengan kedalaman gempa bumi. Berdasarkan Mondol dalam penelitiannya, garis lintang dan garis bujur dapat digunakan untuk memprediksi kedalaman gempa bumi. Garis lintang dan garis bujur dapat menentukan zona seismik dimana frekuensi gempa bumi berulang dapat terjadi. Hal tersebut berguna untuk memprediksi kedalaman gempa bumi di zona seismik [5]. Selain garis lintang dan bujur,



berdasarkan Tehseen et al., apabila semakin banyak energi yang dilepaskan maka kedalaman gempa menjadi dangkal yang artinya ketika gempa bumi semakin besar maka kedalamannya semakin mendekati kerak bumi [9]. Lokasi dan jumlah stasiun juga memiliki hubungan dengan kedalaman gempa bumi, yang mana berdasarkan Siddiquee et al., ketepatan estimasi dari kedalaman gempa sangat bergantung dengan lokasi dan jumlah stasiun [10].

Gempa besar memiliki kaitan dengan kedalaman gempa bumi, dimana semakin besar gempa bumi maka kedalamannya akan semakin mendekati kerak bumi. Jika seismic gap teridentifikasi, maka wilayah tersebut berpeluang besar diperkirakan akan mengalami gempa bumi besar dengan skala yang meningkat seiring waktu setelahnya pada patahan atau batas lempeng tertentu [11].

Namun, dari penelitian terkait, masih sangat sedikit penelitian yang melakukan prediksi kedalaman dari gempa bumi. Sebagian besar penelitian sebelumnya melakukan prediksi magnitudo dan juga potensi gempa bumi. Melakukan prediksi kedalaman gempa bumi sangat penting dilakukan agar masyarakat dapat lebih siaga dengan gempa yang akan terjadi pada kedalaman yang dangkal. Oleh sebab itu, penelitian ini akan melakukan prediksi pada kedalaman gempa bumi tergantung garis lintang, garis bujur, magnitudo, jumlah total stasiun seismik yang digunakan untuk menghitung magnitudo, magnitudo type, jumlah total stasiun seismik yang digunakan untuk menentukan lokasi gempa, dan kesenjangan seismik. Tujuan dari penelitian ini adalah untuk mengevaluasi prediksi dengan mengklasifikasikan gempa bumi berdasarkan kedalamannya menggunakan variabel-variabel penjelas yang digunakan. Setelah itu, mencari metode klasifikasi yang terbaik berdasarkan data yang digunakan dengan cara membandingkan performa klasifikasi yang dihasilkan dari beberapa metode klasifikasi yang digunakan. Tujuan selanjutnya adalah mengevaluasi kebaikan prediksi yang dihasilkan. Manfaat dari penelitian ini adalah agar dapat berkontribusi untuk pemerintah dan badan terkait dalam mitigasi risiko bencana dan peringatan dini. Selain itu, penelitian ini juga dapat berkontribusi untuk penelitian selanjutnya dalam mengembangkan model-model yang lebih baik untuk memprediksi gempa bumi di masa depan.

II. METODE PENELITIAN

2.1. Dataset

Penelitian ini menggunakan data gempa bumi yang bersumber dari website The United States Geological Survey (USGS). Data yang diambil adalah data kejadian gempa bumi di wilayah Indonesia dan perairan sekitarnya seperti yang tampak pada Gambar 1 di bawah. Secara detail, letak geografis yang tercakup adalah pada latitude : -11,58 - 6,208 dan longitude : 94,84 - 141,39 dalam jangka waktu 1 dekade (10 Mei 2015 - 10 Mei 2024) dan 7 hari (26 Mei 2024 - 1 Juni 2024). Data penelitian ini terdiri dari 21 atribut, yaitu Time, Latitude, Longitude, Depth, Magnitude, Magnitude Type, Magnitude Seismic Stations, Azimuthal Gap, Horizontal Distance, Root Mean Square, ID, Updated, Place, Type, Horizontal Error, Depth Error, Magnitude Error, Location Seismic Stations, Status, Location Source, dan Magnitude Source. Jumlah record pada data tersebut adalah sebanyak 18.995 dengan type Earthquake. Berdasarkan hubungan variabel penjelas dengan kedalaman gempa bumi, hanya terdapat 8 atribut yang digunakan pada penelitian ini, antara lain Depth, Latitude, Longitude, Magnitude, Magnitude Seismic Stations, Location Seismic Stations, dan Gap. Variabel Depth digunakan sebagai labeling pada proses pengklasifikasian serta sebagai variabel yang diprediksi.



Gambar 1. Peta Indonesia

2.2. *Naïve Bayes*

Berdasarkan Marathe et al. [12], Naïve Bayes merupakan algoritma machine learning sederhana dengan menggunakan teorema Bayes serta asumsi independensi antar fitur. Hal tersebut berarti algoritma mengasumsikan seluruh variabel penjelas yang digunakan saling independen. Teorema Bayesian memberikan persamaan untuk menghitung probabilitas posterior ($c|x$) dari $P(c)$, $P(x)$, dan $P(x|c)$:

$$P(c|x) = (P(x|c)P(c))/P(x) \quad (1)$$

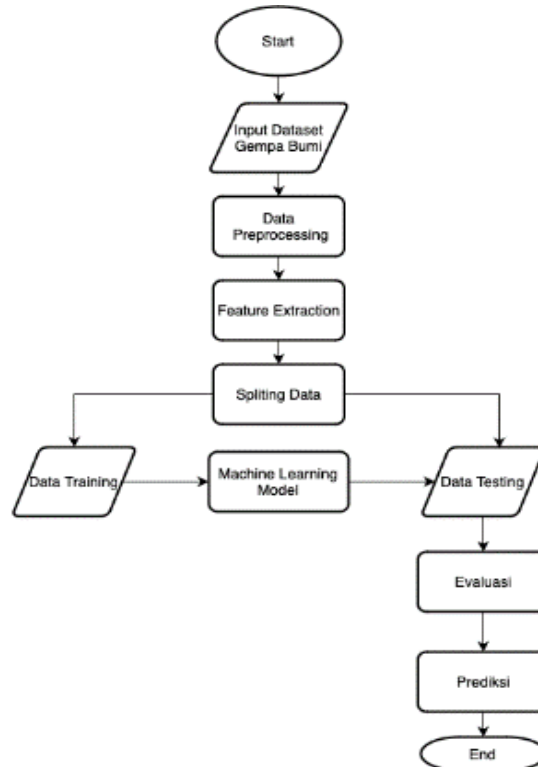
Algoritma Naïve Bayes memiliki kinerja yang baik pada dimensi data tinggi. Alur kerjanya adalah dengan menghitung probabilitas kelompok dengan mengasumsikan setiap atribut independen terhadap kelompoknya. Teorema Bayes kemudian digunakan untuk memperkirakan batas tertentu untuk klasifikasi tiap kelompok. Penelitian ini menganalisis data dengan dimensi tinggi, karena hal tersebut maka penelitian ini menggunakan algoritma Naïve Bayes untuk memprediksi kedalaman gempa bumi.

2.3. *K-Nearest Neighbor (KNN)*

Berdasarkan Fix dan Hodge, K-Nearest Neighbor (KNN) merupakan algoritma yang paling banyak digunakan karena dianggap sederhana. Cara kerja algoritma dari KNN yaitu dengan menghitung jarak antara titik data training dan data testing untuk mengidentifikasi tetangga terdekatnya. Sampel baru kemudian masuk ke kelas tetangga terdekatnya. K merupakan jumlah tetangga terdekatnya sehingga klasifikasi akan sangat tergantung dengan K [13]. KNN sangat efektif untuk data yang mengandung noise. Oleh sebab itu, KNN digunakan dalam penelitian ini untuk memprediksi kedalaman bumi.

2.4. *Random Forest*

Random Forest merupakan algoritma berdasarkan decision tree namun memiliki kinerja yang lebih baik daripada algoritma decision tree. Random Forest membangun decision tree dengan ukuran yang besar dan kemudian menggabungkannya agar mendapatkan prediksi yang akurat. Random Forest sangat efektif untuk menangani data yang overfitting [5]. Oleh karena itu, Random Forest digunakan dalam penelitian ini untuk memprediksi kedalaman gempa bumi.



Gambar 2. Flowchart Tahapan Analisis

Berdasarkan gambar 2, tahapan analisis dapat dijelaskan berikut:

1. Melakukan input dataset yang akan digunakan;
2. Melakukan preprocessing pada data yaitu mendeteksi missing data, melakukan penyesuaian pada tipe data, serta melakukan recode;
3. Melakukan Feature Extraction yaitu melakukan reduksi pada feature atau variabel dari data mentah sehingga mendapatkan feature akhir yang digunakan untuk analisis;
4. Membagi data menjadi data training dan data testing dengan proporsi 75:25, data training merupakan data digunakan untuk mengembangkan model. Sementara itu, data testing merupakan data yang digunakan untuk menguji kualitas klasifikasi model setelah proses training selesai ;
5. Membentuk model dari data training dengan algoritma Naïve Bayes, KNN, dan Random Forest;
6. Menguji model dengan data testing;
7. Melakukan evaluasi dengan menggunakan nilai Accuracy, Sensitivity, Specificity, dan F1 Score untuk mendapatkan model terbaik;
8. Model terbaik akan digunakan untuk memprediksi kedalaman gempa bumi.

2.6. Data Preprocessing

Persiapan data dilakukan dalam beberapa tahap, yakni :

1. Mengubah tipe data menjadi numerik untuk semua variabel yang akan digunakan, karena data yang ada memiliki tipe integer.
2. Melakukan recode untuk data yang bertipe faktor, yaitu data depth berdasarkan klasifikasi jarak hiposentrum dengan permukaan bumi oleh Badan Nasional Penanggulangan Bencana (BNPB),



yaitu gempa bumi dalam (berkode 1; depth > 300 km), gempa bumi menengah (berkode 2; 60 < depth < 300), dan gempa bumi dangkal (berkode 3; depth < 60).

3. Mendeteksi dan menangani missing data; dalam kasus ini karena data yang terdeteksi sebagai outlier hanya 99 dari 18995 data, maka penanganan yang dilakukan adalah melakukan drop (membuang data) terhadap missing data tersebut.

2.7. Training dan Fase Testing

Data yang ada dibagi menjadi data training dan testing untuk mengukur seberapa baik model yang telah dibentuk dengan data training dalam melakukan prediksi klasifikasi pada data testing. Pada penelitian ini terdapat 18.896 data kejadian gempa bumi di wilayah Indonesia selama 10 tahun terakhir, sehingga jumlah data training dan testing berturut-turut adalah 14.172 dan 4.724 yang kemudian akan dimodelkan dan diukur kualitasnya dengan performance measurements.

2.8. Indeks Pengukuran Performa Model

Untuk mengukur seberapa baik kualitas klasifikasi model, confusion matrix akan dibuat untuk data aktual dan prediksi dibentuk untuk menghitung nilai-nilai yang :

- True Positive (Correctly Identified)
- True Negative (Incorrectly Identified)
- False Positive (Correctly Rejected)
- False Negative (Incorrectly Rejected)

[22]

Nilai tersebut kemudian akan dijadikan dasar perhitungan dari performance measurement dengan rumus :

[14]

- Accuracy

$$accuracy = (TP + TN)/(TP + FP + FN + TN) \quad (2)$$

- Precision

$$precision = (TP)/(TP + FP) \quad (3)$$

- Recall

$$recall = (TP)/(TP + FN) \quad (4)$$

- F1 Score

$$F1\ Score = (2 * (precision * recall))/(precision + recall) \quad (5)$$

III. HASIL DAN PEMBAHASAN

Tabel di bawah merupakan hasil dari confusion matrix berbagai metode klasifikasi dengan machine learning. Confusion matrix ini menggunakan data testing, karena bertujuan untuk melakukan evaluasi kebaikan model machine learning.



Tabel 1. Confusion Matrix Metode Klasifikasi Naive Bayes Terhadap Kedalaman Gempa Bumi

Kategori	Predicted		
Observed	1	2	3
1	2	199	55
2	6	1183	622
3	3	1237	1417

Tabel 2. Confusion Matrix Metode Klasifikasi KNN (k=22) Terhadap Kedalaman Gempa Bumi

Kategori	Predicted		
Observed	1	2	3
1	122	100	34
2	15	1214	582
3	8	452	2197

Tabel 3. Confusion Matrix Metode Klasifikasi Random Forest Terhadap Kedalaman Gempa Bumi

Kategori	Predicted		
Observed	1	2	3
1	220	10	8
2	15	1448	284
3	21	353	2365

Untuk mengetahui kualitas suatu metode dalam melakukan klasifikasi, dilakukan beberapa langkah pemeriksaan, yaitu melihat performance measurement dari metode klasifikasi yang digunakan dan deteksi kecocokan atau fitness dari model. Langkah pertama yang dilakukan adalah melakukan penilaian terhadap confusion matrix. Confusion matrix dari suatu model yang sempurna (memiliki performa 100%) akan memiliki nilai di diagonalnya saja, sementara sel lain akan bernilai 0, sementara confusion matrix dari suatu model yang buruk akan mengandung nilai 0 di minimal satu dari diagonalnya [15]. Confusion matrix untuk model klasifikasi yang digunakan pada penelitian ini dapat dilihat seperti pada Tabel 1, 2, dan 3. Terlihat bahwa tidak ada diagonal yang memiliki nilai 0 dan nilai-nilai pada sel selain diagonal tergolong kecil, sehingga dapat dikatakan bahwa model-model yang digunakan sudah baik dan dapat dievaluasi.

Tabel 4. Confusion Matrix Metode Klasifikasi Random Forest Terhadap Kedalaman Gempa Bumi

Metode	Accuracy	Performance Measurement Index		
		Precision	Recall	F1 Score
<i>Naive Bayes</i>	55%	0,78%	18,18%	1,49%
<i>KNN (k=22)</i>	74,78%	47,66%	84,14%	60,84%
<i>Random Forest</i>	85,37%	92,44%	85,94%	89,07%

Setelah diketahui bahwa model klasifikasi yang ada menghasilkan confusion matrix yang baik, confusion matrix tersebut menjadi dasar untuk menghitung beberapa performance measurement index yang akan digunakan untuk memilih model terbaik. Hasil performance measurement di Tabel 4 digunakan untuk memilih metode terbaik dalam melakukan klasifikasi kedalaman gempa bumi. Terlihat bahwa dari perbandingan nilai accuracy, precision, recall, dan F1 score dari tiga model klasifikasi yang digunakan, model Random Forest terpilih menjadi model dengan performance terbaik dalam melakukan klasifikasi kedalaman gempa bumi. Hasil ini sejalan dengan penelitian oleh Mangalathu et al. [16], terkait klasifikasi dampak gempa bumi terhadap bangunan di wilayah bencana, yaitu Kota Napa. Mangalathu et al. menggunakan 4 metode klasifikasi dan dari evaluasi yang dilakukan, diketahui bahwa model Random Forest memiliki performance klasifikasi terbaik, dengan akurasi sebesar 66%.



Penelitian lainnya oleh Aden-Antoniów et al. [17] mengenai pembentukan clustering catalog terkait gempa bumi yang terjadi di Selatan Kalifornia dan Selandia Baru menyatakan bahwa model Random Forest dengan akurasi mendekati sempurna, yaitu 98%, adalah model machine learning terbaik dalam melakukan declustering untuk mendapatkan seismic-based earthquake catalog. Penelitian lain terkait fenomena geografis yakni tanah longsor di Prancis oleh Provost et al. [18] tahun 2017 menyatakan bahwa Random Forest tetap memberikan performa klasifikasi bencana akibat tanah longsor, dengan kategori batu jatuh, gempa, gempa bumi, dan peristiwa antropogeni. Dasar pemilihan model terbaik yang digunakan adalah sensitivity sebesar 93% dengan galat 1,5%.

Tabel 5. Confusion Matrix Metode Klasifikasi Pembentukan Model Random Forest untuk Data Training

Kategori Observed	Predicted		
	1	2	3
1	220	15	21
2	9	1450	352
3	8	283	2365

Tabel 6. Perbandingan Accuracy Metode Klasifikasi Random Forest

Data	Accuracy
Training	85,41%
Testing	85,37%

Berikutnya, pada model terpilih dilakukan pengecekan measurements performance. Kondisi data yang dihindari adalah overfit dan underfit. Kondisi overfit terjadi saat model tidak dapat digeneralisasi dengan baik, sehingga model akan memiliki performa yang sempurna di set data training, namun menurun kualitasnya saat digunakan pada data test. Model dengan kondisi underfit akan cenderung untuk mengingat semua data termasuk noise yang tidak terhindarkan pada data training daripada mempelajari pola dibalik pengklasifikasian data tersebut [19]. Untuk melihat terjadi atau tidaknya kondisi ini, dilakukan perbandingan nilai accuracy data testing dengan data training yang dihitung dengan confusion matrix pada Tabel 5. Pada model terpilih, di Tabel 6 terlihat bahwa selisih nilai accuracy antara data testing dan training tidak berbeda jauh, yakni sebesar 0,00042337 atau 0,042%. Hasil ini menunjukkan model Random Forest sudah berhasil melakukan pengklasifikasian dengan generalisasi yang baik dan minim galat.

Tabel 7. Confusion Matrix Metode Klasifikasi Pembentukan Model Random Forest untuk Data 7 Harian

Kategori Observed	Predicted		
	1	2	3
1	2	0	0
2	0	9	2
3	0	1	15
Accuracy	89,66%		

Kemudian, dilakukan perhitungan untuk melihat apakah performance Random Forest benar memiliki kualitas pengukuran dan relevansi yang baik. Data testing kedua yang berjangka waktu tujuh hari dinilai kualitas klasifikasinya menggunakan model yang sama dengan model klasifikasi data 1 dekade. Dari hasil perhitungan confusion matrix pada Tabel 7, didapatkan nilai performance measurement untuk data kecil. yang didapat tidak jauh berbeda dengan data yang sebelumnya, yakni sebesar 89,66%. Hasil tersebut sesuai dengan penelitian oleh Han et al. [20] mengenai Pemetaan dan Prediksi Klasifikasi Zona Aman Gempa di Kota Gyeongju, Korea Selatan. Penelitian tersebut menyatakan bahwa Random Forest memiliki kualitas terbaik untuk melakukan prediksi klasifikasi



berbahaya atau amannya suatu wilayah dengan success rate sebesar 100% dan akurasi sebesar 96,9%. Penelitian lainnya oleh Qin et al. [21], terkait prediksi penyebab gempa di Oklahoma dengan machine learning. Penelitian ini menyatakan bahwa dengan indikator berupa R Squared, Random Forest berhasil dengan baik melakukan prediksi terhadap kemungkinan terjadinya gempa dalam beberapa periode waktu (30, 60, 90, dan 180 hari) ke depan.

IV. KESIMPULAN

Berdasarkan hasil dan pembahasan sebelumnya, dapat ditarik kesimpulan bahwa confusion matrix yang diperoleh sudah dapat mengevaluasi model dengan baik. Selain itu, berdasarkan nilai accuracy, precision, recall, dan F1 score, didapatkan metode yang terbaik dalam mengklasifikasikan gempa bumi adalah metode Random Forest. Kemudian dari hasil prediksi klasifikasi dengan metode Random Forest berdasarkan R squared, metode Random Forest dapat memprediksi kategori dari kedalaman suatu gempa, baik dengan data testing maupun data kecil (7 harian).

REFERENSI

1. Harsh K. Gupta, Encyclopedia of Earth Sciences Series. 2021.
2. R. Agarwal, “Effects of earthquakes on environment,” 2020, [Online]. Available: <https://www.researchgate.net/publication/342262746>
3. W. Spence, S. A. Sipkin, and G. L. Choy, “USGS,” Earthquakes and Volcanoes.
4. M. Lujan and D. L. Peck, “United States Department Of The Interior,” 1989.
5. M. Mondol, “Analysis and Prediction of Earthquakes using different Machine Learning techniques,” 2021. [Online]. Available: www.kaggle.com/usgs/earthquake-database
6. R. Jain, A. Nayyar, S. Arora, and A. Gupta, “A comprehensive analysis and prediction of earthquake magnitude based on position and depth parameters using machine and deep learning models,” *Multimed Tools Appl*, vol. 80, no. 18, pp. 28419–28438, Jul. 2021, doi: 10.1007/s11042-021-11001-z.
7. H. Tantyoko, D. Kartika Sari, and A. R. Wijaya, “Prediksi Potensial Gempa Bumi Indonesia Menggunakan Metode Random Forest Dan Feature Selection,” 2023. [Online]. Available: <http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/indexHenriTantyoko|http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index>
8. J. Macedo, C. Liu, and F. Soleimani, “Machine-learning-based predictive models for estimating seismically-induced slope displacements,” *Soil Dynamics and Earthquake Engineering*, vol. 148, Sep. 2021, doi: 10.1016/j.soildyn.2021.106795.
9. R. Tehseen, M. S. Farooq, and A. Abid, “Fuzzy expert system for earthquake prediction in western Himalayan range,” *Elektronika ir Elektrotehnika*, vol. 26, no. 3, pp. 4–10, Jun. 2020, doi: 10.5755/j01.eie.26.3.25744.
10. M. A. Siddiquee, V. M. A. Souza, G. E. Baker, and A. Mueen, “Septon: Seismic Depth Estimation Using Hierarchical Neural Networks,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2022, pp. 3889–3897. doi: 10.1145/3534678.3539166
11. Y. Rong, D. D. Jackson, and Y. Y. Kagan, “Seismic gaps and earthquakes,” *J Geophys Res Solid Earth*, vol. 108, no. B10, Oct. 2003, doi: 10.1029/2002jb002334.
12. N. Marathe, S. Gawade, and A. Kanekar, “Prediction of Heart Disease and Diabetes Using Naive Bayes Algorithm,” *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pp. 447–453, May 2021, doi: 10.32628/cseit217399.
13. S. N. Pasha, D. Ramesh, S. Mohmmad, A. Harshavardhan, and Shabana, “Cardiovascular disease prediction using deep learning techniques,” in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing Ltd, 2020. doi: 10.1088/1757- 899X/981/2/022006.
14. M. A. Salam, A. T. Azar, M. S. Elgendy, and K. M. Fouad, “The Effect of Different Dimensionality Reduction Techniques on Machine Learning Overfitting Problem,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 4, pp. 641–655, 2021, doi: 10.14569/IJACSA.2021.0120480.



15. H. Yun, “Prediction model of algal blooms using logistic regression and confusion matrix,” *International Journal of Electrical and Computer Engineering*, vol. 11, no. 3, pp. 2407– 2413, Jun. 2021, doi: 10.11591/ijece.v11i3.pp2407-2413.
16. S. Mangalathu, H. Sun, C. C. Nweke, Z. Yi, and H. V. Burton, “Classifying earthquake damage to buildings using machine learning,” *Earthquake Spectra*, vol. 36, no. 1, pp. 183– 208, Feb. 2020, doi: 10.1177/8755293019878137.
17. F. Aden-Antoniów, W. B. Frank, and L. Seydoux, “An Adaptable Random Forest Model for the Declustering of Earthquake Catalogs,” *J Geophys Res Solid Earth*, vol. 127, no. 2, Feb. 2022, doi: 10.1029/2021JB023254.
18. F. Provost, C. Hibert, and J. P. Malet, “Automatic classification of endogenous landslide seismicity using the Random Forest supervised classifier,” *Geophys Res Lett*, vol. 44, no. 1, pp. 113–120, Jan. 2017, doi: 10.1002/2016GL070709.
19. X. Ying, “An Overview of Overfitting and its Solutions,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Mar. 2019. doi: 10.1088/1742-6596/1168/2/022022.
20. J. Han, J. Kim, S. Park, S. Son, and M. Ryu, “Seismic vulnerability assessment and mapping of Gyeongju, South Korea using frequency ratio, decision tree, and random forest,” *Sustainability (Switzerland)*, vol. 12, no. 18, Sep. 2020, doi: 10.3390/SU12187787.
21. Y. Qin, T. Chen, X. Ma, and X. Chen, “Forecasting induced seismicity in Oklahoma using machine learning methods,” *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022- 13435-3.
22. M, H., & M.N, S. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2), 01–11. <https://doi.org/10.5121/ijdkp.2015.5201>