Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2025. Vol. 4. No. 2; e-ISSN: 2808-4519

Implementation of Logistic Regression Algorithm in Predicting Tsunami Potential on Earthquake Data Parameters

Sofian Wira Hadi[1], Ibnu Alfarobi[2], Irmawati[3*]

[1,2,3] *Universitas Bina Sarana Informatika*
irmawati.iat@bsi.ac.id [3*]

Abstract

This study presents the evaluation and testing of a logistic regression model for predicting earthquake-related features, including earthquake depth, magnitude, and tsunami potential. The model achieved high accuracy in predicting earthquake depth categories (99.82%) and earthquake magnitude (99.84%), but faced challenges with low recall for tsunami prediction (50%) due to class imbalance. Evaluation results showed that the model struggled to predict tsunami occurrence accurately, as the dataset contained a disproportionate number of 'no tsunami' instances. Despite these limitations, the model displayed high accuracy for earthquake depth and magnitude predictions. The testing phase revealed a series of prediction errors, particularly for the tsunami category, influenced by the imbalance in training data. The results emphasize the need for improved handling of imbalanced datasets and the potential for exploring other machine learning algorithms and techniques for better performance in multiclass classification problems. Future research could further refine these models by incorporating additional criteria and exploring other earthquake and tsunami prediction methodologies.

Keywords: *Tsunami; Earthquake; Depth; Magnitude; Logistics Regression;*

1. Introduction

Indonesia is one of the countries with a high risk of natural disasters, particularly earthquakes and tsunamis. This is due to its geographical position at the intersection of three major tectonic plates: the Indo-Australian Plate, the Eurasian Plate, and the Pacific Plate. As a result, Indonesia is a region with very high seismic activity, making the development of disaster mitigation systems highly important [1][2]. Tsunamis, as one of the impacts of tectonic earthquakes, have caused significant losses in both economic terms and human casualties in Indonesia. The 2004 Aceh tsunami tragedy stands as one of the largest events in history, highlighting the critical importance of early warning systems to reduce disaster impacts [3]. Therefore, research aimed at predicting tsunami potential based on earthquake parameters becomes highly relevant [4]. In recent years, machine learning technology has been increasingly used to model natural disaster potentials, including tsunamis. This technology enables the analysis of large datasets to uncover significant patterns that may not be evident through traditional methods [5]. In Indonesia, earthquake data collected by the Meteorology, Climatology, and Geophysics Agency (BMKG) serves as one of the main sources for developing predictive models [6]. This study uses a dataset from BMKG that includes earthquake data in Indonesia during the period from 2008 to 2023. The dataset contains information about parameters such as magnitude, depth, location, and time of earthquake occurrences. With more than 92,000 data points, this dataset provides a strong foundation for building machine learning-based predictive models.

Parameters such as the depth and magnitude of earthquakes are known to have a significant relationship with the potential for tsunamis. Earthquakes with depths of less than 100 km and magnitudes greater than 7.0 on the Richter scale generally have the potential to trigger tsunamis [7]. Therefore, this research develops a predictive model focusing on these parameters. The stages of this research include dataset exploration, data preprocessing, model training, model evaluation, and model implementation. The data exploration step aims to understand the distribution of parameters and identify possible issues such as duplicate data or missing values. Meanwhile, the preprocessing stage involves data transformation and the removal of irrelevant features, such as earthquake focal mechanisms [8]. The logistic regression algorithm was chosen to build the predictive model in this study. This algorithm allows multi-output classification to predict categories of depth, magnitude, and tsunami potential based on earthquake parameters. Additionally, logistic regression has good interpretability, which is crucial in disaster mitigation applications [4][9]. The developed model was evaluated using metrics such as accuracy, precision, and recall to ensure consistent and reliable prediction results. This study aims to produce a model that is not only accurate but also efficient for implementation in tsunami early warning systems[7]. The results of this research are expected to provide a tangible contribution to disaster risk reduction in Indonesia. By leveraging historical earthquake data and machine learning technology, this predictive model is anticipated to become a crucial part of decision-making for tsunami disaster mitigation [10].

2. Research Method

Figure 1 below illustrates the research methodology stages for developing a machine learning model that can predict whether a specific value of earthquake parameters indicates potential for a tsunami. These stages include dataset exploration, data preprocessing, model training, model evaluation, and finally, model deployment.

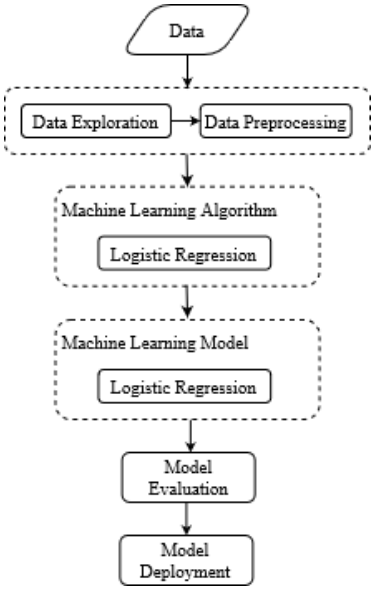


Fig. 1: Research Method

2.1. Data

The dataset used in this study is sourced from secondary data obtained from the Kaggle website [https://www.kaggle.com/datasets/kekavigi/earthquakes-in-indonesia]. The dataset includes several earthquake parameters recorded by BMKG Indonesia over the period from November 1, 2008, to January 26, 2023. This dataset contains 13 features and 92,887 rows of data. The features include 'tgl', 'ot', 'lat', 'lon', 'depth', 'mag', 'remark', 'strike1', 'dip1', 'rake1', 'strike2', 'dip2', and 'rake2'. Table 1 below provides a description of each attribute.

Table 1: Features Description

No	Features	Description
1	tgl	Data of occurence
2	ot	Event timestamp
3	lat	The latitude of the event's epicenter (degrees), ranging from 6N to 11S
4	lon	The longitude of the event's epicenter (degress), ranging from 142E to 94E
5	depth	Depth of occurrence (Km)
6	mag	Magnitude of events, ranging from 1 to 9.5
7	remark	The area of incident
8	strike1	Sometimes, the focal mechanism of event is measured in such a case.
9	dip1	Sometimes, the focal mechanism of event is measured in such a case.
10	rake1	Sometimes, the focal mechanism of event is measured in such a case.
11	strike2	Sometimes, the focal mechanism of event is measured in such a case.
12	dip2	Sometimes, the focal mechanism of event is measured in such a case.
13	rake2	Sometimes, the focal mechanism of event is measured in such a case.

2.2. Data Exploration

The data exploration stage involves checking the content of the data, verifying data types, checking for missing values, identifying duplicate data, and examining the descriptive statistics of the data.

2.3. Data Preprocessing

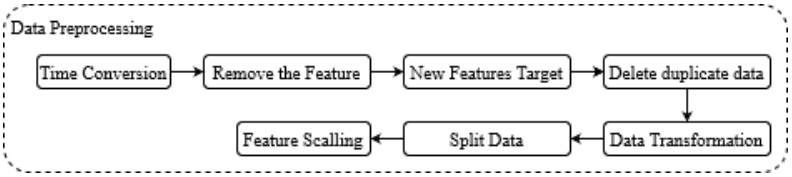


Fig. 2: Data Preprocessing

Figure 2 above illustrates the steps of data preprocessing in this study. The first step in the data preprocessing stage is converting the time feature 'tgl' into new features: 'year', 'month', and 'day'. The purpose of this process is to analyze and better understand the data being used, thereby extracting valuable information. This analysis includes determining the number of earthquakes that occurred each year from November 1, 2008, to January 26, 2023, and analyzing events by month and day over 16 years.

The second data preprocessing step involves removing features that are not significant earthquake parameters and features that contain missing data. The removed features include 'strike1', 'dip1', 'rake1', 'strike2', 'dip2', and 'rake2'.

The third step is creating three new features to be used as target variables for the model. These new features are derived from existing earthquake parameter features, with the modeling approach referred to as a multi-output model. The first feature, 'depth_category,' consists of five unique data categories based on earthquake depth (Km) classification provided by BMKG (Meteorology, Climatology, and Geophysics Agency) and is related to the 'depth' feature [https://inatews.bmkg.go.id/web/realtime]. The second feature, 'mag_category,' contains six unique data categories related to earthquake magnitude in the 'mag' feature. The third feature, 'tsunami,' has two unique categories that indicate whether an earthquake has the potential to trigger a tsunami or not, based on specific conditions: earthquakes with depths less than 100 km and magnitudes greater than 7.0 on the Richter scale [https://stageof-tretes.bmkg.go.id/informasi-tsunami]. The details of these conditions are presented in Table 2 below.

Table 2: Data Conditions of the Target Model Features

Features	Class	Conditions
depth_category	<=50	Data with a depth of less than 50 Km
	<=100	Data with depths between 51 to 100 Km
	<=250	Data with depths between 101 to 250 Km
	<=600	Data with depths between 251 to 600 Km
	>600	Data with a depth of more than 600 Km
mag_category	micro	Data with a magnitude of less than 3.0 on the Richter Scale
	minor	Data with a magnitude of less than 4.0 on the Richter Scale
	light	Data with a magnitude of less than 5.0 on the Richter Scale
	medium	Data with a magnitude of less than 6.0 on the Richter Scale
	robust	Data with a magnitude of less than 7.0 on the Richter Scale
	mayor	Data with a magnitude greater than 7.0 on the Richter Scale
tsunami	tsunami	Data with a magnitude greater than 7.0 on the Richter Scale and a depth of less than 100 Km
	tidak tsunami	Not in accordance with these provisions means that there is no potential for a tsunami

The fourth step involves rechecking whether duplicate data exists after removing the aforementioned features. If duplicates are identified, they are removed.

The fifth data preprocessing step is transforming the target features, namely 'depth_category,' 'mag_category,' and 'tsunami,' to convert categorical string data into numerical categorical data. This transformation is necessary because the logistic regression machine learning algorithm only works with numerical data.

The sixth step is splitting the data into X and Y datasets. X represents the input data for the model, consisting of features 'lat,' 'lon,' 'depth,' and 'mag.' Y represents the target data for the model, consisting of the features 'depth_category,' 'mag_category,' and 'tsunami.' Next, X and Y datasets are further divided into training and testing datasets, with 80% used for training and 20% for testing. The data splitting process is performed using a random state parameter set to 42.

The final data preprocessing step is normalizing the data using the standard scaler function from Scikit-learn, as shown in Equation (1) [Machine Learning for the Model Prediction of Final Semester Assessment (FSA) using the Multiple Linear Regression Method].

$$X_{standard} = \frac{X-\mu}{\sigma} \quad (1)$$

2.4. Model Training

The model training in this study uses 80% of the training data that has undergone data preprocessing. Figure 3 below illustrates the model structure trained using the Multi Output Classifier algorithm, with Logistic Regression as the estimator algorithm. The hyperparameters for Logistic Regression are set as follows: C=1C=1C=1, max_iter=1000\text{max_iter}=1000max_iter=1000, and random_state=42\text{random_state}=42random_state=42. Additionally, the Logistic Regression estimator algorithm includes the following hyperparameters: C=1C=1C=1, max_iter=1000\text{max_iter}=1000max_iter=1000, penalty=L2\text{penalty}=L2penalty=L2, and random_state=42\text{random_state}=42random_state=42.

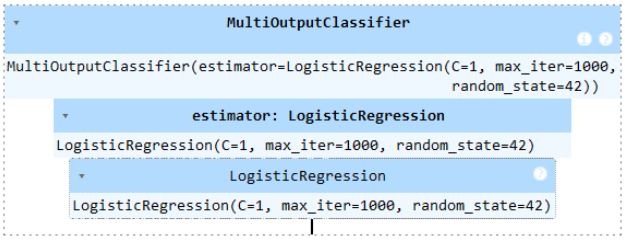


Fig. 3: Model Training

2.5. Model Evaluation

E The model evaluation is conducted using performance metrics commonly employed in machine learning, such as accuracy, recall, precision, and F1-score.
The **accuracy metric** (Equation 2) is calculated as the ratio of correctly predicted data to the total number of predictions. The **recall metric** (Equation 3) answers the question: of all the data that is truly relevant, how much was correctly identified? The **precision metric** (Equation 4) determines, out of the data predicted as relevant by the model, how much of it is actually relevant. Lastly, the **F1-score metric** (Equation 5) provides a weighted average comparison between precision and recall [On the Role of Text Preprocessing in BERT Embedding-based DNNs for Classifying Informal Texts].
These equations reflect the overall performance of the model for the case studied, which involves a multiclass model. The formulas include the following parameters: **TP (True Positive):** The number of correct predictions for a given class, e.g., positive. **TN (True Negative):** The number of correct predictions for another class, e.g., negative. **FP (False Positive):** The number of incorrect predictions, where the model mistakenly predicts a positive class. **FN (False Negative):** The number of incorrect predictions, where a negative class is mistakenly predicted during model evaluation. These metrics collectively provide a comprehensive assessment of the model's performance.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$recall = \frac{TP}{TP+FN} \quad (3)$$

$$precision = \frac{TP}{TP+FP} \quad (4)$$

$$f1 - score = \frac{2 \times precision \times recall}{precision+recall} \quad (5)$$

2.6. Model Deployment

Figure 4 below shows the steps of the process for deploying the model with the goal of testing the model's ability to predict the potential occurrence of a tsunami based on new input data. The system is built using the Streamlit library for the frontend and backend, which contains the deployed model and feature scaling. The results of the input will be directly displayed on the frontend page, and the program is run locally. The process starts with data input into the system, followed by data preprocessing in the form of data normalization using the normalization results from the training data that has been stored. Then, the trained and evaluated model is integrated into the system. Once the user clicks the predict button, if the input data is valid, the system will output the predicted probability values for each class of the model's target features along with the predictions for each feature. If the input data is invalid or any column is left empty, the model's prediction will not be accurate.

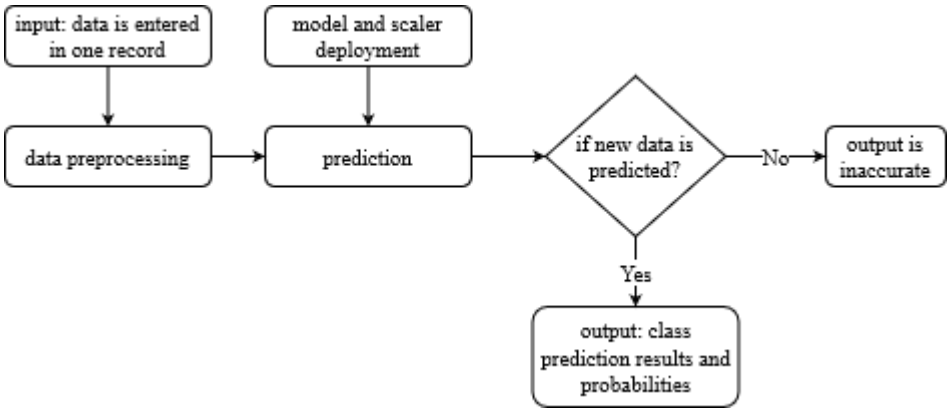


Fig. 4: Model Deployment for Prediction

3. Result and Discussion

3.1. Data Exploration Result

Table 3 provides a detailed overview of the data exploration results, which involved checking the contents of each attribute. The values in the "values of sample of data" column are consistent with their respective data types. The dataset contains 3 features with object or string data type, 9 features with float or decimal data type, and 1 feature with an integer data type. Additionally, there are 90,152 missing data entries in the features 'strike1', 'dip1', 'rake1', 'strike2', 'dip2', 'rake2', and a total of 540,912 missing values across all rows for each feature. Furthermore, no duplicate data was found in the existing dataset source.

Table 3: Data Exploration Results

No	Features	Data Type	Missing Values	Values Sample of Data
1	tgl	Object	0	2008/11/01
2	ot	Object	0	21:02:43.058
3	lat	Float	0	-9.18
4	lon	Float	0	119.06

5	depth	Int	0	10
6	mag	Float	0	4.9
7	remark	Object	0	Sumba Region - Indonesia
8	strike1	Float	90.152	NaN
9	dip1	Float	90.152	NaN
10	rake1	Float	90.152	NaN
11	strike2	Float	90.152	NaN
12	dip2	Float	90.152	NaN
13	rake2	Float	90.152	NaN

3.2. Data Preprocessing Result

Figure 5 shows the number of earthquake incidents in each year that occurred in Indonesia. The highest number of earthquakes occurred in 2018, with 12,345 incidents. The fewest earthquakes were recorded in 2008, with only 316 incidents, as this period only covers the months of November and December. In 2023, there were only 1,176 incidents, all of which occurred in January. Figure 5 also represents the number of earthquake incidents by month over a span of 16 years, from November 1, 2008, to January 26, 2023. The highest number of incidents occurred in December, with a total of 8,990 incidents. Figure 6 shows the month with the fewest earthquakes, which is June, with 6,472 incidents. Finally, Figure 7 shows the number of earthquake incidents by day over 16 years, with the highest number occurring on Monday, totaling 13,725 incidents, and the fewest occurring on Saturday, with 12,549 incidents.

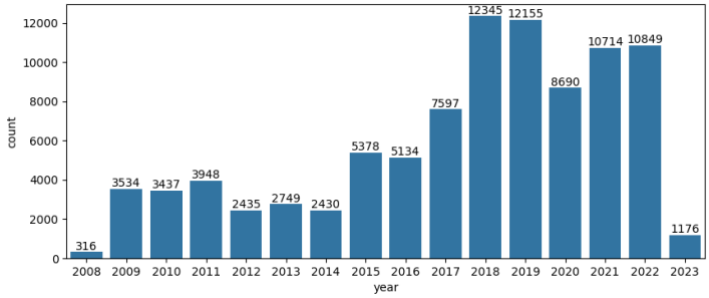


Fig. 5: Amount of Data in a Year

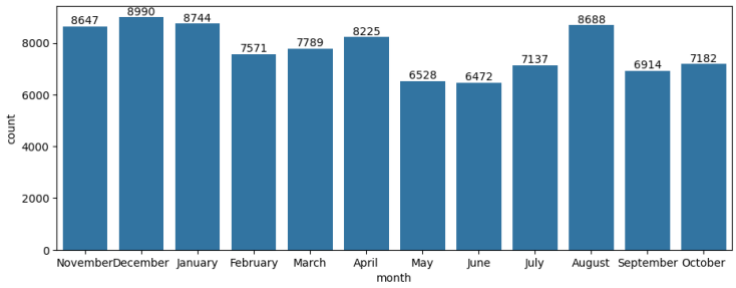


Fig. 6: Amount of Data in a Month

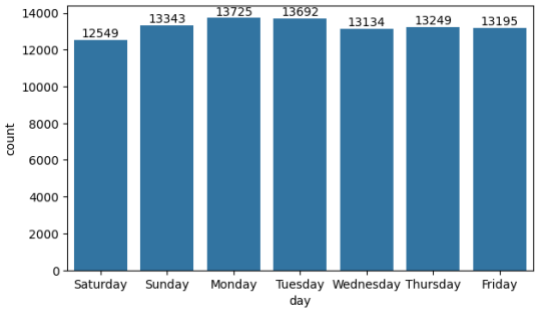


Fig. 7: Amount of Data in a Day

After the removal of 6 features during the data preprocessing stage, 2,735 duplicate entries were found. Table 4 below shows the distribution of data from the model's target features based on class labels. For the 'depth_category' feature, the most frequent data occurred at depths less than or equal to 50 km, with 68,739 entries, while the least frequent data occurred at depths greater than 600 km, with only 138 entries. In terms of earthquake magnitude, the most common occurred in the minor class, with a magnitude of less than 4.0 on the Richter scale, totaling 40,780 entries, while the least common occurred in the major class, with a magnitude greater than 7.0 on the Richter scale, totaling just 39 entries. Finally, based on the parameters for identifying earthquakes that could potentially trigger a tsunami, 92,864 entries were classified as not having tsunami potential, while only 23 entries were classified as capable of triggering a tsunami.

Table 4: Target Features Data Distribution

Feature	Class	Sum
depth_category	<=50	68.739
	<=100	9.837

	<=250	12.046
	<=600	2.127
	>600	138
mag_category	micro	21.775
	minor	40.780
	light	25.256
	medium	4.670
	robust	367
	mayor	39
tsunami	tsunami	23
	tidak tsunami	92.864

The results of the data transformation are presented in Table 5 below in detail. For the target feature 'depth_category,' the labels were transformed as follows: <=100 became 0, <=250 became 1, <=50 became 2, <=600 became 3, and >600 became 4. For the target feature 'mag_category,' the labels were transformed as follows: light became 0, mayor became 1, medium became 2, micro became 3, minor became 4, and robust became 5. Finally, for the target feature 'tsunami,' the label indicating no potential for a tsunami was changed to 0, and the label indicating a potential tsunami was changed to 1.

Table 5: Data Transformation Result

Features	Target Class	Transformation Result
depth_category	<=50	2
	<=100	0
	<=250	1
	<=600	3
	>600	4
mag_category	micro	3
	minor	4
	light	0
	medium	2
	robust	5
	mayor	1
tsunami	tsunami	1
	tidak tsunami	0

Next, it was found that there were 2,742 duplicate entries, and these were removed, reducing the total data to 90,145 entries. Table 6 below shows the details of the train-test data split, where 80% of the data was used for training, resulting in 72,116 entries, and 20% was used for testing, resulting in 18,029 entries. The final step in feature scaling using the Standard Scaler involved transforming the data for each feature into a scale with values ranging from -1 to 1 for the training data input using fit_transform, and for the testing data input using transform. The results of the training data normalization were saved for later use in model deployment.

Table 6: Split Data Result

Train Data			Test Data		
Features	Target Class	Sum	Features	Target Class	Sum
depth_category	<=50	53.822	depth_category	<=50	13.316
	<=100	7.333		<=100	1.923
	<=250	9.241		<=250	2.354
	<=600	1.625		<=600	411
	>600	95		>600	25
mag_category	micro	17.485	mag_category	micro	4267
	minor	32.529		minor	8.150
	light	19.244		light	4.870
	medium	2.657		medium	687
	robust	183		robust	50
	mayor	18		mayor	5
tsunami	tsunami	12	tsunami	tsunami	2
	tidak tsunami	72.104		tidak tsunami	18.027

3.3. Model Evaluation Result

Table 7 below shows the results of the model evaluation using the logistic regression algorithm with the test data. Based on these results, it is observed that the evaluation accuracy of the model for the earthquake depth feature, 'depth_category,' achieved a very good accuracy of 99.82%, which is fairly balanced with precision, recall, and F1-score values of 99%, 91%, and 94%, respectively. For the earthquake magnitude feature, 'mag_category,' the model achieved an accuracy of 99.84%, but the recall value was quite low at only 75% compared to the accuracy. Lastly, the evaluation for predicting the potential occurrence of a tsunami achieved an accuracy of 99.98%, but the recall, precision, and F1-score values were much lower, with each being only 50%.

Table 7: Evaluation Model Result

Fetures Target Model	Accuracy	Recall	Precision	F1-score
depth_category	99,82%	91%	99%	94%
mag_category	99,84%	75%	80%	77%
tsunami	99,98%	50%	50%	50%

3.4. Model Testing Result

Figure 8 below shows the initial page display of the frontend system created with Streamlit. It features an input column for entering data on parameters such as latitude, longitude, depth, and earthquake magnitude, along with a "Predict" button to obtain the prediction results.

Fig. 8: System Input Page

Table 8 shows the prediction results where the first 5 rows use data from the dataset with known target class labels, and the sixth row tests the model with random input data. The first to third rows of data, based on the input, have actual prediction results indicating a potential 'tsunami,' with a magnitude strength of 'mayor' in the range greater than 7.0 on the Richter scale, and earthquake depth levels for row one being less than or equal to 50 km, and for rows two and three, the earthquake depth is less than or equal to 100 km. The prediction result for the 'depth_category' feature for the first row is '<=50' km with a prediction accuracy of 100%, which matches the actual data. However, for the 'mag_category' feature, the prediction is 'robust' with an accuracy of 93.71%, which is incorrect as the actual data is 'mayor.' Additionally, the prediction for the 'tsunami' feature also results in an error with the label 'no tsunami' and an accuracy of 83.29%, while the actual data is 'tsunami.' Similarly, rows two and three show prediction errors for the 'mag_category' feature with predictions of 'robust,' and for the 'tsunami' feature with 'no tsunami,' while the actual magnitude data is 'mayor,' and the potential for a 'tsunami' is present. However, the model successfully predicted the second row with an accuracy of 99.67% for the 'depth_category' label of '<=100'. In contrast, for the third row, although the prediction was correct, it achieved a relatively low accuracy of 52.93% for the label '<=100'. The errors in predicting the potential for a tsunami or not, with high accuracy for the wrong predictions, may be influenced by the imbalance in the data between the two labels, and the errors in predicting the magnitude strength are also impacted by the imbalance between the labels, leading to a biased model. This is demonstrated by the evaluation results, where the recall and precision values are significantly different from the model's accuracy.

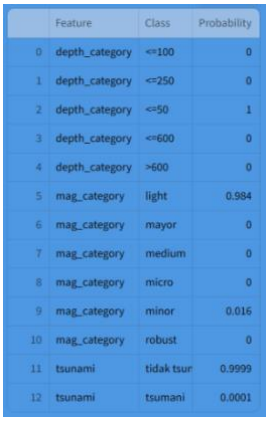
In the fourth and fifth rows of data, the actual prediction data indicates that there is no potential for a 'tsunami,' with an earthquake magnitude of less than 5.0 on the Richter scale in the 'light' label. The actual earthquake depth for row four is less than or equal to 50 km, and for row five, the depth is between 101 and 250 km, with a range of 101 to 250 km. The prediction results for these two rows are quite accurate, with correct predictions of 'no tsunami' and a high level of accuracy. The prediction for the magnitude feature was also correct, with the 'light' label and a high accuracy. Additionally, the prediction for the depth level was accurate, with row four predicting '<=50' and row five predicting '<=250,' achieving good accuracy. These prediction results are also influenced by the fact that the input values dominate in these labels.
In the sixth row, the input data is random for the 'depth' and 'magnitude' features. However, for the 'latitude' and 'longitude' features, the model predicts a potential 'tsunami.' During testing, the model predicted that there was no indication of tsunami potential for this input.

Table 8: Testing Model Result

Input Model				Output Model (Class Target and Probabilitas Value		
Latitude	Longitude	Depth	Magnitudo	deph_category	mag_category	tsunami
-7.59	122.23	10	7.3	<=50 (100%)	Robust (93,71%)	Tidak tsunami (83,29%)
5.88	126.82	75	7.1	<=100 (99,67%)	Robust (81,34%)	Tidak tsunami (88,25%)
-0.82	133.41	51	7.5	<=100 (52,93%)	Robust (84,45%)	Tidak tsunami (85,29%)
-9.18	119.86	10	4.9	<=50 (100%)	Light (98,4%)	Tidak tsunami (99,9%)
-7.07	129.67	135	4.08	<=250 (99,98%)	Light (97,98%)	Tidak tsunami (100%)
-0.82	133.41	269	7.4	<=600 (81,92%)	Robust (82%)	Tidak tsunami (96,7%)

Figure 9 below shows the display of the page when data is entered. In the image, this corresponds to the data in row four, and Figure 10 shows the display of the prediction results for this data.

Fig. 9: Prediction Result (Bar Chart)



	Feature	Class	Probability
0	depth_category	<=100	0
1	depth_category	<=250	0
2	depth_category	<=50	1
3	depth_category	<=600	0
4	depth_category	>600	0
5	mag_category	light	0.984
6	mag_category	mayor	0
7	mag_category	medium	0
8	mag_category	micro	0
9	mag_category	minor	0.016
10	mag_category	robust	0
11	tsunami	tidak tsur	0.9999
12	tsunami	tsumani	0.0001

Fig. 10: Prediction Result (Pie Chart)

4. Conclusion

Based on the analysis of the model testing results, it can be concluded that the model still struggles to identify input data that predicts the potential for a 'tsunami,' due to the extremely small number of data instances indicating a tsunami, which amounts to only 23 entries, compared to 92,864 entries indicating no tsunami potential. However, this data is considered valid because the determination of tsunami potential is based on two criteria provided by BMKG.

Future research suggestions include exploring other machine learning algorithms specifically designed for multiclass classification problems and using methods for handling imbalanced data. Additionally, future research could consider incorporating criteria such as the earthquake category based on its epicenter being located in the open sea and earthquakes with an upfault or downfault pattern, in accordance with BMKG's guidelines for tsunami potential.

References

[1] A. Taufan Maulana And A. Andriansyah, "Mitigasi Bencana Di Indonesia," *Comserva J. Penelit. Dan Pengabdi. Masy.*, Vol. 3, No. 10, Pp. 3996–4012, 2024, Doi: 10.59141/Comserva.V3i10.1213.

[2] A. Z. Z. Abidin, "Implementasi Algoritma C 4.5 Untuk Menentukan Tingkat Bahaya Tsunami," *Semin. Nas. Inform. 2011 (Semnasif 2011)*, Vol. 2011, No. Semnasif, Pp. 29–36, 2011, [Online]. Available: Http://Repository.Upnyk.Ac.Id/620/

[3] Firdaus Laia, Tobias Duha, Mitranikasih Laia, Amirudin Khorul Huda, And Agung Jasuma, "Klasifikasi Data Gempa Bumi Di Pulau Sumatera Menggunakan Algoritma Naïve Bayes," *J. Inform.*, Vol. 2, No. 1, Pp. 23–27, 2023.

[4] M. Iqbal Ramadhan, "Penerapan Data Mining Untuk Analisis Data Bencana Milik Bnpb Menggunakan Algoritma K-Means Dan Linear Regression," *J. Inform. Dan Komput.*, Vol. 22, No. 1, Pp. 57–65, 2017.

[5] D. P. Utomo And B. Purba, "Prosiding Seminar Nasional Riset Information Science (Senaris) Penerapan Datamining Pada Data Gempa Bumi Terhadap Potensi Tsunami Di Indonesia," *Pros. Semin. Nas. Ris. Inf.*, No. September, Pp. 846–853, 2019.

[6] A. Prasetio, M. M. Effendi, And M. N. Dwi M, "Analisis Gempa Bumi Di Indonesia Dengan Metode Clustering," *Bull. Inf. Technol.*, Vol. 4, No. 3, Pp. 338–343, 2023, Doi: 10.47065/Bit.V4i3.820.

[7] B. Muslim, B. Sunardi, S. Rohadi, Y. H. Perdana, And J. Effendi, "Posibilitas Penentuan Magnitude Gempabumi Real Time Menggunakan Gps Untuk Sistem Peringatan Tsunami Di Indonesia A Possibility Of Real Time Earthquake Magnitude Determination Using Gps For Tsunaminwarning System In Indonesia," Vol. 7, No. 2, Pp. 71–76, 2018.

[8] M. Rijal, M. Hamsar, F. Tempola, S. Do Abdullah, And A. A. H. Usman, "Jurnal Sustainable : Jurnal Hasil Penelitian Dan Industri Terapan Penerapan Algoritma Mknn Pada Data Historis Gempa Bumi Yang Berpotensi Tsunami," Vol. 11, No. 01, Pp. 26–33, 2022.

[9] H. Saoud, A. Ghadi, M. Ghailani, And B. A. Abdelhakim, "Using Feature Selection Techniques To Improve The Accuracy Of Breast Cancer Classification," In *Lecture Notes In Intelligent Transportation And Infrastructure*, Vol. Part F1405, Springer Nature, 2019, Pp. 307–315. Doi: 10.1007/978-3-030-11196-0_28.

[10] A. S. Assiri, S. Nazir, And S. A. Velastin, "Breast Tumor Classification Using An Ensemble Machine Learning Method," *J. Imaging*, Vol. 6, No. 6, May 2020, Doi: 10.3390/Jimaging6060039.