Implementation of the Fuzzy K-Nearest Neighbor in Every Class (FKNNC) Method for Tsunami Disaster Potential Classification

Zainal Mu'arif*, Rais Rais, and Mohammad Fajri

Department of Statistics, Faculty of Mathematics and Natural Sciences, Tadulako University, Palu, Indonesia
zainalmuarifstory@gmail.com

Abstract. Tsunamis are sea waves generated by sudden disturbances in the ocean, resulting from changes in the seabed's shape, causing sea water to rise to the surface. Almost every year earthquakes with the potential for tsunamis always occur in all countries in the world. No one knows when a tsunami will occur, but it is possible to predict with classification techniques the potential for tsunami occurrence through the application of science and technology. Through the data contained in historical earthquakes, there is information that can be analysis and processed so that it can determine the potential for a tsunami or not after an earthquake. Analysis and processing of earthquake history data can be done with the data mining process, namely classification techniques. Therefore, this study aims to classify tsunami disaster data based on historical earthquakes with variable attributes of earthquake depth, magnitude, and intensity. The classification technique employed in this study is Fuzzy K-Nearest Neighbor in Every Class (FKNNC) which is a method with the ability to consider the ambiguous nature of the data and give strength to the decision of a class because it has a degree of membership value. In this study using data divided into training and testing data. Obtained the best accuracy value is 90.47%, this accuracy value shows good performance in classifying potential tsunami disaster data. Thus, this research helps identify and provide information on potential tsunami disasters based on historical earthquakes that have occurred.

Keywords: Classification · FKNNC · Tsunami

1 Introduction

Tsunamis are waves in the ocean that result from abrupt disruptions in the sea, triggered by modifications in the seabed's structure. Changes in the shape of the seabed can be caused by tectonic earthquakes, volcanic eruptions or landslides on the seabed. Tectonic earthquakes are the main cause of tsunamis worldwide [1]. Tectonic earthquakes are the most common earth-shaking events, vibrations resulting from the fracture of rocks due to the slow collision of two plates where the accumulated energy from the collision exceeds the strength of the rocks. The resulting energy is radiated in all directions in the form of waves so that the effect is felt to the earth's surface [2].

Tsunamis that occur around the world no one knows when they will occur, but to predict by classification techniques the potential for tsunamis to occur can be done through the application of science and technology. The data contained in historical earthquakes has information that can be analyzed and processed so as to determine the

Y. Yuyun et al. (eds.), *Proceedings of the 5th International Seminar on Science and Technology (ISST 2023)*, Advances in Physics Research 10,
https://doi.org/10.2991/978-94-6463-520-1_11

potential for a tsunami to occur or not after an earthquake. Analysis and processing of historical earthquake data can be done with the data mining process [3]. One of the data mining techniques, namely classification, can be used to determine whether an earthquake has the potential to cause a tsunami or not. One method commonly used in classification techniques is K-Nearest Neighbor (KNN). The KNN method will classify new data that has no known class [4].

With KNN, earthquake data is classified strictly or fully in one of the prediction classes, namely tsunami potential and no tsunami potential, so that the results do not have membership values or probability values in each prediction class [5]. To improve the weakness of K-NN, it can be combined with fuzzy logic so that it becomes the Fuzzy K-Nearest Neighbor in Every Class (FKNNC) method. The FKNNC method pays attention to the value of each class label, allowing for better classification results [6]. Therefore, in this study, the FKNNC method will be applied to classify potential tsunami natural disasters using the parameters of earthquake depth, earthquake magnitude, and earthquake intensity.

2 Materials

2.1 Tsunami

Tsunamis are large waves from the ocean that occur due to impulsive disturbances in the ocean that hit coastal areas. The impulsive disturbance occurs because of a sudden change in the shape of the seabed. Tsunamis are invisible when they are still far away in the middle of the ocean, but once they reach shallow areas they take the form of waves that move faster and get bigger. The wave moves at high speed and then increases in height until it reaches the coastal area [7].

2.2 Fuzzy K-Nearest Neighbor in Every Class

Fuzzy K-Nearest Neighbor in Every Class (FKNNC) is an extension method of FKNN that pays attention to the probability value in each class [8]. The FKNNC method can provide greater flexibility, enabling the model to adjust to varying patterns for each category. The FKNNC method utilizes a set number of K nearest neighbors within each class of test data, rather than using K nearest neighbors as done in KNN and FKNN.

3 Methods

In this research, the data mining process model implemented is FKNNC. The research methodology is carried out through the following steps.

3.1 Data Collection

The data to be used in this study are data on earthquake depth, earthquake magnitude, earthquake intensity, and tsunami occurrence from 2016 to 2022 around the world sourced from the National Geophysical Data Center (NGDC) website [9].

3.2 Analysis of Fuzzy K-Nearest Neighbor in Every Class

The FKNNC framework is built upon FKNN, where test data have a membership value in each class within the interval 0 to 1. The following is the algorithm of the FKNNC method [10].

1. Input data to be classified according to attributes and class values and divide the data into two, namely train and test data. Train data is used for the learning process in classification, and test data is used for prediction and classifying class results on data.
2. Normalize the data as in the following formula.
$$V' = \frac{v(x) - \min(x)}{Range(x)} \quad (1)$$
where, V' : normalization result whose value ranges between 0 and 1, *V(x)* : the value of the attribute to be normalized, *Max(x)* : maximum of an attribute, *Min(x)* : minimum of an attribute and *Range(x)*: the value of maximum x minus minimum x.
3. KNN calculation process, to find the value of KNN using Euclidean Distance as in the following equation.
$$d_i = \sqrt{\sum_{i=1}^{p} (x_{2i} - x_{1i})^2} \quad (2)$$
where, d_i : proximity distance, p : number of attributes, x_1 : training data, and x_2 : test data
4. Calculate S as the sum of all distances from C x K neighbors of each class, using the following equation.
$$S_{ij} = \sum_{r=1}^{K} d(x_i, x_r) \quad (3)$$
where S_{ij} is the accumulated neighbor distance K from neighbor C x K.
5. Calculate D as the accumulation of all distances from neighbors using the following formula.
$$D = \sum_{j=1}^{C} (S_{ij})^{\frac{-2}{m-1}} \quad (4)$$
where D is the sum of all K distances from neighbor C x K. In this context, m represents the weight exponent, the value of m is greater than 1.
6. Calculate U as the data membership value in each class, using the following formula.
$$U_{ij} = \frac{S_{ij}}{D_{ij}} \quad (5)$$
U_{ij} is the data membership value. Choose the max membership value using the following equation.
$$y' = \max(U_{ij}) \quad (6)$$
assign the class with the max membership value result to be the prediction class for the test data.
7. Evaluate the accuracy of classification results through accuracy using the following equation.
$$Accuracy = \frac{TP+TN}{Total} x\ 100 \quad (7)$$

4 Results and Discussion

Table 1. Classification Accuracy Results

K	Sensitivity	Specitivity	Accuracy
2	0,81	0,96	90,47
3	0,80	0,92	88,09
4	0,80	0,92	88,09
5	0,80	0,92	88,09

Based on Table 1, it is observed that the classification results using the FKNNC method, with the number of K ranging from 2 to 5, the highest accuracy value is 90.47%, with a sensitivity value of 0.81 and a specificity value of 0.961. Meanwhile, the lowest accuracy value is 88.09%, with a sensitivity value of 0.80 and a specificity value of 0.925.

5 Conclusion

Based on the analysis conducted using the FKNNC method, a high accuracy rate of 90.47% was achieved. With this good accuracy value, all earthquake parameters used, namely earthquake depth, magnitude, and earthquake intensity, can provide information in the classification process of determining earthquakes that will have the potential for a tsunami disaster so that they can be used as a reference for government agencies to general information for the community.

Acknowledgements. We would like to express our gratitude to the Faculty of Mathematics and Natural Sciences at Tadulako University for providing the research funding.

References

1. Mustofa, N. A., Gempa Bumi, Tsunami dan Mitigasinya, Balai Informasi dan Konservasi Kebumian Karangsabung LIPI Kebumen, 7(1), pp. 66-73 (2010).
2. Triana, W. L., Penetuan Zonasi Resiko Tsunami di Kabupaten Banyuwangi, Institut Teknologi Nasional Malang, Malang (2017).
3. Dito, P. U., and Bister, P., Penerapan Data Mining pada Data Gempa Bumi Terhadap Potensi Tsunami, Prosiding Seminar Nasional Riset Information Sciences (SENARIS), ISSN 2686-0260, pp. 846-853 (2019).
4. Resa, A. Y., Kusrini, and Hanif, A. F., Analisis Pengaruh Tingkat Akurasi Klasifikasi Citra Wayang dengan Algoritma CNN, Jurnal Teknologi Informasi, 4(2), pp. 182-190 (2020).
5. Hendrik, S., and Ratih, F., Implementasi Metode Fuzzy K-Nearest Neighbor Dalam Memprediksi Tingkat Kelulusan Mahasiswa, Jurnal Duniailmu.org, 1(1), pp. 1-10 (2021).
6. Prasetyo, E., Fuzzy K-Nearest Neighbor in Every Class Untuk Klasifikasi Data, Seminar Nasional Teknik Informatika (SANTIKA 2012), pp 57-60 (2012).
7. Dian, P., and Arniza, F., Analisis Potensial Penjalaran Gelombang Tsunami di Pesisir Barat Lampung Indonesia, Jurnal Teknik Sipil ITP, 8(1), pp. 29-37 (2021).
8. Izzah, R. U., Wahyuningsih, S., and Amijaya, F. D. T., Classificationn of Nutritional Status of Todlers Using Fuzzy Knearest Neighbor in Every Class (FK-NNC), Journal of Physics Conf. Ser. 1277 012050, pp. 1-11 (2019).
9. National Geophysical Data Center, Earthquake Events, https://www.ngdc.noaa.gov, accesed on October (2023).
10. Keller, J. M., Gray, M. R., and Givens, J. A., Fuzzy K-Nearest Neighbor Algorithm, IEEE Transcition on System, Man, and Cybernetics, 15(4), pp. 580-585 (1985).