Coastal Vulnerability Classification of the North Coast of Java using K-Nearest Neighbor

Ayu Novitasari Saputri*, Arita Witanti
Department of Informatics, Faculty of Information Technology, University of Mercu Buana Yogyakarta, Indonesia

Abstract

Indonesia has many beaches that attract the attention of tourists, including the north coast of Java. In addition to visual appeal, there are also other potentials such as settlements, agriculture, fisheries, ports, ponds, and other resources. However, there is also a threat to coastal damage caused by, among other things, wave action, tides, abrasion, and tidal flooding. For this reason, in developing coastal areas on the north coast of Java, it is necessary to consider the potential for damage to the coast based on the physical condition of the coast and a system that can classify the vulnerability levels of coastal areas is also needed. The Coastal Vulnerability Index (CVI) can be determined and classified using for example the Gornitz formula, or using data driven model and machine learning based on the coastal parameter data. This study demonstrates how the K-Nearest Neighbor, also known as K-NN algorithm, can be used to classify the level of vulnerability of the coastal areas. This study uses 290 points (locations) along the northern coasts of Java. The parameters that determine the coastal vulnerability are mean sea level (MSL), mean significant wave height (MSWH), mean tidal range (MTR), shoreline changes, landforms and slopes. In this study, the classification of coastal vulnerability levels is classified into 4, namely "low, moderate, high and very high". The K-NN system uses 80% of the data for training and 20% for testing, with the value of K = 1 to 10. The test results show that the K-NN method is capable of classifying the vulnerability levels of the North Coast of Java. From the test results for values of K = 1 to K = 10, and by randomizing the training data and test data gives an average accuracy rate of 86.21% to 97.13%, with the best K value obtained at K = 2.

INTRODUCTION

Indonesia is an archipelago located between the Asian continents and the Australian continent, the Indian Ocean and the Pacific Ocean with coordinates of 6°N - 11°S and 95°E - 141°W [1]. Indonesia has a coastline of more than 95,181 kilometers, making it the world's second longest behind Canada [2][3]. Beaches in Indonesia, especially the north coast of Java, have a lot of attractions, especially from a visual perspective. In addition, it has many other potentials such as residential areas, tourism, agriculture, fisheries, ports, ponds, and other resources.

However, behind this potential, the threat to coastal damage is getting bigger from time to time, for example, on the northern coasts of Java. The threat is in the form of coastal damage caused by wave action, tides, abrasion, and tidal flooding [4][5]. The threat of coastal damage must be taken into considerations for sustainable development of coastal areas. The potential for coastal damage can be classified based on the physical condition of the beach which is

described in physical coastal data such as ground level, sea level rise, significant wave height, coastal geomorphology, and so on [6].

The definition of vulnerability in general is the level of a system that leads to a decrease in resilience as a result of external influences. Coastal vulnerability describes the condition of a coastal area which can be subject to technical damage such as abrasion or accretion, which if it occurs repeatedly will result in a decrease in the land area along the coast [7]. The level of coastal vulnerability can be classified into 4 categories, namely low, moderate, high and very high. This classification is based on the coastal vulnerability index (CVI), which can be obtained using the Gornitz method (explained later in the article), using the previously mentioned parameters: mean sea level (MSL), mean significant wave height (MSWH), mean tidal range (MTR), shoreline changes, landforms and coastal slopes. However, this study tries to demonstrate whether the K-nearest neighbor (K-NN) algorithm can perform to classify the vulnerability level, in this case, the northern coasts of Java.

The rapid development of computing technology facilitates the utilization of machine learning in all areas including the use of K-NN. Data mining discovers new patterns in very large datasets, including techniques from statistic, machine learning, database systems, and artificial intelligence [14][15], which are used to find clear relationships and produce conclusions that were not previously known or understood, but useful for data owners [10]. In general, data mining is used for two purposes, namely descriptive and predictive. Several groups of methods that are often used in data mining include [11]: prediction, classification, estimation, clustering and association. Meanwhile, the Knowledge Discovery in Databases (KDD) includes several stages, from the database, data warehouse, the mining process, patterns and knowledge [8].

This study uses a classification process that finds models by describing and distinguishing data classes, with the goal of allowing object classes to predict previously unknown classes [11]. This classification process has several main components, namely: Classes as dependent variables that function as labels/targets from the classification results, Predictors as characteristics of data attributes that are classified as independent variables in the model, Training datasets as the complete dataset containing the classes and predictions to be trained, so that the model can be grouped into the correct or appropriate class, and the last is the Test data sets, that contain a set of new data against which the model is classified and whose purpose is to determine the accuracy of the established model [10].

The use of K-NN has been proven in many cases, including conducting K-NN research to classify consumer interest in an insurance company [12], determining the level of forest fire hazard using 252 data, with an accuracy rate of 80.16% [13]. In the world of education, [14] conducted research aimed at assisting schools in determining which students will be accepted based on zoning and non-zoning based on the regulations of the Minister of Education and Culture. This research was successful with an accuracy rate of 83.36% at K = 5. Other examples of research using K-NN have also been carried out, which aims to classify the level of sales of Lombok Vape On per month [15], regarding the selection of student thesis topics [16] and the news classification based on topics such as health, sports, technology, and so on [17].

The KNN algorithm is widely used in data mining for classification, or estimation and prediction [13]. The K-NN algorithm is a supervised learning technique that labels data and groups it by label, storing all training data, and delaying the formation of a classification model until the test data is provided predictions. This method classifies data based on their similarity or proximity to the majority of a test data's neighbors. Data points that are close to each other are called "neighbors" [18]. The number of the nearest neighbor data is determined

and expressed in K and the best K value can be determined by the parameter optimization [19]. The lower the number of K chosen, the more data there is. However, the bigger the K number used, the larger the data dimension [20]. There are a number of techniques for computing neighbor distance, which is an important part of K-NN algorithms, such as Manhattan or city block distance and Minkowski distance, and the most widely used is the Euclidean distance, which is a technique of comparing the differences of two data [21, 22, 23, 24, 25, 26].

To summarize, this study aims to develop a system that can classify the vulnerability levels of coastal areas on the north coast of Java by implementing the K-Nearest Neighbor (K-NN) algorithm, as well as obtaining the level of accuracy and reliability of the system. This research is expected to assist users or policy-making authorities in classifying the vulnerability levels of coastal areas in general if the physical condition parameters of the coast are known, so as to assist the authorities in planning for coastal area development.

MATERIALS AND METHOD

The materials used in this study were coastal physical parameter data taken at 290 points (locations) along the north coast of Java. These points line up from the Serang, Banten to Sidoarjo, East Java, as shown in Figure 1.

The equipment used in this study included a computer system with Windows 10 Pro OS with Intel Core i7 gen 10 CPU @1.8GHz, 16GB RAM, 500GB HDD with a 14" display. The software used is MS Excel 2021, Matlab v14 and Global Mapper v13.

The physical parameters of the coast that are considered important as a determinant of coastal vulnerability are not always the same. This usually depends on the needs and location of the study. In the case of the north coast of Java, where a lot of activities occur, erosion of the coastlines must be considered. The process of damage to the northern coast of Java is attributed to the wave height, sea level rise, tidal range, and coastal slope [6]. Several physical parameters of the coast are used to estimate coastal vulnerability. The physical data of the coast used in this study are tabulated in Table 1 [27, 28, 29, 30, 31, 32].
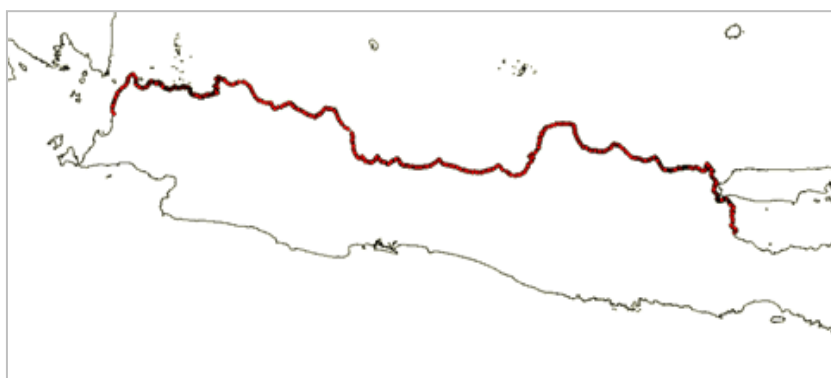
Figure 1. The distribution of coastal parameter data along north coast of Java

Table 1. Physical Parameters Used for Coastal Vulnerability Assessment [32]

No	Parameter	Source
1	Mean Sea Level (m)	Copernicus, 1993 - 2021
2	Mean Significant Wave Height (m)	Copernicus, 2020 – 2021
3	Mean Tide Range (m)	Kementerian Kelautan dan Perikanan (KKP), 2009
4	Coastline Change (m/yr)	Landsat data – USGS, 1990 - 2020
5	Landforms	Global SRTM Landforms – USGS, 2011
6	Slope (deg)	DEM SRTM – USGS, 2018

The same parameters have also been used by the USGS in estimating the level of vulnerability of beaches in the United States to the sea level rise. Meanwhile, [33] also used 6 parameters that were almost the same in determining the vulnerability index of beaches in Bangladesh, namely the average tide range, height of the significant wave, change in shoreline, cyclone track density, slope of the coastal topography, and the rise in sea level.

After the physical parameter data are collected, the next step is to transform the data, where the physical values of the coastal parameters are classified into several classes of vulnerability, namely: low, moderate, high and very high. The transformation was carried out according to [34] with a few scale adjustments, as shown in Table 2.

The level of coastal vulnerability is then expressed by Coastal Vulnerability Index or CVI, which is determined using the Gornitz method, where each parameter is classified into several classes, namely low, moderate, high, and very high, on a scale of 1 to 4. The CVI value is obtained by calculating the root of the mean multiplication of the class value of all of these parameters [33], according to (1) [27, 28, 29, 30, 31]. Finally, the CVI category is obtained by transforming the calculated result according to Table 2.

$$CVI = \sqrt{((a * b * c * d * e * f)/6}$$
(1)

where:
a = mean sea level (MSL) d = shoreline change (erosion/accretion rate)
b = mean wave height (MSWH) e = landform
c = mean tide range (MTR) f = slope

The next stage is to classify the vulnerability level using the K-NN algorithm, after examining that the data contain all levels of CVI categories, with reasonable number of occurrences for each category [26, 35, 36, 37]. In this case, the 290 data were used, each of which consists of the 6 coastal parameters and its CVI value. The data were divided into 80% for training and 20% for testing, and using values of K=1 to K=10. Figure 2 shows the steps performed in this study.

In this study, the K-NN uses Euclidean distance to calculate the distance between each parameter of the test data and the corresponding parameter of the training data [38], and determined using the following (2).

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$
(2)

where x_i and y_i represents the values of training and testing data respectively, i represents the data index, and d(x,y) represents the distance between training and test data, and represents the number of data. Meanwhile, the level of accuracy of the classification results is calculated using the following (3) [39].

Table 2. Vulnerability Classification Based on Physical Coast Parameters [34]

Parameter	Index of Coastal Vulnerability			
	"Low"	"Medium"	"High"	"Very High"
	1	2	3	4
Mean Sea Level (m/yr)	< 0,53	0,53 – 0,68	0,68 – 0,82	> 0,82
Mean Height of Significant Wave (m)	< 0,29	0,29 – 0,52	0,52 – 0,63	> 0,63
Mean Tide Range (m)	< 0,40	0,40 – 0,80	0,80 – 1,20	> 1,20
Shoreline Change (m/yr)	> 79,99	20,36 – 79,99	13.03 – 20,36	< (-13,03)
Landforms	11 - 15	21 - 24	31 - 32	33 - 42
Slope (°)	> 12,87	6,11 – 12,87	2,32 – 6,11	0,00 – 2,32
CVI	1 - 5	5 - 10	10 - 15	> 15
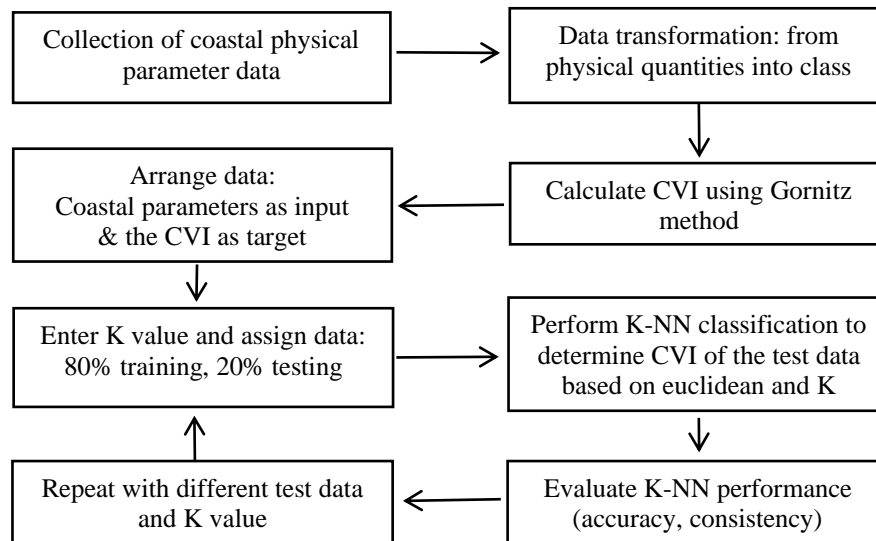
Figure 2. Steps of research

In this case, the accuracy means how the CVI class of the testing data obtained from the Gornitz method is replicated by the K-NN.

$$Accuracy = \frac{Number\ of\ Correct\ Result}{Number\ of\ Test\ Data}\ x\ 100\%$$

(3)

The K-NN performance consistency is also tested several times, each time by changing the data sequence, in other words, by shuffling or randomizing the order of the data. Therefore, each round of training-and-testing will use different sets of data points, eventhough maintaining the ratio of 80:20.

In this study, the system was designed using UML (Unified Modeling Language), which includes various elements such as diagrams, scenarios, and the GUI Design [40][41].

RESULTS AND DISCUSSION

From the 290 data along the north coast of Java, the CVI is calculated using the Gornitz formula, which results in the following CVI class distribution: 91 low, 105 moderate, 67 high and 27 very high, as shown in Figure 3.
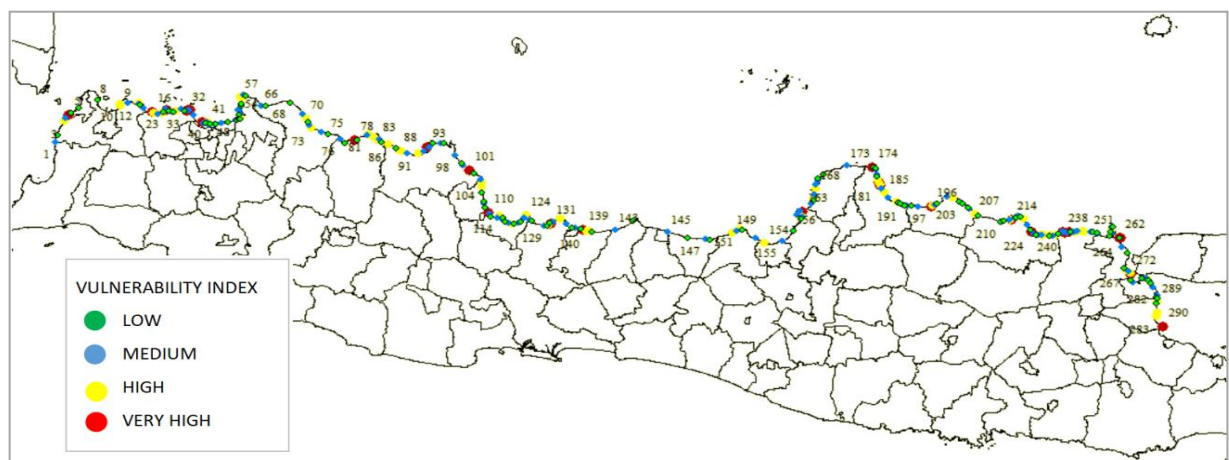
Figure 3. The distribution of CVI data for the K-NN Classification System

In this study, we used a K-NN algorithm with the values of K ranging from 1 to 10 on 290 data splits into 80% (232) for training and 20% (58) for testing. The attributes used are MSL, MSWH, MTR, Shoreline Change (SC), Landforms, and Slope.

Each data is assigned an identity (ID) from 001 to 290, sequentially from west to east. Data with ID 001 is located in Serang Regency, Banten, while data with ID 290 is in Sidoarjo Regency, East Java. Therefore, for the initial case, the training data consist of those data from Serang to a location in Tuban (001-232), while the testing data are those located from Tuban to Sidoarjo (233-290). Table 3 and Table 4 show the distribution of the test data and training data in original forms (in physical quantities). Meanwhile, Table 5 and Table 6 shows the transformed training data and testing data (in categories). As mentioned earlier, the CVI category is obtained by transforming the result from the Gornitz formula (1), which is calculated using the transformed values of the parameters (Table 2).

For the purposes of this study, a K-NN classifier program with a GUI was written using MATLAB and implemented on prepared data. Figure 4 shows the display after successful data entry and Figure 5 shows the resulting display of K-NN classification with the value K=1. When the RUN button is pressed, the data that have been entered will be used in sequence as the initial ordering.

Table 3. Training Data with Parameter (Attribute) Values

ID	Lon	Lat	MSL	MSWH	MTR	SC	Landforms	Slope	Region
1	105.835	-6.244	0.54414505	0.64	0	-1.3622259	24	0.6106694	Serang
2	105.846	-6.198	0.52522355	0.15296684	0	30.537035	24	0.5811676	Serang
3	105.887	-6.095	0.54414505	0.64	1,5	-0.45047653	24	1.8776451	Serang
					s/d				
232	112.146	-6.901	0.53470594	0.49819785	0	-4.46	24	1.1775075	Tuban

Table 4. Test Data with Parameter (Attribute) Values

ID	Lon	Lat	MSL	MSWH	MTR	SC	Landforms	Slope	Region
233	112.154	-6.898	0.53470594	0.49819785	1.5	-7.29	34	0.24059798	Tuban
234	112.167	-6.9	0.5780206	0.7224264	0	-3.9506848	34	0.119982705	Tuban
235	112.187	-6.893	0.5734428	0.9828424	1.5	22.79613	34	0.28499106	Lamongan
					s/d				
290	112.868	-7.542	0.57222205	0.64213836	2	1.5711467	34	0.11753825	Sidoarjo

Table 5. Classification of Training Data, With CVI Values According to Gornitz

ID	MSL	MSWH	MTR	SC	Landforms	Slope	CVI	Region
1	2	4	1	3	2	4	moderate	Serang
2	1	1	1	2	2	4	low	Serang
3	2	4	4	3	2	4	high	Serang
10	2	3	4	3	4	4	high	Serang
				s/d				
232	2	2	1	3	2	4	low	Tuban

Table 6. Classification of Test Data, With CVI Values According to Gornitz

ID	MSL	MSWH	MTR	SC	Landforms	Slope	CVI	Region
233	2	2	4	3	4	4	high	Tuban
234	2	4	1	3	4	4	moderate	Tuban
235	2	3	4	2	4	4	high	Lamongan
				s/d				
290	2	4	4	3	4	4	very high	Sidoarjo

The training data (232 data) used are data from ID001 to ID232, and the test data (58 data) used are data from ID233 to ID290. Meanwhile, to test system reliability and the validity of the classification results, the user can randomize the data that has been input by selecting the RANDOMIZE button and then selecting the RUN button. In this case, 232 training and 58 test data used are no longer in the order they were originally entered.

Accuracy Result of K-NN classification

The Euclidean distance of each test data parameter to the training data is used to calculate K-NN in this study. After the eEuclidean distance is obtained, the distance is sorted in ascending order. The vulnerability index CVI is then calculated using the K value.

Using sequential data, the results of K-NN classification are shown in Table 7. In all cases, 80 percent (232 data) and 20 percent (58 data) were used for training and testing respectively. By using values of K=1 to K=10, an average accuracy of 87.07%, and the highest accuracy of 91.38% were obtained at K=3 and K=4.
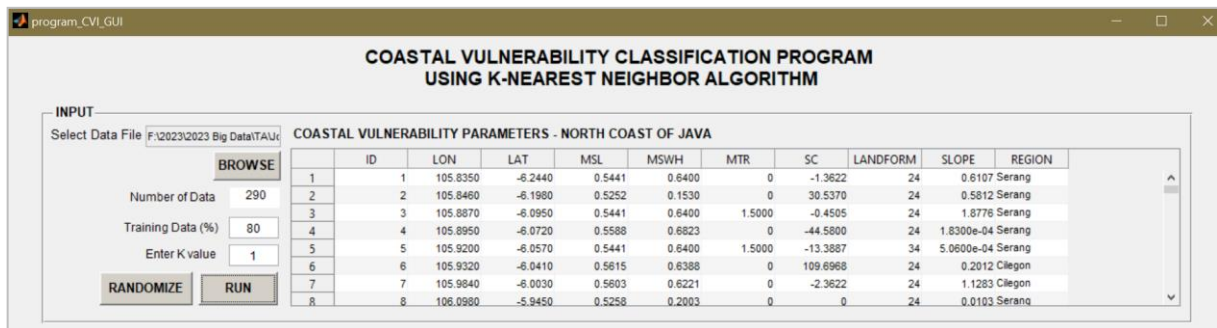
Figure 4. Input Data on The GUI Program
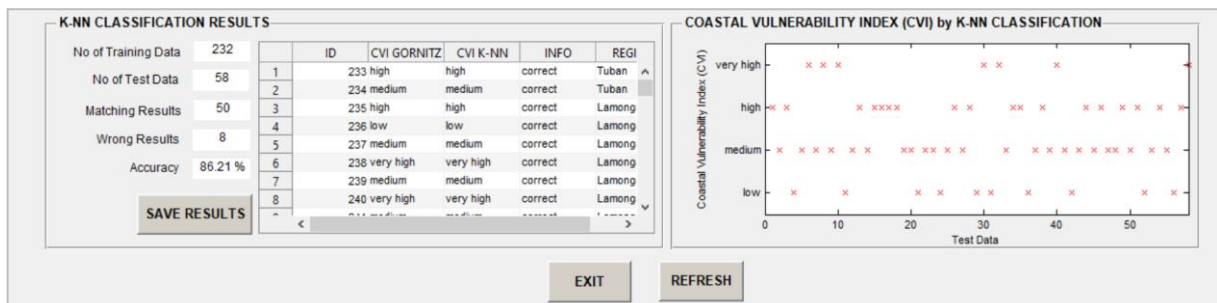
Figure 5. K-NN Classification Results on The GUI Program

Table 7. The accuracy of K-NN testing using the initial data sequence

K value	Number of data	No of correct prediction	Accuracy (%)
1		50	86.21
2		50	86.21
3		53	91.38
4		53	91.38
5	58	52	89.66
6		52	89.66
7		51	87.93
8		51	87.93
9		46	79.31
10		47	81.03
		Average accuracy	**87.07**

In this case, the classification results of the K-NN algorithms showed that 5 out of 58 test data do not match Gornitz's vulnerability data (CVI), namely data ID 248 (Lamongan), 266 (Gresik), 271 (Surabaya), 279 (Surabaya), and 281 (Surabaya). Figure 6 shows an illustration of the results of this test.

The consistency of the K-NN performance has also been tested by running the system for K=1 to K=10, by randomizing the data sequence 3 times for each value of K. Table 8 shows the accuracy of the KNN classification with random data.

Based on Table 8, for the test using random data from K=1 to K=10, each of which was carried out 3 times, resulting in an average accuracy rate of 91.84%. The highest average accuracy rate of 97.13% is obtained at the value of K = 2. This shows that the K-NN system performs consistency for different sets of test data. Even though CVI value can be determined directly using the Gornitz method, this study demonstrates that machine learning algorithms such as K-NN can be used to classify the coastal vulnerability levels.

ID	LOKASI	CVI-Gornitz	CVI KNN	KET
233	Tuban	tinggi	tinggi	sesuai
234	Tuban	sedang	sedang	sesuai
235	Lamongan	tinggi	tinggi	sesuai
236	Lamongan	rendah	rendah	sesuai
237	Lamongan	sedang	sedang	sesuai
238	Lamongan	sgt tinggi	sgt tinggi	sesuai
239	Lamongan	sedang	sedang	sesuai
240	Lamongan	sgt tinggi	sgt tinggi	sesuai
241	Lamongan	sedang	sedang	sesuai
242	Lamongan	sgt tinggi	sgt tinggi	sesuai
243	Lamongan	rendah	rendah	sesuai
244	Lamongan	sedang	sedang	sesuai
245	Lamongan	tinggi	tinggi	sesuai
246	Lamongan	sedang	sedang	sesuai
247	Lamongan	tinggi	tinggi	sesuai
248	Lamongan	tinggi	sedang	BERBEDA
249	Lamongan	tinggi	tinggi	sesuai
250	Lamongan	tinggi	tinggi	sesuai
251	Lamongan	sedang	sedang	sesuai
252	Lamongan	rendah	rendah	sesuai
253	Gresik	rendah	rendah	sesuai
254	Gresik	sedang	sedang	sesuai
255	Gresik	rendah	rendah	sesuai
256	Gresik	rendah	rendah	sesuai
257	Gresik	sedang	sedang	sesuai
258	Gresik	tinggi	tinggi	sesuai
259	Gresik	sedang	sedang	sesuai
260	Gresik	tinggi	tinggi	sesuai
261	Gresik	rendah	rendah	sesuai
262	Gresik	sgt tinggi	sgt tinggi	sesuai
263	Gresik	rendah	rendah	sesuai
264	Gresik	sgt tinggi	sgt tinggi	sesuai
265	Gresik	sedang	sedang	sesuai
266	Gresik	tinggi	rendah	BERBEDA
267	Gresik	tinggi	tinggi	sesuai
268	Gresik	rendah	rendah	sesuai
269	Gresik	sedang	sedang	sesuai
270	Gresik	tinggi	tinggi	sesuai
271	Surabaya	sedang	rendah	BERBEDA
272	Gresik	sgt tinggi	sgt tinggi	sesuai
273	Surabaya	sedang	sedang	sesuai
274	Surabaya	rendah	rendah	sesuai
275	Surabaya	sedang	sedang	sesuai
276	Surabaya	tinggi	tinggi	sesuai
277	Surabaya	sedang	sedang	sesuai
278	Surabaya	tinggi	tinggi	sesuai
279	Surabaya	sedang	rendah	BERBEDA
280	Surabaya	sedang	sedang	sesuai
281	Surabaya	tinggi	rendah	BERBEDA
282	Surabaya	sedang	sedang	sesuai
283	Sidoarjo	tinggi	tinggi	sesuai
284	Surabaya	rendah	rendah	sesuai
285	Sidoarjo	rendah	rendah	sesuai
286	Sidoarjo	tinggi	tinggi	sesuai
287	Surabaya	sedang	sedang	sesuai
288	Sidoarjo	rendah	rendah	sesuai
289	Sidoarjo	tinggi	tinggi	sesuai
290	Sidoarjo	sgt tinggi	sgt tinggi	sesuai

(a) classification table

(b) Gornitz CVI

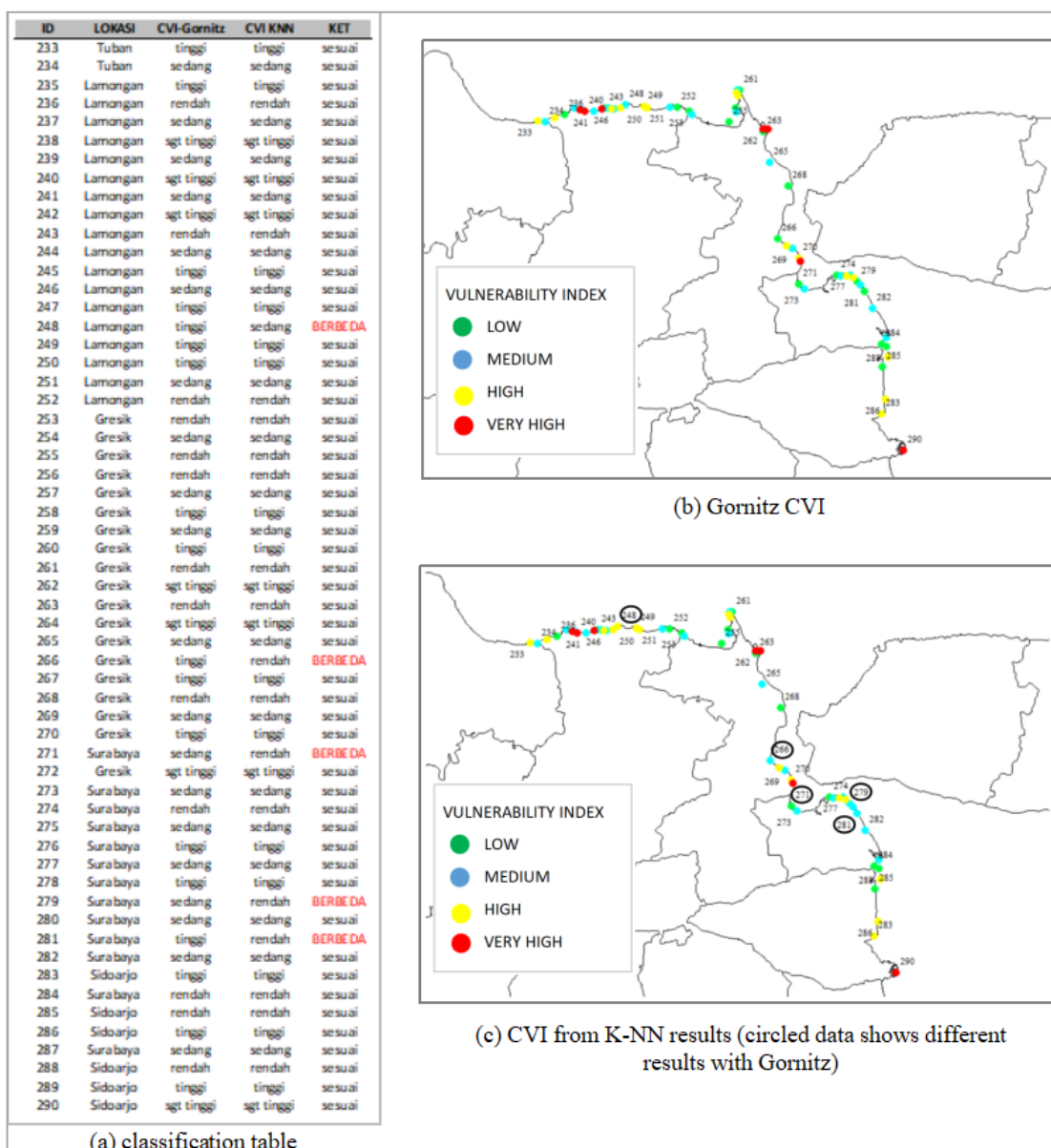(c) CVI from K-NN results (circled data shows different results with Gornitz)

Figure 6. The comparison of the CVI from K-NN vs CVI Gornitz

Table 8. The accuracy of the K-NN testing with random data sequences, for values K=1 to 10

K value	No of test data	Test no	No of correct prediction	Accuracy (%)
1	58	1	57	98.28
		2	56	96.55
		3	55	94.83
		Average accuracy for K = 1		**96.55**
2	58	1	56	96.55
		2	57	98.28
		3	56	96.55
		Average accuracy for K = 2		**97.13**
3	58	1	54	93.10
		2	49	84.48
		3	56	96.55
		Average accuracy for K = 3		**91.38**
4	58	1	55	94.83
		2	55	94.83
		3	54	93.10
		Average accuracy for K = 4		**94.25**
5	58	1	50	86.21
		2	56	96.55
		3	54	93.10
		Average accuracy for K = 5		**91.95**
6	58	1	50	86.21
		2	56	96.55
		3	55	94.83
		Average accuracy for K = 6		**92.53**
7	58	1	54	93.10
		2	53	91.38
		3	50	86.21
		Average accuracy for K = 7		**90.23**
8	58	1	52	89.66
		2	53	91.38
		3	48	82.76
		Average accuracy for K = 8		**87.93**
9	58	1	48	82.76
		2	54	93.10
		3	48	82.76
		Average accuracy for K = 9		**86.21**
10	58	1	55	94.83
		2	47	81.03
		3	53	91.38
		Average accuracy for K = 10		**89.08**
		Average accuracy		**91.84**

CONCLUSIONS

The K-NN method has been applied in the vulnerability level classification in the North Coast of Java, using 290 data, dividing 80% and 20% of the data for training and testing respectively. From the 58-testing data (cases), this system is able to predict the low CVI at 17 locations, moderate at 19 locations, high at 15 locations and very high at 7 locations. These results agree with the Gornitz vulnerability index (CVI), except for one location in Lamongan, one in Gresik and 3 in Surabaya. The average accuracy of classification test using K-NN with sequential data was 87.07%, with the highest accuracy of 91.38% for K=3 and K=4. In testing

with random data, the average accuracy of the K-NN classification test is 91.84%, with the highest accuracy of 97.13% at K=2. To improve the accuracy, further studies may consider reducing the study area, such as applying the same parameters to a coastal cell about 100-200 km long.

ACKNOWLEDGMENT

REFERENCES

[1] NN, "Geografi." *Embassy of The Republic of Indonesia in Astana, Kazakhstan,* Ministry of Foreign Affairs of The Republic of Indonesia, 2019. [Online]. Available: https://kemlu.go.id/nur-sultan/id/pages/geografi/41/etc-menu

[2] NN, "Laut Masa Depan Bangsa, Mari Jaga Bersama," *Ministry of Maritime Affairs and Fisheries of The Republic of Indonesia*, 2019. [Online]. Available: https://kkp.go.id/artikel/12993-laut-masa-depan-bangsa-mari-jaga-bersama

[3] A. Syetiawan, D. Arini, R. S. Dewi, H. Gularso, and W. Nugraha, "Potential of Lidar Bathymetry Technology for Coastline Large-Scale Mapping Acceleration in Indonesia".

[4] J. Kalther and A. Itaya, "Coastline changes and their effects on land use and cover in Subang, Indonesia," *Journal of Coastal Conservation*, vol. 24, no. 2, p. 16, Apr. 2020, doi: 10.1007/s11852-020-00736-w.

[5] S. A. Harahap, N. P. Purba, and M. L. Syamsuddin, "Trend of Coastline Change for Twenty Years (1994-2014) in Cirebon, Indonesia," *World Scientific News*, vol. 138, no. 2, pp. 79-92, 2019.

[6] W. A. Gunawan and D. N. Handiani, "Kerentanan Pesisir Di Pantai Utara Jawa Barat," *FTSP Series: Seminar Nasional dan Diseminasi Tugas Akhir*, Institut Teknologi Nasional Bandung (ITENAS), 2021.

[7] R. R. Husaini and N. E. Darfia, "Analisis Kerentanan Pantai Pulau Rupat Provinsi Riau Berdasarkan Metode Indeks Kerentanan Pantai," *Jurnal Teknik Sipil Institut Teknologi Padang (JTSITP)*, vol. 8, no. 1, pp. 29-37, Jan. 2021, doi: 10.21063/jts.2021.V801.06.

[8] J. Tripathy, R. Dash, and B. K. Pattanayak, "A Case Study on Data Classification Approach Using K-Nearest Neighbor," in *2021 International Conference in Advances in Power, Signal, and Information Technology (APSIT)*, Bhubaneswar, India, Oct. 2021, pp. 1–6. doi: 10.1109/APSIT52773.2021.9641209.

[9] Suyanto, *Data Mining Untuk Klasifikasi dan Klasterisasi Data (Edisi Revisi)*. Informatika Bandung, 2019.

[10] R. T. Handayanto and Herlawati, *Data Mining dan Machine Learning Menggunakan Matlab dan Python*. Informatika Bandung, 2020.

[11] S. Bahri, A. Wibowo, R. Wajhillah, and S. Suhada, *Data Mining Algoritma Klasifikasi & Penerapannya Dalam Aplikasi*, Pertama. Graha Ilmu, 2019.

[12] W. O. N. Kadir, B. Pramono, and Statiswaty, "Penerapan Data Mining Dengan Metode K-Nearest Neighbor (KNN) Untuk Mengelompokan Minat Konsumen Asuransi (PT. Jasaraharja Putera)," *semanTIK*, vol. 5, pp. 97–104, Jun. 2019, doi: 10.5281/zenodo.3116132.

[13] M. R. Noviansyah, T. Rismawan, and D. M. Midyanti, "Penerapan Data Mining Menggunakan Metode K-Nearest Neighbor Untuk Klasifikasi Indeks Cuaca Kebakaran Berdasarkan Data AWS (Automatic Weather Station) (Studi Kasus: Kabupaten Kubu Raya)," *Jurnal Coding Sistem Komputer Untan*, vol. 06, no. 2, pp. 48–56, 2018, doi: 10.26418/coding.v6i2.26672

[14] D. Kurniawan and A. Saputra, "Penerapan K-Nearest Neighbour Dalam Penerimaan Peserta Didik Dengan Sistem Zonasi," *Jurnal Sistem Informasi Bisnis*, vol. 9, no. 2, pp. 212-219, Nov. 2019, doi: 10.21456/vol9iss2pp212-219.

[15] Yahya and W. P. Hidayanti, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada 'Lombok Vape On,'" *Jurnal Informatika dan Teknologi*, vol. 3, no. 2, pp. 104–114, Jul. 2020, doi: 10.29408/jit.v3i2.2279

[16] A. E. Gumanti, Taslim, S. Handayani, and D. Toresa, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Topik Skripsi Mahasiswa di Fakultas Ilmu Komputer," *JITaCS: Journal of Information Technology and Computer Science*, vol. 1, no. 2, pp. 44–50, Jun. 2022, doi: 10.31849/jitacs.v1i2.10037

[17] A. Y. Muniar, Pasnur, and K. R. Lestari, "Penerapan Algoritma K-Nearest Neighbor Pada Pengklasifikasian Dokumen Berita Online," *Jurnal Teknologi Informasi dan Komunikasi AKBA*, vol. 10, no. 2, pp. 137–144, Desember 2020, doi: 10.35585/inspir.v10i2.2570

[18] L. Afifah, "Algoritma K-Nearest Neighbor (KNN) untuk Klasifikasi." [Online]. Available: https://ilmudatapy.com/algoritma-k-nearest-neighbor-knn-untuk-klasifikasi/

[19] E. S. Susanto, K. Kusrini, and H. A. Fatta, "Prediksi Kelulusan Mahasiswa Magister Teknik Informatika Universitas Amikom Yogyakarta Menggunakan Metode K-Nearest Neighbor," *JTI Respati (J-urnal T-eknologi I-nformasi Respati)*, vol. 13, no. 2, Jul. 2018, doi: 10.35842/jtir.v13i2.260.

[20] A. Advernesia, "Pengertian dan Cara Kerja Algoritma K-Nearest Neighbors (KNN)." [Online]. Available: https://www.advernesia.com/blog/data-science/pengertian-dan-cara-kerja-algoritma-k-nearest-neighbours-knn/

[21] A. Kataria and M. D. Singh, "A Review of Data Classification Using K-Nearest Neighbour Algorithm," *International Journal of Emerging Technology and Advanced Engineering*, vol. 3, no. 6, pp. 354-360, 2013.

[22] Okfalisa, I. Gazalba, Mustakim, and N. G. I. Reza, "Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification," in *2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, Yogyakarta: Indonesia, Nov. 2017, pp. 294–298. doi: 10.1109/ICITISEE.2017.8285514.

[23] C. Eyupoglu, "Breast Cancer Classification Using K-Nearest Neighbors Algorithm," *The Online Journal of Science and Technology*, vol. 8, no. 3, pp. 29-34, 2018.

[24] N. Damastuti, A. Siti Aisjah, and A. A. Masroeri, "Classification of Ship-Based Automatic Identification Systems Using K-Nearest Neighbors," in *2019 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Semarang, Indonesia, Sep. 2019, pp. 331–335. doi: 10.1109/ISEMANTIC.2019.8884328.

[25] S. R. Sannasi Chakravarthy, N. Bharanidharan, and H. Rajaguru, "Deep Learning-Based Metaheuristic Weighted K-Nearest Neighbor Algorithm for the Severity Classification of Breast Cancer," *Innovation and Research in BioMedical engineering (IRBM)*, vol. 44, no. 3, p. 100749, Jun. 2023, doi: 10.1016/j.irbm.2022.100749.

[26] J. C. Macuácua, J. A. S. Centeno, and C. Amisse, "Data mining approach for dry bean seeds classification," *Smart Agricultural Technology*, vol. 5, pp. 100240, Oct. 2023, doi: 10.1016/j.atech.2023.100240.

[27] E. Widura and D. Mardiatno, "Assessment of the Coastal Vulnerability Index (CVI) for disaster mitigation strategies in some coastal tourism areas in Gunungkidul, Yogyakarta-Indonesia," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 989, no. 1, p. 012014, Feb. 2022, doi: 10.1088/1755-1315/989/1/012014.

[28] A. Sumardi, E. Fatimah, and N. Nizamuddin, "Study of Physical Vulnerability Mapping of the Coastal Areas of North - East Aceh," *International Journal of Multicultural and Multireligious Understanding (IJMMU)*, vol. 6, no. 4, pp. 107-114, Aug. 2019, doi: 10.18415/ijmmu.v6i4.944.

[29] B. Rumahorbo, M. Warpur, R. Tanjung, and B. Hamuna, "Spatial Analysis of Coastal Vulnerability Index to Sea Level Rise in Biak Numfor Regency (Indonesia)," *Journal of Ecological Engineering*, vol. 24, no. 3, pp. 113–125, Mar. 2023, doi: 10.12911/22998993/157539.

[30] A. Koroglu, R. Ranasinghe, J. A. Jiménez, and A. Dastgheib, "Comparison of Coastal Vulnerability Index applications for Barcelona Province," *Ocean & Coastal Management*, vol. 178, p. 104799, Aug. 2019, doi: 10.1016/j.ocecoaman.2019.05.001.

[31] A. I. A. Hamid, A. H. M. Din, N. M. Abdullah, N. Yusof, M. R. A. Hamid, and A. M. Shah, "Exploring space geodetic technology for physical coastal vulnerability index and management strategies: A review," *Ocean & Coastal Management*, vol. 214, p. 105916, Nov. 2021, doi: 10.1016/j.ocecoaman.2021.105916.

[32] F. Yulianto *et al.*, "Dataset Indeks Kerentanan Pesisir Pantai Utara Pulau Jawa," *Surat Pencatatan Ciptaan Hak Cipta,* No. 000507353, Badan Riset dan Inovasi Nasional (BRIN), 2023.

[33] S. M. S. Islam, A. H. Tanim, and Md. R. A. Mullick, "Vulnerability Assessment of Bangladesh Coastline Using Gornitz Method," in *Water, Flood Management and Water Security Under a Changing Climate*, A. Haque and A. I. A. Chowdhury, Eds., Cham: Springer International Publishing, 2020, pp. 301–313. doi: 10.1007/978-3-030-47786-8_21.

[34] S. A. Hossain, I. Mondal, S. Thakur, and A. M. Fadhil Al-Quraishi, "Coastal vulnerability assessment of India's Purba Medinipur-Balasore coastal stretch: A comparative study using empirical models," *International Journal of Disaster Risk Reduction*, vol. 77, p. 103065, Jul. 2022, doi: 10.1016/j.ijdrr.2022.103065.

[35] L. Sun, J. Zhang, W. Ding, and J. Xu, "Feature reduction for imbalanced data classification using similarity-based feature clustering with adaptive weighted K-nearest neighbors," *Information Sciences*, vol. 593, pp. 591–613, May 2022, doi: 10.1016/j.ins.2022.02.004.

[36] M. Bach, A. Werner, and M. Palt, "The Proposal of Undersampling Method for Learning from Imbalanced Datasets," *Procedia Computer Science*, vol. 159, pp. 125–134, 2019, doi: 10.1016/j.procs.2019.09.167.

[37] M. Bach, "New Undersampling Method Based on the kNN Approach," *Procedia Computer Science*, vol. 207, pp. 3403–3412, 2022, doi: 10.1016/j.procs.2022.09.399.

[38] Z. Chen, L. J. Zhou, X. D. Li, J. N. Zhang, and W. J. Huo, "The Lao Text Classification Method Based on KNN," *Procedia Computer Science*, vol. 166, pp. 523–528, 2020, doi: 10.1016/j.procs.2020.02.053.

[39] M. Imron and S. A. Kusumah, "Application of Data Mining Classification Method for Student Graduation Prediction Using K-Nearest Neighbor (K-NN) Algorithm," *Indonesian Journal of Interdisciplinary Islamic Studies (IJIIS)*, vol. 1, no. 1, pp. 1–8, Sep. 2018, doi: 10.47738/ijiis.v1i1.17.

[40] G. Bergström *et al.*, "Evaluating the layout quality of UML class diagrams using machine learning," *Journal of Systems and Software*, vol. 192, p. 111413, Oct. 2022, doi: 10.1016/j.jss.2022.111413.

[41] S. H. Jayady and H. Antong, "Theme Identification using Machine Learning Techniques," *Journal of Integrated and Advanced Engineering (JIAE)*, vol. 1, no. 2, pp. 123-134, 2021, doi: 10.51662/jiae.v1i2.24