WAVES WITH MULTI-SENSOR SYSTEM BASED ON WEB APPLICATION USING NAIVE BAYES ALGORITHM

Jhosua Parningotan Sianipar[1], Randy Erfa Saputra[2], Casi Setianingsih[3]

[1,2,3] Telkom University, Bandung
jhosuasianipar@student.telkomuniversity.ac.id[1], resaputra@telkomuniversity.ac.id[2], setiacasie@telkomuniversity.ac.id[3]

***Abstract*—Indonesia is the largest archipelago in the world and has wider waters than land. Indonesia is also one of the countries most vulnerable to tsunami natural disasters. Tsunami is a natural disaster event in which sea waves with large size and high speed hit the coastal surface. Tsunamis usually occur due to tectonic earthquakes that occur on the seabed or the eruption of Mount Merapi located near the sea. Tsunamis don't just happen. This disaster has a process or anomaly in a ocean wave, such as the rapid receding of sea water. Currently the tsunami detection system already exists but this system is still very minimal in Indonesia because of the relatively high price.**

The purpose of this study is so that the public can monitor the current sea conditions. Created a web application with the help of the Laravel framework that can provide information on sea conditions in real time. This web application can also classify sea conditions in the form of "Potential" and "Not Potential" by using the Naive Bayes Algorithm which can provide 95% accuracy. Therefore, with this application can make it easier for the community to monitor the condition of the sea.

***keyword*—Naive Bayes Algorithm, Web Application, Tsunami.**

I. INTRODUCTION

Indonesia is the largest island nation in the world which certainly has a very large sea. Indonesia has 17,499 islands with Indonesia's total area of about 7.81 million km^2. Of Indonesia's total area, 3.25 million km^2 is the ocean, 2.55 million km^2 is the Exclusive Economic Zone (EEZ), and only about 2.01 km^2 is land[1]. With the wider territorial waters in Indonesia compared to the land, of course, the greater the threat of natural disasters in the sea such as tsunamis. The threat of such a disaster is certainly very troubling for people, especially those living on the coast [2].

With the large threat of disasters occurring at sea, an application system is needed that can monitor and classify sea conditions in real-time and has a high level of accuracy so that the public can monitor sea conditions in real-time and is expected to be more quickly aware in the event of a disaster and can more quickly evacuate to a safer place.

With the above problems, for that to create a system that aims to monitor the condition of a sea in real-time where the system is a website application that can inform sea conditions such as sea depth, wave speed, to vibrations that occur and can also classify a sea condition with a high degree of accuracy so that the public can more easily monitor a sea condition, and can anticipate quickly if sea conditions are worsening to reduce casualties if sea conditions are worsening.

II. RELATED WORK

A. *Tsunami Early Warning Detection Using Bayesian Classifier (Dewi Yanti Liliana and Dewi Yanti Liliana,2019)*

Dewi Yanti Liliana et al, conducting research on the tsunami early warning detection system. The system processes data using naïve bayes algorithm with classifications as many as two namely "Tsunami" and "NotTsunami" by using parameters as many as three, namely: earthquake, magnitude, epicenter, and place. This study aims to warn the public if there are signs of a tsunami. The results of this study succeeded in creating a system using naïve bayes algorithm by providing optimistic value; Using varying training data, the system obtained 92.37% accuracy and 0.98 on an F1 score[3].

B. *Sentiment Analysis System on Product Reviews Using the Naive Bayes Method (Billy Gunawan, Helen Sasty Pratiwi, Enda Esyudha Pratama, 2018)*

Billy Gunawan et al researched the online product review analysis system. In the study they used Bayes' naïve method of classifying review classes on products, there were five classes of data classified, namely very negative, negative, neutral, positive, and very positive. Tests were conducted in three classes (negative, neutral, and positive) the best results were obtained from 90% of training data, and 10% of test data managed to get an accuracy score of 77.78%, recall 93.33%, and precision of 77.78%. In testing five classes the system managed to get an ad accuracy score of 59.33%, recall 58.33%, and precision 59.33%[4].

III. LITERATURE REVIEW

A. *Ocean Waves*

Ocean waves are the occurrence of an energy transfer that is characterized by the rise and fall of seawater with a certain period and wavelength[5]. The energy transfer that occurs in the sea is caused by several factors, namely wind factors, earth-moon-sun attraction, earthquakes in the sea, and waves caused by ships[6].

B. *Tsunami*

Tsunami waves are one of the natural disasters that can cause severe damage to the point of fatalities. The word tsunami comes from the Japanese language which consists of two words, namely Tsu meaning port, Nami means wave. Initially, a tsunami was a large wave that hit the port, so the word was eventually used around the world to refer to the natural disaster that hit the coast[3].

In general, tsunamis occur due to disturbances on the seafloor such as tectonic earthquakes that cause tectonic plates to shift vertically so that seawater is sucked in and the occurrence of water shrinkage is quite a lot. At a certain time, the water that is sucked in all coastal directions at a very high speed reaches 600-900 Km / h. At the time of the tsunami at first, the waves of seawater have a small amplitude, but by the time the wave approaches the shore, the wave amplitude becomes so large that it damages whatever the wave passes[7].

C. Machine Learning

Machine learning is a machine that was developed to be able to learn by itself without having to be directed by users. Machine learning was developed with the disciplines of statistics, mathematics, and data. The term machine learning is basically a machine process for learning from data[8]. Therefore, in machine learning, there are 2 data, namely data training and data testing. Data training is data used by machines to learn and train algorithms while data testing is data used to test the results of performance from algorithms that have been trained and studied by previous machines, to get results in the form of models. The model obtained in the form of predictions is divided into two forms, namely, if the output is discrete then it is called classification[9], while the output is continuous then it is called regression.

D. Naive Bayes Algorithm

Naive Bayes is an algorithm that classifications data by calculating group probabilities and summing up combinations of values from the data sets that have been collected [10]. The Naive Bayes classification has great and stable accuracy results. However, Naive Bayes results in some conditions that will result in a low accuracy value if the features and parameters of the data are added [11].

In this study the processed data is numerical data, for that in calculating the probability value of the class can use the standard function of probability density. In the probability function density can represent the distribution of known data, here is the formula of Gauss Density[12].

$$P(X_i = x_i | Q_i = q_i) = \frac{1}{\sqrt{2\Pi\sigma_{ij}}} e^{-\frac{(xi-uij)^2}{2\sigma^2 ij}} \quad (1)$$

Information:

P	:	Probability
X_i	:	n Feature
x_i	:	n Feature Value
Q	:	Related classes
q_i	:	Related subclasses
u	:	Average value on related features
σ	:	Standard deviation on related features

To search for the mean value or mean value on a related feature, it can be searched using the following formula.

$$u = \frac{1}{n}\sum_{i=1}^{n} xi \quad (2)$$

Information:

u	:	Average value on related features
x_i	:	n Feature Value
n	:	Number of features

To find the standard deviation value of each related feature, it can be searched using the following formula.

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(xi-u)^2} \quad (3)$$

Information:

u	:	Average value on related features
x_i	:	n Feature Value
n	:	Number of features
σ	:	Standard deviation on related features

In the Naive Bayes Algorithm, there are several things that must be sought before classifying a data, as follows:

1. *Calculate the probability value of each prior class*

$$P(Q_i) = \frac{NQ_i}{N} \quad (4)$$

Information:

$P(Q_i)$	:	Probability value of a condition
NQ_i	:	Number of conditions
N	:	number of all conditions

2. *Calculate the likelihood value of each class*

$$L(Q_i) = P(X_1|Q_i)x \dots xP(X_n|Q_i)xP(Q_i) \quad (5)$$

Information:

$L(Q_i)$	:	The likelihood value of a condition
$P(X_1\|Q_i)$	:	Probability value of a feature in a condition
$P(Q_i)$	:	Probability value of a condition

3. *Calculate the posterior of each class*

$$Posterior(Q) = \frac{L(Q_i)}{P(Q_i)} \quad (6)$$

Information:

$L(Q_i)$	:	The likelihood value of a condition
$P(Q_i)$	:	Probability value of a condition

E. Earthquake Intensity Scale

The Earthquake Intensity Scale of the Meteorology, Climatology, and Geophysics Agency is a guide in measuring the impact of damage caused by earthquake disasters compiled by the Meteorology, Climatology, and Geophysics Agency Indonesia[13]. The arranged scale consists of 5 scales that can be seen in figure 1

Scale	Color	Simple Description	Detailed Description	MMI scale	PGA (gal)
I	White	Not Felt	Not felt or felt by only a few people but recorded by tools.	I-II	<2.9
II	Green	Felt	Felt by the crowd but did not cause any damage. Light objects were hung swaying and glass windows shook.	III-V	2.9-88
III	Yellow	Slight Damage	Non-structural parts of the building suffered minor damage, such as hair cracking on the walls, the roof shifting downwards and partly falling.	VI	89-167
IV	Orange	Moderate Damage	Many cracks occur in the walls of simple buildings, some collapse, glass breaks. Some of the wall plaster came off. Most of the roof shifts down or falls. The structure of the building suffered minor to moderate damage.	VII-VIII	168-564
V	Red	Heavy Damage	Most of the walls of the building permanently collapsed. The structure of the building suffered heavy damage. The railway tracks are curved.	IX-XII	>564

Fig 1. Earthquake Intensity Scale*[13]*

III. RESEARCH METHOD

A. *System Overview*

Systems created to produce a classification of sea conditions can be seen in Figure 1 where data is retrieved from sensors and sent on the ThingSpeak database and web applications will request data through the ThingSpeak API and ThingSpeak will send data in JSON format and data processed on machine learning with Naive Bayes Algorithm after the data is successfully processed then the results of classification of marine conditions will be displayed on the web application page. Web applications and machine learning are created using the PHP programming language and support from the Laravel framework.

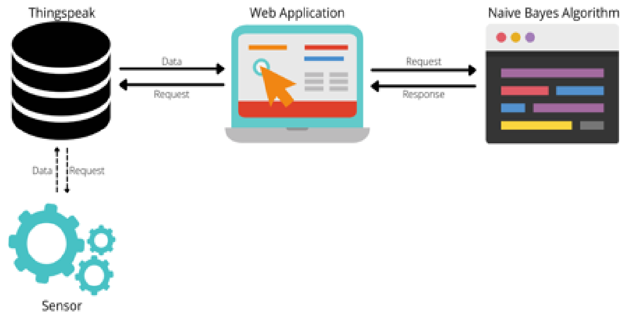

Fig 2. System Overview

B. *Overview of Naïve Bayes Classification*

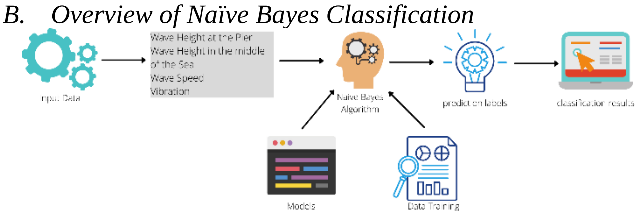

Fig 3. Overview of Naïve Bayes Classification

Figure 3.2 shown the system process of the Naive Bayes classification in determining ocean conditions, here are the processes the system occurs:

- Data Input: The process of receiving data from sensors through ThingSpeak in the form of JSON data. The parameters of the data taken are wave height at the pier and in the middle, speed, and vibration.
- Classification of Naive Bayes: Incoming data will be processed on a model based on training data using machine learning with the Naive Bayes Algorithm.
- Prediction Label: The data that has been processed will be given a prediction label according to the highest probability value. So that the results of the classification will be displayed on the web page.

C. *Naïve Bayes Algorithm Implementation*

Naive Bayes is an algorithm that classifies data by calculating the probability value of data. Naive Bayes predicts the odds of a condition based on experience or learning from training data that has been entered. In figure 3.6 can be seen in the first stage of the training data will be entered and calculated the probability value of the condition of the training data, after which the test data will be entered sent from the ThingSpeak database then calculated the probability value of each training data that has been entered, and finally will be done posterior probability calculation to get the classification results of the test data.

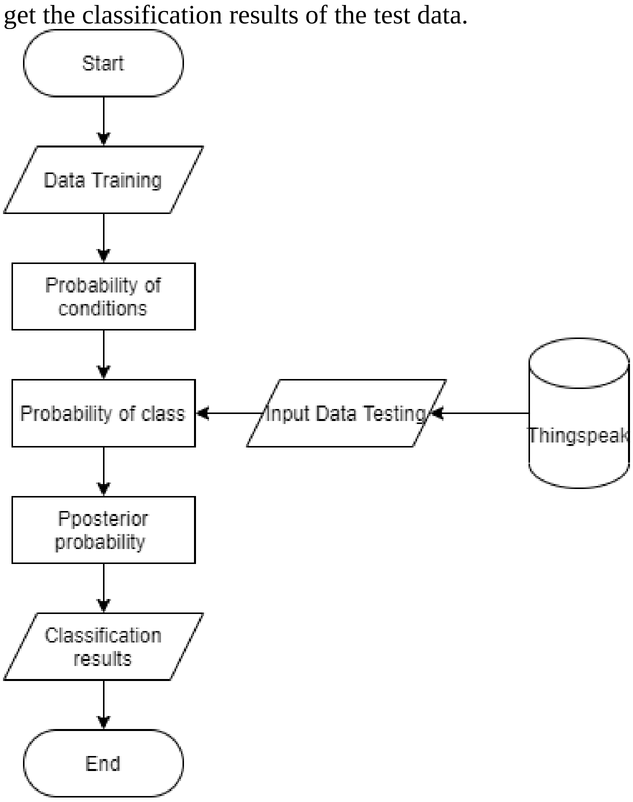

Fig 4 Naive Bayes Flow Chart

1) *Data Training*

Table 1. Data Training

Id	TD (m)	TT (m)	KC (m/s)	GN (Gal)	Status
1	0.74	2.21	4.56	0.21	not potential
2	0.21	1.15	2.91	107.76	potential
3	0.11	1.41	1.57	682.96	potential

Id	TD (m)	TT (m)	KC (m/s)	GN (Gal)	Status
4	0.17	1.16	14.16	520.6	potential
5	0.72	1.61	4.44	63.07	not potential
...	...	...	...	...	...
500	1	1.85	8.15	0.82	not potential

As can be seen in Table 1, in the training data 500 data are consisting of TD (Wave Height at the Pier), TT (Wave Height in the middle of the Sea), KC (Wave Speed), GN (Vibration) which is divided into two classifications namely "not potential", and "potential". Of the 500 data in the training data, there are 218 data with "not potential" status and 282 data with "potential" status. To calculate the probability of each status can use the following formula 4. Probability results from each class are obtained.

$$P(not\ potential) = \frac{218}{500} = 0.436$$

$$P(potential) = \frac{282}{500} = 0.564$$

Table 2. Probability Value of Each Condition

not potential	potential
0.436	0.564

2) *Class Probability*

Before looking for a class probability value, the mean and standard deviation values of each class must be sought first. To find the mean value you can use the formula (2). The mean value of each class is obtained as follows

$$u(TD|not\ potential) = \frac{307.48}{218} = 1.410$$

$$u(TD|potential) = \frac{365.96}{282} = 1.297$$

Table 3. Mean Value of Each Class

	TD	TT	KC	GN
u(not potential)	1.410	2.415	8.010	219.488
u(potential)	1.297	2.438	9.192	333.903

To find the standard deviation value of each class can use Formula (3). The standard deviation value of each class is obtained as follows

$$\sigma(TD|not\ potential) = \sqrt{\frac{(307.48 - 1.410)^2}{218 - 1}} = 0.379457$$

$$\sigma(TD|potential) = \sqrt{\frac{(365.96 - 1.297)^2}{282 - 1}} = 0.62051$$

Table 4. Standard Deviation Value of Each Class

	TD	TT	KC	GN
σ(not potential)	0.379	0.660	4.939	26.9428
σ(potential)	0.620	0.734	5.245	223.468

After obtaining the mean and standard deviation values of each class, the probability values of each class are obtained. Known values of TD=1.4 m, TT=2.8 m, KC=7.3 m/s, and GN=12.4 Gal.

$$P(1.4|not\ potential) = \frac{1}{\sqrt{2\Pi * 0.379}} e^{-\frac{-(1.4-1.410)^2}{2(0.379)^2}} = 0.647$$

$$P(1.4|potential) = \frac{1}{\sqrt{2\Pi * 0.620}} e^{-\frac{-(1.4-1.297)^2}{2(0.620)^2}} = 0.499$$

Table 5. Probability Value of Each Class

	TD	TT	KC	GN
Data testing	1.4	2.8	7.3	12.4
P(not potential)	0.647	0.414	0.177	0.074
P(potential)	0.499	0.412	0.163	0.009

3) *Posterior Probability*

After obtaining the probability value of each class next look for the posterior probability value of each condition, but before looking for the posterior value the probability must look for the likelihood value of each condition first by using the following formula 5. The likelihood of each class is obtained.

$$L(not\ potential) = 0.647551x0.414691x0.177699x0.0074262 = 0.001545$$

$$L(potential) = 0.499744x0.412622x0.163248x0.009483 = 0.00018$$

Table 6. Likelihood Value of Each Class

L(not potential)	L(potential)
0.001545	0.00018

After obtaining the likelihood value of each class, the posterior probability score of each class is obtained using the following formula 6. The posterior of each class is obtained.

$$posterior(not\ potential) = \frac{0.001545}{0.436} = 0.0035435$$

$$posterior(potential) = \frac{0.00018}{0.564} = 0.0003191$$

posterior(not potential)	posterior(potential)
0.0035435	0.0003191

4) *Classification Results*

After obtaining posterior grades from each class, the classification results of each class are obtained as follows:

$$P(not\ potential) = \frac{0.0035435}{0.0035435 + 0.0003191} x100 = 91.73\%$$

$$P(potential) = \frac{0.0003191}{0.0035435 + 0.0003191} x100 = 8.27\%$$

Because the value of P(not potential) is greater than the value of P(potential) then the classification result of the testing data with the value TD = 1.4, TT = 2.8, KC = 7.3, GN = 1.2 is not potential.

IV. TESTING AND RESULT

A. *Naïve Bayes Algorithm Testing*

Data classification testing is carried out as many as 20 times. Here is a table of data classification tests from the algorithm.

Table 7. Data Classification Testing

No	Data Testing				Prob		Inf
	TD	TT	KC	GN	TB	B	
1	0.7	2.2	4.5	0.2	78	22	V
2	0.2	1.1	2.9	107	1	99	V
3	2.1	4.0	5.2	611	0	100	V
4	2.1	3.7	3.4	97	3	97	V
5	1.4	2.4	15	0	85	15	V
6	0.7	1.4	0.2	0.8	75	25	V
7	0.1	1.1	2.6	264	0	100	V
8	0.2	1.1	1.1	100	1	99	V
9	0.4	1.1	8.9	131	1	99	V
10	0.6	1.2	4.0	1.4	72	28	V
11	0.6	1.1	16	2.7	53	47	V
12	0.6	1.1	3.2	0.8	70	30	V
13	0.9	1.9	0.2	56	70	30	V
14	1.0	2.0	10	1.8	86	14	V
15	0.9	1.6	5.4	80	25	75	N
16	0.1	1.6	14	503	0	100	V
17	0.3	1.4	10	69	7	93	V
18	0.1	1.6	17	102	1	99	V
19	1.1	1.8	6.8	15	90	10	V
20	1.1	1.9	3.3	1.7	90	10	V

Information:
TD : Wave Height at the Pier
TT : Wave Height in the middle of the Sea
KC : Wave Speed
GN : Vibration
TB : Not potential
B : Potential
V : Valid
N : Invalid

The results of the experiment above valid and invalid values are obtained from the comparison of data on the rules of classification of sea status that have been validated from the Indonesian Meteorology, Climatology, and Geophysics Agency. From the above experiment, the accuracy of the Naïve Bayes Algorithm is obtained as follows:

$$Accuracy = \frac{19}{20} x100 = 95\%$$

B. *Duration of Bayes Naive Algorithm Process*

To find out the duration in processing data using Naive Bayes running on applications, ranging from data stored on variables to processed to get output results. Here is the duration process obtained:

Table 8. Duration of Bayes Naive Algorithm Process

No	Data Testing				Duration
	TD	TT	KC	GN	
1	0.7	2.2	4.5	0.2	1.81
2	0.2	1.1	2.9	107	2.33
3	2.1	4.0	5.2	611	1.76
4	2.1	3.7	3.4	97	1.78
5	1.4	2.4	15	0	1.77
6	0.7	1.4	0.2	0.8	1.68
7	0.1	1.1	2.6	264	2.19
8	0.2	1.1	1.1	100	1.7
9	0.4	1.1	8.9	131	1.92
10	0.6	1.2	4.0	1.4	1.62
11	0.6	1.1	16	2.7	1.59
12	0.6	1.1	3.2	0.8	1.82
13	0.9	1.9	0.2	56	1.71
14	1.0	2.0	10	1.8	1.75
15	0.9	1.6	5.4	80	1.85
16	0.1	1.6	14	503	1.67
17	0.3	1.4	10	69	1.8
18	0.1	1.6	17	102	1.66
19	1.1	1.8	6.8	15	1.72
20	1.1	1.9	3.3	1.7	1.95

From the results of the test as many as 30 times, it can be calculated the average duration of the process of the Naive Bayes Algorithm as follows:

$$Average\ duration = \frac{\sum Duration}{\sum experiment}$$

$$Average\ duration = \frac{36.8}{20} = 1.804$$

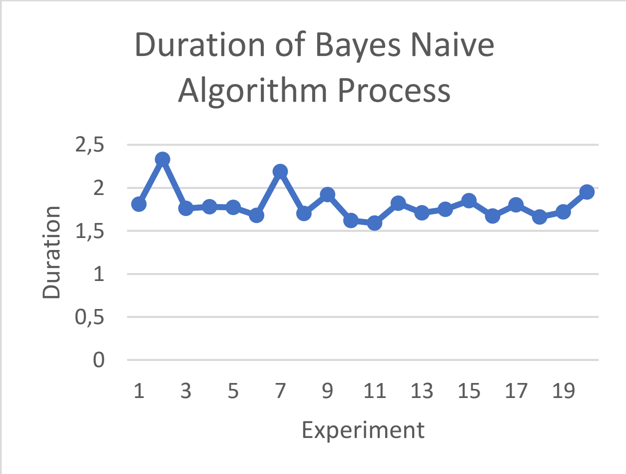

Fig 5. Bayes Naive Algorithm Duration Testing Graph

V. CONCLUSION

Based on the results of testing and analysis on this final task research, it can be concluded as follows.

1. Tsunami warning mitigates system with nautical wave height parameters, wave speed, and vibration using web-based Naive Bayes Algorithm with "no-hassle" and "potential" classifications has an accuracy of 95%.

2. Naive Bayes algorithm in performing the data classification process requires an average duration of 1,804 seconds.

References

[1] "KKP | Kementerian Kelautan dan Perikanan." https://kkp.go.id/djprl/artikel/21045-konservasi-perairan-sebagai-upaya-menjaga-potensi-kelautan-dan-perikanan-indonesia (accessed Aug. 01, 2021).

[2] A. Ariansyah, "Potensi Ancaman Bencana," *BNPB*. https://bnpb.go.id/potensi-ancaman-bencana (accessed Aug. 18, 2021).

[3] D. Y. Liliana and D. Priharsari, "Tsunami Early Warning Detection using Bayesian Classifier," in *2019 2nd International Conference of Computer and Informatics Engineering (IC2IE)*, Banyuwangi, Indonesia, Sep. 2019, pp. 44–48. doi: 10.1109/IC2IE47452.2019.8940823.

[4] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *JEPIN*, vol. 4, no. 2, p. 113, Dec. 2018, doi: 10.26418/jp.v4i2.27526.

[5] M. Z. Arifin and J. I. Manengkey, "Modul : Meteorologi Dan Oseonografi." http://www.pusdik.kkp.go.id/elearning/index.php/modul/read/190116-091446uraian-c-materi (accessed Aug. 01, 2021).

[6] W. L. Dhanista, "Gelombang Laut - Departemen Teknik Kelautan." https://www.its.ac.id/tkelautan/gelombang-laut/ (accessed Aug. 07, 2021).

[7] T. Ilyas, "MITIGASI GEMPA DAN TSUNAMI DIDAERAH PERKOTAAN," presented at the Seminar Bidang Kerekayasaan Fatek-Unsrat 2006.

[8] D. Intern, "Apa itu Machine Learning? Beserta Pengertian dan Cara Kerjanya," *Dicoding Blog*, Aug. 18, 2020. https://www.dicoding.com/blog/machine-learning-adalah/ (accessed Aug. 07, 2021).

[9] A. Ahmad, "Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning," p. 6, 2017.

[10] A. H. jahromi, "A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features," presented at the 2017 Artificial Intelligence and Signal Processing Conference (AISP, Shiraz,Iran.

[11] J. Chen, Z. Dai, J. Duan, H. Matzinger, and I. Popescu, "Naive Bayes with Correlation Factor for Text Classification Problem," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, Boca Raton, FL, USA, Dec. 2019, pp. 1051–1056. doi: 10.1109/ICMLA.2019.00177.

[12] W. Li and Y. Jia, "Distributed Gaussian sum filter for discrete-time nonlinear systems with Gaussian mixture noise," in *2016 35th Chinese Control Conference (CCC)*, Chengdu, China, Jul. 2016, pp. 1831–1836. doi: 10.1109/ChiCC.2016.7553360.

[13] BMKG, "Skala Intensitas Gempabumi (SIG) | BMKG," *BMKG | Badan Meteorologi, Klimatologi, dan Geofisika*. https://www.bmkg.go.id/gempabumi/skala-intensitas-gempabumi.bmkg (accessed Aug. 23, 2021).