Optimization Naive Bayes using Particle Swarm Optimization in Volcanic Activities

F Tempola¹, A Mubarak¹

¹Department Of Informatics, Faculty Of Engineering, Universitas Khairun, Indonesia

*Corresponding author: firman.tempola@unkhair.ac.id

Abstract. This study is a continuation of previous studies that apply Naive Bayes classifier algorithm for predicting the status of volcanoes in Indonesia based on factors of seismicity. There are 5 criteria used in predicting the status of the mountain, namely the status of normal, alert and standby. The results of the study showed that the system accuracy produced was only 79.31%, in other words, it was still at the stage of fair classification. To overcome these weaknesses so that accuracy increases, optimization is done by giving the weight of criteria or attributes using particle swarm optimization. From the results of research by applying the same data using Particle Swarm optimization methods optimization, the accuracy of the resulting system increase of to 95.65%, where the number of particles is initialized 50 and the weight range [0 2].

1. Introduction

In Indonesia, volcanic eruptions occur almost every year, because of many there is volcanic active in Indonesia. Not only that, position Indonesia's geography is located on the Asian and Australian plates as well to be wrong one factor often occurrence disaster caused tectonic from volcanic eruption. Interestingly, from 127 active volcanoes in Indonesia, only 69 were monitored. In addition, when the volcano erupted the impact on the community was still so great. This is because many people who live not far from the volcano. Call it Mount Gamalama, which is in the city of Ternate. Which is where the residents of Ternate city reside right below the foot of Mount Gamalama.

In determining the status of volcanoes in Indonesia, the institution that played a role in issuing recommendations was the Center for Volcanology and Geological Disaster Mitigation (PVBMG). PVMBG in issuing recommendations on volcanic activity status based on data monitored from the activities of each volcanic. There are two ways of monitoring that are done based on visual observation and seismic factors.
In the current global era, developments in the world of computer science continue to grow. Computer science is here to solve and facilitate various problems, including the recommendation of the status of volcanoes. As has been done by [1] [2] Especialy for volcanic activities in Indonesia has been done by [3], by comparing the two algorithms k-nn and naive bayes, the result is that naive bayes accuracy is higher than that of knives. However, the accuracy obtained is still at the level of fair classification [4]. To improve accuracy on Naive Bayes, optimization needs to be done by giving weights to criteria or attributes [5]. Naive Bayes algorithm optimization so that accuracy can be improved has been done by [5] where the optimization algorithm is done is particle swarm optimization (PSO). Likewise, what was done by [6], [7] conducted optimization of Naive Bayes with Particle Swarm Optimization, which is a difference in the results of accuracy between those who only use Naive Bayes and optimization of Naive Bayes with PSO. When using optimization with PSO system accuracy increases. Therefore in this study, will be optimized Naive Bayes with particle swarm optimization in the determination of the status of the volcano in Indonesia.

2. Methods

This research is experimental, namely by optimizing the algorithm naive bayes classifier using particle swarm optimization on volcanic activity data in Indonesia with the aim of increasing system accuracy. The steps of this study are shown in Figure 1.
2. Dataset
The data in this study were obtained from public data released on the official site of the Volcanology and Volcanic Disaster Mitigation Center, Ministry of Energy and Mineral Resources. These data are as shown in Table 1, then processed and tested on the system that has been built.

Table 1. Dataset of volcanic activities

| No | Shallow volcanic | Deep Tectonics | In Volcanic | Earthquake blowing | Previous Status | Recommended Status |
|----|-----------------|----------------|-------------|--------------------|-----------------|-------------------|
| 1  | 2               | 68             | 1           | 41                 | Standby         | Standby           |
| 2  | 0               | 35             | 10          | 0                  | Standby         | Normal            |
| 3  | 19              | 54             | 62          | 114                | Standby         | Alert             |
| 4  | 2               | 10             | 7           | 13                 | Alert           | Standby           |
| 5  | 2               | 15             | 1           | 63                 | Standby         | Alert             |


|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 6 | 144 | 131 | 17 | 53 | Standby |
| 67 | 78 | 135 | 17 | 0 | Normal |
| 68 | 1 | 13 | 5 | 0 | Standby |
| 69 | 17 | 25 | 5 | 157 | Alert |

2.2 Naive Bayes Algorithm

**Naive Bayesian Classifier** (NBC) is one method of probabilistic reasoning, which is based on the Bayes theorem. NBC is a classification algorithm that is very effective (getting the right results) and efficient (the reasoning process is done using existing inputs in a relatively fast way). Another advantage of NBC can handle data both discrete and continuous. The workings of Naive Bayes is by calculating a set of probabilities by summing the frequencies and combinations of values from the given dataset. The general form of the Bayes theorem as shown in Equation 1.

\[
P(H|X) = \frac{P(X|H)P(H)}{P(X)}
\]  

(1)

Where :

- X = data with an unknown class
- H = The hypothesis of data X is a specific class
- \( P(H|X) \) = Probability of Hypothesis H Based on Condition x (Posteriori probability)
- \( P(H) \) = Probability of H hypothesis (Prior Probability)

The conditional opportunities of categorical attributes are expressed in the form of equation 2.

\[
P(A_i|C_j) = \frac{|A_{ij}|}{N_{c,j}}
\]

(2)

Where \(|A_{ij}|\) is the number of training examples of class \( A_i \) that receives the value of \( C_j \). While \( N_{c,j} \) is the amount of training data from class \( A_i \).

Whereas for opportunities with a continuous requirement, it is expressed with the density of gauss as in Equation 3.

\[
P(\chi) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(\chi-m)^2}{2\sigma^2}}
\]

(3)

2.3 Particle Swarm Optimization

**Particle Swarm Optimization** (PSO) is a global optimization method introduced by Kennedy and Eberhart in 1995 based on research on the behavior of flocks of birds and fish in foraging. The theory is based on three main principles, namely Evaluation, Compare and Imitate. Because the principle of Particle Swarm Optimization can improve the accuracy of the previous algorithm because it is able to compare before executing [8]

In this PSO algorithm, the search for solutions is carried out by a population consisting of several particles. Populations are randomly generated with the smallest and largest limits. Each particle represents the position or solution of the problem at hand. Each particle searches for an optimal
solution by traversing the search space. This is done by way of each particle adjusting to the best position of the particle (local best) and adjusting to the position of the best particles of all flocks (global best) while crossing the search space. Thus, the spread of experience or information occurs within the particle itself and between a particle and the best particles of all flocks during the process of finding a solution. After that, a search process is carried out to find the best position of each particle in a certain number of iterations until a relatively steady position is reached or reaches a predetermined iteration limit. In each iteration, each solution represented by the position of the particle is evaluated for its performance by entering the solution into the fitness function [9] The process of the PSO algorithm is as follows:

a. Initialization
   - The first-speed initialization
     In iteration 0, it can be ascertained that the value of the initial velocity of all particles is 0
   - Initialization of the first particle position
     In the 0 iterations, the first position of the particles is generated by equation 4:
     \[ x = x_m + r \times [0,1] \times (x_m - x_m) \] (4)
   - pBest and gBest initialization
     In the 0 iterations, pBest will be equaled to the initial position value of the particle. While gBest is selected from one pBest with the highest fitness

b. Update speed: To update speed, using Equation 5.
   \[ v_{i,j}^{t+1} = w \times v_{i,j}^t + c_1 \times r_1 \times \left[ b \times (t_j^t - x_{i,j}^t) \right] + c_2 \times r_2 \times \left[ G \times (g_{i,j}^t - x_{i,j}^t) \right] \] (5)

c. Update position and calculate fitness: To update the position, using Equation 6.
   \[ x_{i,j}^{t+1} = x_{i,j}^t + v_{i,j}^{t+1} \] (6)

d. Update pBest and gBest
   Do the comparison between pBest in the previous iteration with the results of the position update. Higher fitness will be new pBest. the newest pBest that has the highest fitness value will be the new gBest.

3. Result and Discussion

This research will be done by optimizing attributes by giving weight using particle swarm optimization algorithm. Before optimization with particle swarm optimization, the initial step is to divide the dataset into two parts, namely training data and testing data. Then in the test divided several tests to find the highest accuracy. for two types of testing, namely particle number testing and weight limit testing. In particle testing with multiples of 10, the weight limits are made in the range [0 1], [0 2] and [0 3].

The results of the first testing of the number of particles that were estimated as many as 10 and the weight range namely [0 1] obtained system accuracy of 73.91% with the final weight of each criterion namely shallow volcanic = 0.572, distant tectonics = 0.152, volcanic within = 1, earthquake blowing = 1, previous status = 1. Figure 2 is the final result of each criterion in the first test. The same is true for the first test, in the second test with particle number 20, the third test with 30 particle counts, the fourth test with 40 particle counts, testing the five particle counts 50 resulting in system accuracy that is 73.91% different only the final weight of each criterion.
Different results were obtained in the sixth test with particles initialized 10, 20, 30 and 40 with weight ranges [0 2] the system accuracy obtained was 91.3 % with the final results of each criteria weight shown in Table 2.

**Table 2.** Comparison of each testing

| Particle amount | Shallow volcanic | Deep Tectonics | In Vulcanic | Earthquake Blowing | Previous status |
|-----------------|-----------------|----------------|------------|--------------------|-----------------|
| 10              | 0.572           | 0.152          | 1          | 1                  | 1               |
| 20              | 0.654           | 0.159          | 1          | 1                  | 1               |
| 30              | 0.683           | 0.162          | 1          | 1                  | 1               |
| 40              | 0.698           | 0.163          | 1          | 1                  | 1               |
| 50              | 0.707           | 0.164          | 1          | 1                  | 1               |
| 10              | 1.333           | 0.44           | 2          | 1,729              | 2               |
| 20              | 1.333           | 0.44           | 2          | 1,729              | 2               |

**Figure 2.** The Final result of criteria weights
The next test was carried out the initialization of 50 particles with weight ranges [0 2] accuracy increased by 4.35%, so that obtained system accuracy of 95.65%. Then testing again with 50 particle counts and weight ranges [0 3] the accuracy of the system has not changed. As the graph of the test results is shown in Figure 3. Thus the test is stopped because it has the highest accuracy.

![Figure 3. result of accuracy every testing](image)

In accordance with the results obtained that the accuracy of the system obtained when performing optimization with Naive Bayes has increased namely when the weight range between [0 2] and the number of initialized particles is 30 to 50 particles. This was also obtained by [10] namely when performing Naive Bayes optimization with PSO on microaneurysm detection in fundus images. It's just that there is a difference with [10] that is optimized is the selection of features not on the weight, as well as that done by [11] and [12] which is optimized, is the value of the feature, not the weight. Whereas in the research [5] the accuracy of the system also increased only in research [5] in the naive Bayes optimization test with Particle swarm optimization which was done only in determining the number of particles not the initial weight range initialization test was performed.

### 4. Conclusions and Recommendations

Based on the results of testing the optimization system Naive Bayes with particle swarm optimization can be concluded, namely:
1. Optimization with Particle swarm optimization on naive bayes classifier for volcanic activity data can improve system accuracy
2. Initializing the weight limit and number of particles in particle swarm optimization can affect the results of system accuracy
3. Optimization naive bayes with PSO can be tried in other cases where the system accuracy is still low.
4. Experiments can be done by using another optimization algorithm to compare with the particle swarm optimization

References
[1] Reath K A, Ramsey M S, Dehn J, and Webley P W 2016 “Predicting eruptions from precursory activity using remote sensing data hybridization,” J. Volcanol. Geotherm. Res., vol. 321, pp. 18–30.
[2] Tempola F, Muhammad M, and Khairan A, 2018 “Naive Bayes Classifier for Prediction of Volcanic Status in Indonesia,” Proc. - 2018 5th Int. Conf. Inf. Technol. Comput. Electr. Eng. ICITACEE 2018, pp. 365–369
[3] Tempola F, Muhammad M, and Khairan A, 2018 “Perbandingan Klasifikasi Antara Knn Dan Naive Bayes Pada Penentuan Status Gunung Berapi Dengan K-Fold Cross Validation Comparison of Classification Between Knn and Naive Bayes At the Determination of the Volcanic Status With K-Fold Cross,” J. Teknol. Inf. dan Ilmu Komput., vol. 5, no. 5, pp. 577–584.
[4] Gorunescu F, 2011 Data Mining: Concepts, Models and Techniques. London: Springer, 2011.
[5] Muhamad H et al., 2017 “Optimasi Naïve Bayes Classifier Dengan Menggunakan Particle,” J. Teknol. Inf. dan Ilmu Komput., vol. 4, no. 3, pp. 180–184
[6] Agustina C, 2017 “Optimasi Naive Bayes Menggunakan Particle Swarm Optimization Untuk Meningkatkan Akurasi Deteksi Autisme Spectrum Disorder,” J. Speed – Sentra Penelit. Eng. dan Edukasi, vol. 10, no. 2, pp. 1–5.
[7] Idrus A, Brawijaya H, and Maruloh, 2018 “Sentiment Analysis of State Officials News on Online Media Based on Public Opinion Using Naive Bayes Classifier Algorithm and Particle Swarm Optimization,” 2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018, no. Citsm, pp. 1–7.
[8] Miao C.-L. L. S.-W and Fei, Y.-B 2009 “Chinese Grain Production Forecasting Method Based on Particle Swarm Optimization-based Support Vector Machine,” Recent Patents Eng., vol. 3, pp. 8–12.
[9] Santosa B and Willy P, Metoda metaheuristik, Konsep dan Implementasi. Yogyakarta: Graha Ilmu, 2011.
[10] Kumar S N, Dinesh D, Siddharth T, Ramkumar S, Nikhill S, and Lavanya R, 2017 “Selection of features using particle swarm optimization for microaneurysm detection in fundus images,” in Proceedings of the 2017 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET, vol. 2018-Janua, pp. 140–144.
[11] K. Agarwal, “approach of Naïve Bayes and Particle Swarm Optimization,” 2018 Second Int. Conf. Intell. Comput. Control Syst., no. Iciccs, pp. 685–690.

[12] A. Pandhu Wijaya and H. Agus Santoso, 2018 “Improving the Accuracy of Naïve Bayes Algorithm for Hoax Classification Using Particle Swarm Optimization,” Proc. - 2018 Int. Semin. Appl. Technol. Inf. Commun. Creat. Technol. Hum. Life, iSemantic 2018, pp. 482–487.