JPPI Vol 9 No 2 (2019) 97 - 103

Jurnal Penelitian Pos dan Informatika

32a/E/KPT/2017

e-ISSN 2476-9266
p-ISSN: 2088-9402

[Doi:10.17933/jppi.2019.090202](https://doi.org/10.17933/jppi.2019.090202)

IMPLEMENTED PSO-NBC AND PSO-SVM TO HELP DETERMINE STATUS OF VOLCANOES

IMPLEMENTASI PSO-NBC DAN PSO-SVM UNTUK MEMBANTU PENENTUAN STATUS GUNUNG BERAPI

Firman Tempola

Teknik Informatika Universitas Khairun Ternate
Jati , Ternate Selatan, Kota Ternate Maluku Utara 97716, Indonesia

Firman.tempola@unkhair.ac.id

Naskah Diterima: 2 February 2018; Direvisi : 26 November 2019; Disetujui : 4 December 2019

Abstract

This research is a continuation of previous research that applied the Naive Bayes classifier algorithm to predict the status of volcanoes in Indonesia based on seismic factors. There are five attributes used in predicting the status of volcanoes, namely the status of the normal, standby and alerts. The results Showed the accuracy of the resulted prediction was only 79.31%, or fell into fair classification. To overcome these weaknesses and in order to increase accuracy, optimization is done by giving criteria or attribute weights using particle swarm optimization. This research compared the optimization of Naive Bayes algorithm to vector machine support using particle swarm optimization. The research found improvement on system after application of PSO-NBC to that of 91.3 % and 92.86% after applying PSO-SVM.

Keywords: Naive Bayes, Support Vector Machine, Particle Swarm Optimization, Volcanoes,

INTRODUCTION

Indoensia is one of the countries in the world with the most volcanoes. The fact makes Indonesia known as the ring of fire. There are at least 127 active volcanoes in Indonesia and only 64 are monitored. Indonesia experiences volcanic eruptions every year, as it is reported on various mainstream media such as television, newspaper, or electronic media.

Currently, the use of machine learning to solve problems have seen an increasing trends in various fields. The field of natural disasters was no exception. In volcanic disasters data related to volcanic eruptions, in particular, the Indonesian government has a Center for Vulcanology and Geological Hazard Mitigation (PVMBG) whose task is to carry out research, investigation, engineering and provide services in the field of volcanology and geological hazard mitigation. PVMBG publishes recommendations on the status of volcanic activity based on data that is monitored from each volcano.

In determining the status of volcanoes, the Center for Volcanology and Geological Mitigation Mitigation (PVMBG) monitors the volcanic activities in two approaches, namely visual observations and seismic factors. Publications related to volcanic disasters have also been published in several previous researchers (Pratomo, 2006), (Reath, 2017) (Tempola, 2018).

This research is a continuation of the previous research (Tempola, 2018) in which Naive Bayes Classfier algorithm was used. In this study, a different classification algorithm was applied, namely the support vector machine. Diffetent algorithm was used considering the level of accuracy obtained in the previous research was still at the level of fair classification (Gorunescu, 2011). To improve accuracy on Naive Bayes optimization, more weight to the criteria or attributes (Muhammad H, et al 2017) on the Naive Bayes classification algorithm are given, along with the implementation of support vector machine.

Classification algorithm optimization was conducted to improve system accuracy as done by Kumar, et al in 2017, where the optimization algorithm performed was particle swarm optimization (PSO). Likewise, Agustina (2018) and Idrus, et al (2018) have conducted the Naive Bayes optimization with Particle Swarm Optimization, resulted in a difference of accuracy between those using only Naive Bayes and Naive Bayes optimization with PSO. When optimized with PSO, the system accuracy increased. This research will optimize the classification algorithm of Naive Bayes and Sortport Vector Machines with particle swarm optimization to help determine the status of volcanoes in Indonesia.

METHODOLOGY

This research is experimental, namely by optimizing the Naive Bayes classifier classification algorithm and support vector machine with particle swarm optimization on volcanic activity data in Indonesia, with the aim to improve system accuracy.

Dataset

Data processed in the classification using Naive Bayes and support vector machines are public data which can be accessed openly by anyone at the official website of Energy and Mineral Resources Ministry's PVMBG. Data as shown in Table 1, are then processed and tested on a system that has been developed.

Table 1. Dataset of Volcanic Activity

No	SV	DT	IV	EB	PS	RS
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	S Alert
67	78	135	17	0	Normal	Standby
68	1	13	5	0	Standby	Normal
69	17	25	5	157	Alert	Alert

Descriptions:

SV = Shallow Volcanic

DT = Deep Tectonics

IV = in Volcanic

EB = Earthquake Blowing

PS = Previous Status

RS = Recommended status

Machine Learning

Machine learning is a part of artificial intelligence that works by requiring data. Without data, no output can be produced (Harrington, 2012). In machine learning, several algorithms are grouped. One of them is the classification algorithm. Applied in this study are the classification algorithm Support vector machine and Naive Bayes classifier.

Naive Bayes Classification Algorithm

Naive Bayesia Classifier (NBC) is one of the methods in machine learning, which is based on the Bayes theorem. NBC is a classification algorithm that is very effective (in acquiring the right results) and efficient (the reasoning process is done by utilizing existing *inputs* in a relatively fast manner). Another advantage of NBC is that it can handle both numerical and discrete data. Naive bayes works by calculating the set of probabilities of each attribute by adding up the frequency and combination of values from the given dataset. The general form of the Bayes theorem is shown in Equation 1.

$$P(H|X) = P(X|HP(H)\frac{p\ (H)}{P(X)} \qquad (1)$$

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 (Posteriory probability)

P (H) = Probability of H hypothesis (Prior Probability)

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

$$P(A_i|C_j = \frac{|A_{ij}|}{N_{cj}} \qquad (2)$$

Where $|A_{ij}|$ is the number of training examples of class A_i that receives the value of Cj. While N_{cj} is 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.

$$F(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad (3)$$

SVM Classification Algorithm

Support vector machine (SVM) is a relatively new technique (1995) for prediction, classification and regression, which has been very popular in recent decades. Support vector machine (SVM) is the *best stock classifier* . SVM has good decision making for data points outside the training set. The classification method that is now being developed and implemented is *support vector machine.* This method is rooted in statistical learning theory whose results are very promising. SVM is divided into two linear and non-linear SVM models (Machine learning in action) . The concept of SVM can be explained simply as an attempt to find the best hyperplane that functions as a separator of two classes in the input space The classification problem can be translated by trying to find a line (hyperplane) that separates the two dataset using SVM multiclass.

Particle Swarm Optimization

The PSO method was introduced by Doctors Kennedy and Elbert in 1995 based on research conducted on the behavior of flocks of birds and fish and is a *global heuristic* optimization method. PSO is a population-based iterative algorithm. The population consists of many particles, which are initialized with a population of random solutions and used to solve optimization problems (Abraham, Grosan, & Ramos, 2006). Each particle represents the candidate's solution and moves toward the optimal position by changing its position according to the speed of the particle flying through the search space with a dynamic speed adjusted for historical behavior. Therefore, particles have a tendency to fly toward better and better search areas during the search process (Abraham et al., 2006).

PSO algorithm, the search for solutions is carried out by a population consisting of several particles. The population is generated randomly with the smallest and largest value limits. Each particle represents the position or solution of the problem at hand. Each particle searches for an optimal solution by crossing the search space. This is done by adjustments made by each particle to the best position of the particle (local best) and to the best particle position of the whole herd (global best) while crossing the search space. Thus, the spread of experience or information takes place inside the particle itself and between a particle and the best particles of the whole herd during the process of finding a solution. After that, the search process is carried out to find the best position of each particle in a certain number of iterations to obtain a relatively steady position or reach a predetermined iteration limit. At each iteration, each solution is represented by the position of the particle, its performance is evaluated by inserting the solution into the fitness function (Budi Santosa and Pauly Willy, 2011) . 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:

$$x = x_{min} + rand[0,1]x(x_{max} - x_{min}) \quad (4)$$

   - 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.

$$v_{i,j}^{t+1} = w.v_{i,j}^{t} + c_1.r_1\left(Pbest_{i,j}^{t} - x_{i,j}^{t}\right) + c_2.r_2$$
(5)

c. Update position and calculate fitnes: To update the position, using Equation 6.

$$x_{i,j}^{t+1} = x_{i,j}^{t} + v_{i,j}^{t+1} \quad (6)$$

(6)

d. Update pBest and gBest
   Do 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 fitnes value will be the new gBest.

Measurement Performance system

For the measurement of classification performance is by comparing all test data that are classified correctly with the number of test data. Equation 7 is a model used to measure classification performance.

$$Performance = \frac{\sum predict\ true}{\sum\ data\ testing}\ x\ 100\% \quad (7)$$

RESULTS AND DISCUSSION

This study compares two classification algorithms, namely Naive Bayes and support vector machines which are optimized using particle swarm optimazation to help determine the status of volcanoes in Indonesia. Before doing optimization with particle swarm optimization, the initial stage was to divide the dataset into two groups, namely training data and testing data. And then, the test phase was divided into two tests, namely testing with Naive Bayes optimization using PSO (PSO-NBC) and testing with optimizing support vector machine with particle swarm optimizition (PSO-SVM). Each test was calculated to obtain the accuracy of each system and followed by data validation using k-fold cross validation and comparing the results of its accuracy. As the test model is shown in Figure 2 .

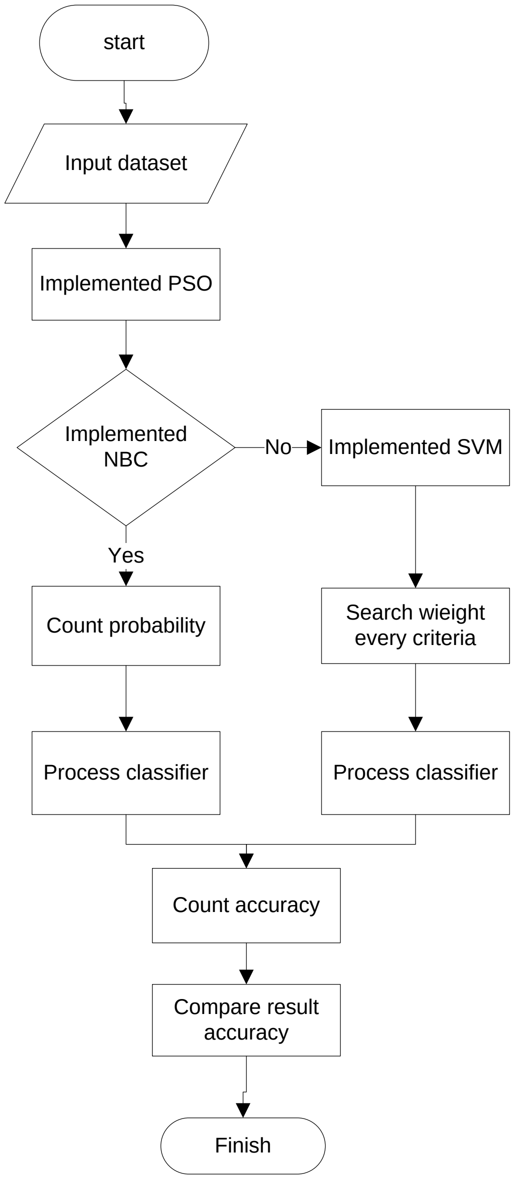

Figure 1. steps of testing model

Result of PSO-NBC testing

The results of testing by optimizing Naive Bayes by using PSO with the number of particles initialized as much as 10 and the weight range that is [0 1] obtained a system accuracy of 73.91% with the final weight of each criterion namely shallow volcanic = 0.572, distant tectonic = 0.152, volcanic in = 1, earthquake gusts = 1, the previous status = 1. Figure 2 is the final result b o bot each of the criteria in the first test. Likewise in the first test, in the second test with the number of particle at 20, the third test with the number of particle at 30, the fourth test with the number of particle at 40, the fifth test with the number of particle at 50 the system accuracy obtained was 73.91% which is only the final weight of each criterion.

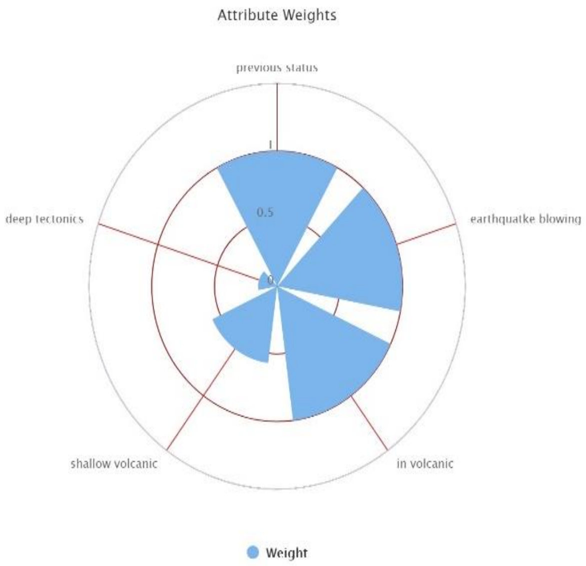

Figure 2. final result for attribute weights

Different results obtained in the sixth test with particle initialized at 10, 20, 30 and 40 with a weight range of [0 2]. The obtained system accuracy was equal to 91,3% with the final results of each criterion weights shown in Table 2.

Table 2. Final Results of Each Criterion Weight

Particle	weights				
Amount	X1	X2	X3	X4	X5
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
30	1,333	0,44	2	1,729	2
50	1,001	0,33	1,882	2	2
50	0,837	0,63	2,985	1,842	2,941
50	0,395	2	0	0,154	0

Result Of testing PSO-SVM

The second testing model is similar to the one completed on the first test. Namely by initializing different particles and weight ranges. The obtained final weight of each attribute is shallow volcanic = 0.572, distant tectonic = 0.152, deep volcanic = 1, earthquake gusts = 1, previous status = 1. While the highest accuracy obtained was 92.86%. This is different from testing without using PSO-SVM where the test results only reached 83.3%.

Based on the results of system testing conducted using the two methods of system accuracy, acquired system accuracy was above 90% or can be categorized as excelent classification, different from without optimization which was only categorized as fair classfication. In addition, it can also be seen that the PSO-SVM method is still better than the PSO-NBC on volcanic activity data. As shown in Figure 3 As also found in other studies.

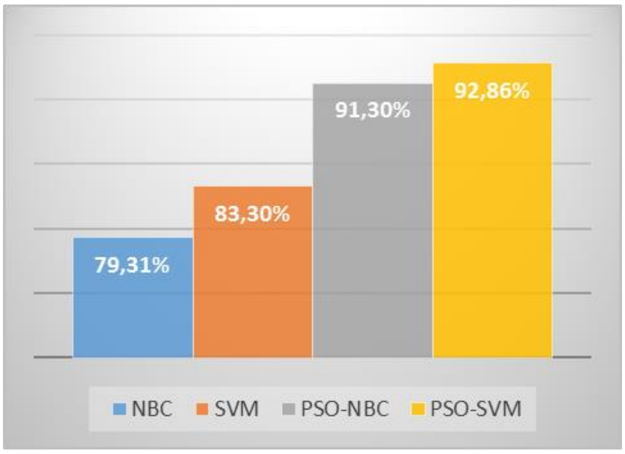

Figure. 3 accuracy comparison

CONCLUSIONS

Based on the results of testing in the comparison of Naive Bayes optimization systems with support vector machines using the particle swarm optimazition it can be concluded, that the use of Naive Bayes optimization with PSO can improve system accuracy. in addition, there are differences in the results of the system accuracy between naive bayes and support vector machines where the results of system accuracy in volcanic activity data when support vector machines **was used with** PSO optimization algorithms are better than the Naive Bayes classifier.

REFERENCES

Abraham, A., Grosan, C., & Ramos, V. (2006). *Swarm Intelegence in Data Mining*. London: Verlag Berlin Heidelberg, Springer.

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.

Gorunescu F, (2011) *Data Mining: Concepts, Models and Techniques*. London: Springer, 2011.

Harrington, P., (2012). *Machine Learning in Action*. USA: Manning Publication.

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.

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.

Muhamad H *et al.*, 2017 "Optimization Naïve Bayes Classifier using Particle swarm optimization," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, pp. 180–184

Pratomo, I. (2006). Klasifikasi Gunung Api di Indonesia, Studi Kasus Dari Beberapa Letusan Gunung Api Dalam Sejarah. *Jurnal Geologi Indonesia*, *vol. 1, No. 4, Desember 2006,* 2009-227.

Reath, K. A., & et al. (2016). Predicting eruptions from precursory activity using remote sensing data hybridization. *Journal of Volcanology and Geothermal Research*. *Vol. 321*, 18-30. diakses dari [https://www.sciencedirect.com.ezproxy.ugm.ac.id/science/article/pii/S0377027316300695] tanggal 3 Februari 2018

Santosa B and Willy P, (2011) *Method metaheuristic, concept and Implementation*. Yogyakarta: Graha Ilmu.

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

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" *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, pp. 577–584.