Indonesian Physical Review

Volume 6 Issue 1, January 2023

P-ISSN: 2615-1278, E-ISSN: 2614-7904

Application of Support Vector Machine on Drought Code Classification in North Sumatra Indonesia

Kartika Dewi Butar-Butar[1*], Poltak Sihombing[2*], Tulus[3 *]

[1] *Master of Informatics Program, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia.*

[2] *Department of Computer Science, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara.*

[3] *Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Sumatera Utara*

Corresponding Authors E-mail: poltak@usu.ac.id , dewi90kartika@gmail.com, tulus@usu.ac.id

Article Info

Article info:

Received: 28-10-2022

Revised: 16-12-2022

Accepted: 19-12-2022

Keywords:

SVM Method; Clasification; CFFDRS; FWI; Drought Code

How To Cite:

K. D. Butar-Butar, P. Sihombing and Tulus. "Application of Support Vector Machine on Drought Code Classification in North Sumatra Indonesia", Indonesian Physical Review, vol. 6, no. 1, p 51-59, 2023

DOI:

https://doi.org/10.29303/ipr.v6i1.196.

Abstract

This study aims to classify the Drought Code in North Sumatra. Drought Code is part of Fire Danger Rating System (FDRS) developed in Canada. One of the sub-systems of FDRS is Fire Weather Index (FWI) that aims to evaluate fire hazards from current and past weather conditions. Drought Code is classified by using Support Vector Machine. Support Vector Machine is widely used in the data classification process. One of the advantages of Support Vector Machine methods is it has ability in classifying large amount of data and classifying more than two classes or multi-classes. Weather parameters used in this study are rainfall and temperature in North Sumatra. The data used are from 8 (eight) meteorological observation stations in North Sumatra from 2017 to 2021. Drought Code is carried out with several tests using several kernels contained in SVM.

Introduction

Drought Code (DC) is one of the parameters of the Fire Weather Index. The Fire Weather Index (FWI) is one of the subsystems of the Fire Danger Rating System (FDRS) which was first developed in Canada [1]. FWI aims to evaluate the fire hazard from current and past weather conditions [2]. The FWI system depends on weather factors consisting of air temperature, relative humidity, rainfall and wind speed. FWI includes three humidity codes namely Fine Fuel Moisture Code (FFMC), Drought Moisture Code (DMC), Drought Code (DC). FWI produces three output indexes, namely Initial Spread Index (ISI), Build Up Index (BUI) and Fire Weather Index (FWI) [3]. The FWI structure can be seen in Figure 1.

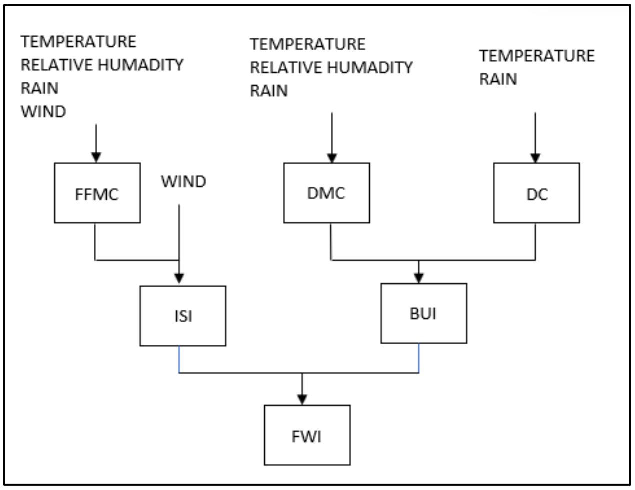

Figure 1. Fire Weather Index Structure

DC is a numerical level of water content contained in the organic layer below the soil surface. DC is used as an indicator of the level of potential ease of occurrence of forest fires in terms of weather parameters [4]. The highest DC ratings occur more frequently in peatland fires. DC represents the water content of dense organic soil with a medium depth of about 18 cm [5]. n determining the DC value requires two weather parameters, namely rainfall and air temperature. DC has four class classifications namely low, moderate, high and extreme [6]. Classification is a technique in the field of data mining that is used in making predictive models to produce new data [7]. The Support Vector Machine (SVM) method developed by Boser, Guyon, and Vapnik can be used to classify DC ratings.

SVM calculations focus on how to find the best hyperplane of two or more classes. Margin is the distance between the hyperplane and the closest data to each class. The nearest data is referred to as a support vector. SVM has advantages in linear and nonlinear classification problems [8]. In the Linear SVM classifier problem used on data that can be separated linearly or can be classified into two classes using a single straight line [9]. Whereas in nonlinear classification problems the data cannot be separated using a single straight line, so SVM requires a kernel to carry out the classification [10]. SVM kernel types are Linear, Polynomial, Sigmoid and Radial Basis Function (RBF) [11].

The application of the SVM method in the DC classification in North Sumatra can produce the best SVM model, in which DC classification predictions can be made using the SVM model. SVM model accuracy is calculated based on training data and test data. The best SVM model in predicting DC classification, can be used as an early warning to the public to be more vigilant at locations that have a high DC rating or in a high to extreme class classification.

Method

Data used

The data used was obtained from a data request at the Medan BMKG, namely rainfall and air temperature from January 2017 to December 2021. The data used was 14.068 data. The following is a list of observation stations used in this study which can be seen in Table 1.

Table 1. List of Observation Stations

No	Station Name	Code
1	BMKG Medan	bw1
2	Kualanamu Meteorological Station	kno
3	Deli Serdang Geophysical Station	tsi
4	Belawan Marine Meteorological Station	blw
5	Deli Serdang Climatology Station	kds
6	Aek Godang Meteorological Station	agd
7	F.L Tobing Meteorological Station	sba
8	Binaka Gunungsitoli Meteorological Station	gsi

The observation station used is an observation station located in North Sumatra. North Sumatra Province is located at 1° - 4° North Latitude and 98° - 100° East Longitude. The land area of North Sumatra Province is 72,981.23 km². In North Sumatra, each region has a different topography, so it has different weather conditions [12]. Each observation station represents the area being researched, such as bw1 station representing urban areas, blw and kno stations representing east coast areas, tsi stations representing highland areas, sba stations representing west coast areas, gsi stations representing nias islands, kds stations representing plains areas low and agd stations represent relatively mountainous and steep areas.

Rainfall and air temperature data that have been collected will be calculated using the basic DC calculation [13]. The following equation is used to get the DC value.

$$Rd = 0.83 * ro - 1.27 \tag{1}$$

$$Qo = 800 \exp\left(\frac{-Do}{400}\right) \tag{2}$$

$$Qr = Qo + 3.937 * Rd \tag{3}$$

$$Dr = 400 \ln\left(\frac{800}{Qr}\right) \tag{4}$$

$$V = 0.36(T + 2.8) + lf \tag{5}$$

$$DC = Do(orDr) + 0.5V \tag{6}$$

Description :

ro	: accumulated rainfall in one day, where ro > 2,8
Rd	: effective rainfall
D_0/D_T	: previous day's DC
Q_0	: moisture equivalent of previous days's DC
Q_r	: moisture equivalent after rain
Lf	: adjusted day length
V	: evapotranspiration
T	: air temperature
DC	: drought code value

Lf (adjusted day length) value is different for each observation station. Lf value is calculated from the monthly mean value of evapotranspiration and the monthly average value of the maximum air temperature. Lf value of the observation station used in this study can be seen in Table 2.

Table 2. Lf Value of Observation Station

Month	SBA	BW1	BLW	GSI	TSI	KDS	AGD	KNO
January	4,0	2,2	2,8	4,1	2,2	3,2	4,3	3,6
February	4,0	2,7	2,8	4,0	2,5	3,4	4,7	4,2
March	4,2	2,4	3,1	4,3	2,6	3,3	5,0	4,1
April	4,0	2,3	2,5	3,8	2,9	3,1	4,7	3,8
May	4,0	2,1	2,7	3,4	2,1	3,0	4,5	3,7
June	3,9	2,3	2,6	3,8	2,1	3,6	5,1	3,8
July	4,2	2,4	2,8	4,3	2,3	4,0	5,5	4,2
August	3,7	2,2	2,6	4,2	2,3	3,7	5,2	4,0
September	4,0	1,8	2,2	4,0	2,3	3,9	4,6	3,8
October	3,7	2,2	2,4	4,0	2,0	3,5	5,3	3,8
November	4,3	2,0	2,3	3,2	2,3	3,2	4,5	3,7
December	4,0	2,4	2,5	4,0	2,3	3,1	4,3	3,8

The basic calculation of DC using equations 1 to 6 will produce a DC value. The DC value is classified according to the class based on the DC value obtained. Class classification in DC can be seen in Table 3.

Table 3. DC Class Classification

Class	Range	Description
Low	< 200	Wet conditions, smoke haze does not occur.
Moderate	200 – 300	Under normal conditions, combustion should be monitored.
High	300 – 400	In peak dry season conditions, burning should be monitored.
Extreme	>400	Drought hazard conditions, burning completely prohibited, potential for haze.

Proposed model of Support Vector Machine (SVM)

This study uses the SVM method with nonlinear classification, so it uses kernels such as linear, polynomial, sigmoid and RBF kernels so as to get the best SVM model. The commonly used kernel functions is stated in Table 4:

Table 4. SVM Kernels

Kernel	Formula
Linear	$K(x_i, x_j) = x_i^T x_j$
Polynomial	$K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, > 0$
Radial Basis Radial (RBF)	$K(x_i, x_j) = exp(-\gamma \lVert x_i - x_j{}^2, > 0$
Sigmoid	$K(x_i, x_j) = tanh(\gamma x_i^T x_j + r)$

Linear kernels are usually used in linear data classification which is a simple kernel function, polynomial kernels are used in classifications which usually have normalized training datasets, sigmoid kernels are generally used for neural networks and radial basis function (rbf) kernels are used in mapping samples into high dimensions. Testing the SVM method uses training data and test data, then calculates the accuracy value of the SVM method. The evaluation used in the SVM method is accuracy, precision, and recall using the confusion matrix [14]. The range of accuracy, precision and recall in terms of values is from 1 to 100%. Here are the equations for calculating accuracy, precision and recall.

$$Accuracy = \frac{TP}{TN+TP+FN+FP} \text{x } 100\% \quad (7)$$

$$Precision = \frac{TP}{TP+FP} \text{x } 100\% \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \text{x } 100\% \quad (9)$$

Description:
TP : True Positive
FP : False Positive
FN : False Negative
TN: True Negative

Result and Discussion

According to basic calculation of Drought Code, the weather parameters used are rainfall and air temperature will produce the Drought Code value along with its category classification in csv form Table 5 is the result of the DC calculation.

Table 5. The result of the DC calculation

year	mth	day	temp	prcp	dc	category
2017	1	1	26	51	13	low
2017	1	2	26	49	13	low
2017	1	3	25	54	13	low
2017	1	4	25	63	14	low
2017	1	5	25	33	15	low
2017	1	6	25	12	15	low
2017	1	7	24	58	13	low
2017	1	8	26	0	12	low
2017	1	9	26	0	10	low
2017	1	10	26	0	11	low
2017	1	11	26	0	9	low
2017	1	12	26	11	10	low
2017	1	13	26	15	12	low
2017	1	14	27	1	15	low
2017	1	15	25	2	16	low
2017	1	16	25	0	23	low
2017	1	17	26	0	23	low
2017	1	18	27	2	22	low
2017	1	19	23	35	29	low
2017	1	20	25	3	28	low
⋮						
2021	12	30	26	0	197	Low
2021	12	31	25	1	200	Moderate

The results of these calculations are used as test data with the kernel contained in SVM, namely linear, polynomial, sigmoid and also RBF. The testing process is divided into 3 division models for training data and testing data. With the aim of knowing the best model that produces the best accuracy, precision and recall values. Before conducting the test, the data is divided into two parts, namely training data and testing data. training data is used to train and develop the model and also compare the performance of different models such as using the kernel on SVM. If data testing is used to check whether the model used is working correctly. Model testing is done by dividing the training data by 80% and testing data by 20%. then carry out tests with 70% training data and 30% testing data and finally carry out tests by separating 60% training data and 40% testing data. Table 6 shows the results of the accuracy of testing the SVM model with several kernels and the data distribution used.

Table 6. SVM Kernel Accuracy Value

Kernel	Linear	Polynomial	Sigmoid	RBF
80% training data and 20% testing data	100%	99%	90%	98%
70% training data and 30% testing data	100%	98%	90%	98%
60% training data and 40% testing data	100%	98%	88%	98%

Table 6 shows the percentage of accuracy values by testing several kernels on SVM. The linear kernel is constantly has accuracy value of 100%. This explains that the linear kernel can perform data distribution well during the data mapping process. While in the polynomial, sigmoid and RBF kernels, there is variety in the accuracy value according to the data split. Evaluation of the SVM model also calculates the precision and recall values of each test. The values of precision and recall assessed in each category of DC are shown in Figure 2 and Figure 3.

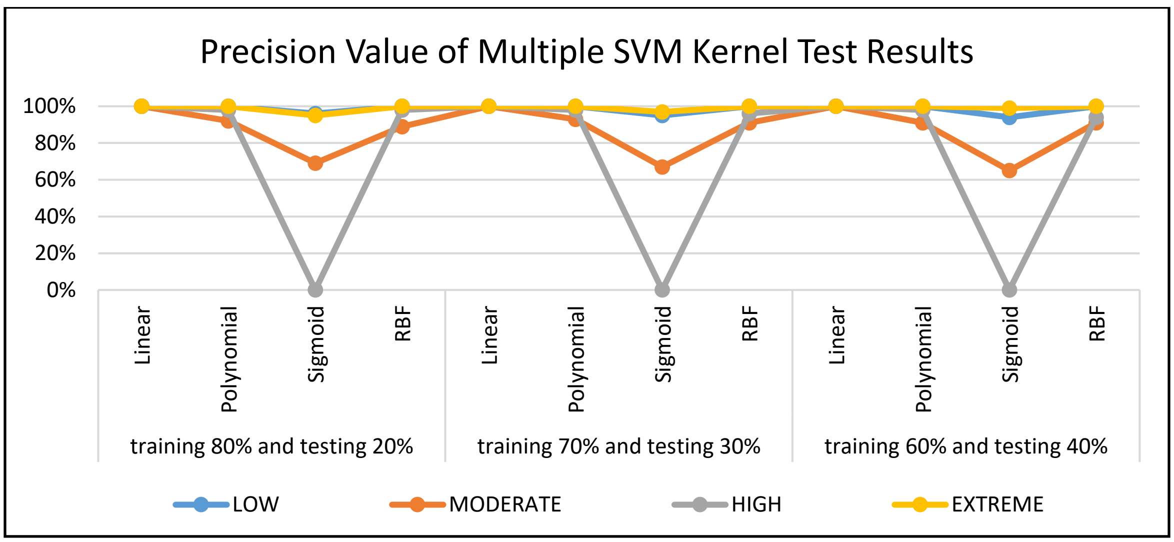

Figure 2. Precision Value Chart

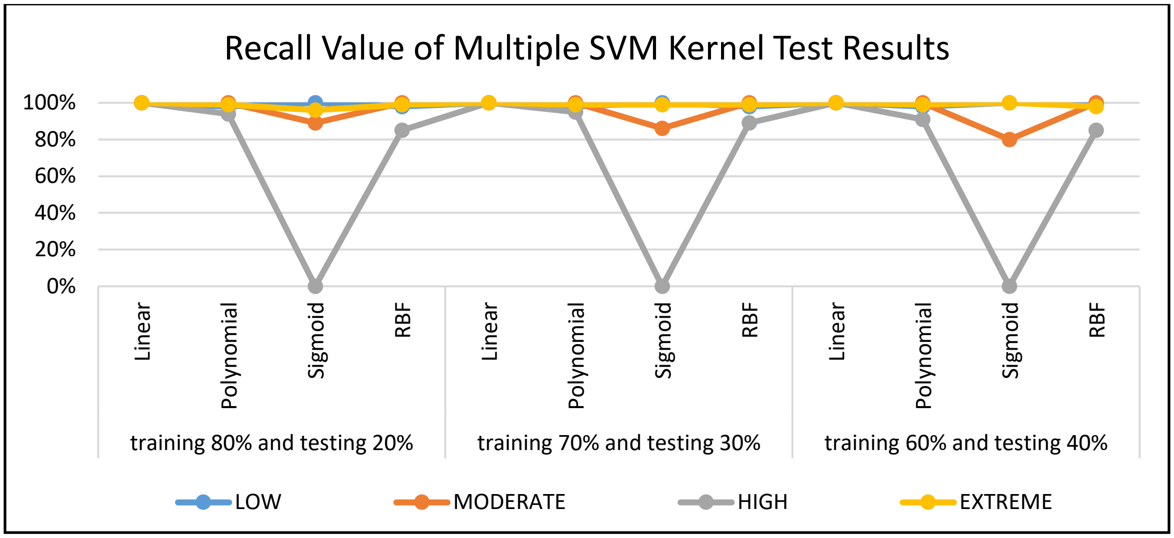

Figure 3. Recall Value Chart

After obtaining the accuracy value by dividing the SVM model, it can be seen in Figure 2 and Figure 3 that the precision and recall values using the linear kernel have a value of 100%. This shows that solving problems in the data classification process in this study is very good using linear kernels compared to sigmoid kernels. sigmoid kernel is not suitable for use in DC classification because sigmoid kernel is usually used in artificial neural network problems.

Conclusion

The DC classification in this study was carried out in North Sumatra by adjusting the corrected day length values at each meteorological observation station and conducting tests with the SVM kernel. From the results of the tests that carried out with the distribution of 80% training data : 20% testing data, 70% training data : 30% testing data, and 60% training data : 40% testing data, the highest accuracy results are using the linear kernel and the lowest is the sigmoid kernel. The value of precision and recall by performing tests, the highest value in each category uses a linear kernel (100%), while the value of precision and recall using the sigmoid kernel in the high category is 0%. The results of the highest accuracy, precision and recall are in the linear kernel, because the data used in the test is already separated linearly or already has a relationship between one variable and another. While the results of accuracy, precision and recall on the sigmoid kernel are lower, because the sigmoid calculation concept is the hyperbolic tangent which is not quite right for the data being tested.

References

[1] R. H. Ziel, P. A. Bieniek, U. S. Bhatt, H. Strader, T. S. Rupp, and A. York, “A comparison of fire weather indices with MODIS fire days for the natural regions of alaska,” *Forests*, vol. 11, no. 5, 2020, doi: 10.3390/F11050516.

[2] Suciarti, “Sistem Informasi Tingkat Bahaya Kebakaran Hutan dan Lahan Dengan Menggunakan Fire Weather Index (FWI) dan SIG Arcview,” *J. Tek. Elektro Univ. Tanjungpura*, vol. 1, no. 1, pp. 1–9, 2013.

[3] B. D. Amiro *et al.*, "Fire weather index system components for large fires in the Canadian boreal forest," *Int. J. Wildl. Fire*, vol. 13, no. 4, pp. 391–400, 2004, doi: 10.1071/WF03066.

[4] D. M. Midyanti, "Combination of SOM-RBF for drought code prediction using rainfall and air temperature data," *J. Teknol. dan Sist. Komput.*, vol. 8, no. 1, pp. 64–68, 2020, doi: 10.14710/jtsiskom.8.1.2020.64-68.

[5] Y. Yang, M. Uddstrom, G. Pearce, and M. Revell, "Reformulation of the drought code in the Canadian Fire Weather Index System implemented in New Zealand," *J. Appl. Meteorol. Climatol.*, vol. 54, no. 7, pp. 1523–1537, 2015, doi: 10.1175/JAMC-D-14-0090.1.

[6] Varela, "Fire Weather Index (FWI) classification for fire danger assessment applied in Greece," *Tethys, J. Weather Clim. West. Mediterrania*, no. 1994, pp. 31–40, 2015, doi: 10.3369/tethys.2018.15.03.

[7] E. Saputra, U. Khaira, and Z. Abidin, "Implementasi Algoritme Support Vector Machines untuk Klasifikasi Area Terbakar di Lahan Gambut (Implementation of Support Vector Machines Algorithm to Classify Burn Area of Peat)," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 1, pp. 19–24, 2021.

[8] H. Harafani, "Optimasi Parameter Pada Metode Support Vector Machine Berbasis Algoritma Genetika untuk Estimasi Kebakaran Hutan," *J. Intell. Syst.*, vol. 1, no. 2, pp. 82–90, 2015.

[9] E. E. Pratama and B. R. Trilaksono, "Klasifikasi Topik Keluhan Pelanggan Berdasarkan Tweet dengan Menggunakan Penggabungan Feature Hasil Ekstraksi pada Metode Support Vector Machine (SVM)," *J. Edukasi dan Penelit. Inform.*, vol. 1, no. 2, 2015, doi: 10.26418/jp.v1i2.11023.

[10] W. Purnami, A. M. Regresi, and L. Ordinal, "Perbandingan Klasifikasi Tingkat Keganasan Breast Cancer Dengan Menggunakan Regresi Logistik Ordinal Dan Support Vector Machine (SVM)," *J. Sains Dan Seni Its*, vol. 1, no. 1, 2012.

[11] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, "A Practical Guide to Support Vector Classification," *BJU Int.*, vol. 101, no. 1, pp. 1396–1400, 2008.

[12] Y. Darmawan, H. H. Hsu, and J. Y. Yu, "Article characteristics of large-scale circulation affecting the inter-annual precipitation variability in northern sumatra island during boreal summer," *Atmosphere (Basel).*, vol. 12, no. 2, pp. 1–20, 2021, doi: 10.3390/atmos12020136.

[13] C. E. Van Wagner, P. Forest, and E. Station, "Structure of the Canadian," *Can. For. Serv. Publ.*, no. 1333, 1974.

[14] M. Awad and R. Khanna, "Efficient learning machines: Theories, concepts, and applications for engineers and system designers," *Effic. Learn. Mach. Theor. Concepts, Appl. Eng. Syst. Des.*, no. January, pp. 1–248, 2015, doi: 10.1007/978-1-4302-5990-9.