Parameter: Journal of Statistics
Available at https://bestjournal.untad.ac.id/index.php/parameter
Vol. 4 No. 1 2024, 23-29
DOI: *https://doi.org/10.22487/27765660.2024.v4.i1.17142*
eISSN: 2776-5660

PREDICTING DROUGHT IN EAST NUSA TENGGARA: A NOVEL APPROACH USING WAVELET FUZZY LOGIC AND SUPPORT VECTOR MACHINES

Hartayuni Sain[1*], Firda Fadri[2]

[1]Department of Mathematics, Universitas Tadulako, Palu, Indonesia
[2]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Jember, Indonesia

***e-mail**: [1*]*hartayuni.sain@untad.ac.id*, [2]*firdafadri@unej.ac.id*

ABSTRACT

The water crisis, or what is hereinafter referred to as drought, has become a systemic and crucial problem in several regions in Indonesia. Indonesia is an agricultural country, where the presence of water is very influential so that drought can become a natural disaster if it starts to cause an area to lose its source of income due to disturbances in agriculture and the ecosystem it causes. Drought forecasting can provide support solutions in preventing the impact of drought. In this paper, we compare the performance of wavelet fuzzy logic and the support vector machine (SVM) as a supervised learning method for drought forecasting in Indonesia. This study examines the monthly rainfall data for 1999-2015 which is the basis for calculating the drought index based on the Standardized Precipitation Index (SPI). The SPI value used is SPI-3 at a station in East Nusa Tenggara Province. The performance of models is compared based on R^2. The results showed that the R^2 value of SVM with 60:40 data partition was 65.07%, for 75:25 data partition the R^2 value was 76.54% and for 90:10 data partition the R^2 value was 79.57%. Meanwhile, the R^2 value of wavelet fuzzy logic is 53.6%. So it can be concluded that forecasting using SVM is better than wavelet fuzzy logic because the R^2 value for data partition 90:10 is 79.57%.

Keywords: *Drought forecasting, SPI, Wavelet Fuzzy Logic, Support Vector Machine.*

Cite: Sain, H., & Fadri, F. (2024). *Predicting Drought in East Nusa Tenggara: A Novel Approach Using Wavelet Fuzzy Logic and Support Vector Machines*. *Parameter: Journal of Statistics*, *4*(1), 23-29, https://doi.org/10.22487/27765660.2024.v4.i1.17142.

[*]Received: May 2024; Reviewed: May 2024; Published: Jun 2024

INTRODUCTION

Drought is a global environmental disaster that occurs when precipitation falls below a certain level for an extended length of time. The economy, agriculture, water resources, tourism, and environments are all negatively impacted [1][2][3]. In this sense, the most effective method for precise drought prediction needs to be discovered in order to reduce its negative consequences on the environment and wildlife. This study's goal is to demonstrate how artificial intelligence techniques are frequently applied to drought forecasting. The Standard Precipitation Index (SPI), which is produced by (McKee et al. 1993), is frequently used to anticipate drought within specific precipitation time periods. According to the SPI, positive values imply wet circumstances, while higher negative values suggest severe drought. higher negative values indicate moderate drought.

In recent years, many models of drought forecasting have been developed [5][6]. Cancelliere et al. (2005) used an SPI time series autocovariance matrix to forecast the values of the Standard Precipitation Index (SPI). Autoregressive integrated moving average (ARIMA) and multiplicative seasonal ARIMA were created by Mishra and Desai (2005) in order to use the SPI value to anticipate droughts. Zhang et al. (2019) compared the ability of the ARIMA, WNN, and SVM models to forecast the occurrence of drought using the SPEI index in Sanjiang Plain, China. In order to anticipate drought, Mokhtarzad et al. (2017) compared a number of techniques, including ANN, ANFIS, and SVM. SPI is the output parameter, and the following input parameters are used: temperature, humidity, and precipitation. Their results show that the SVM model offers more accurate values for forecasting, claiming that the SVM model is more precise than ANN and ANFIS. Neural Networks, Wavelet Neural Networks, and Support Vector Regression to Forecast Drought are used by Belayneh and Adamowski (2012). Their findings demonstrate that the combined wavelet neural network (WN) model is the best model for predicting the value of SPI over multiple waiting times in the Awash River Basin in Ethiopia..

In this paper, we compare the performance of wavelet fuzzy logic and the support vector machine (SVM) for drought forecasting in Indonesia. This study examines the monthly rainfall data for 1999-2015 which is the basis for calculating the drought index based on the Standardized Precipitation Index (SPI). The SPI value used is SPI-3 at a station in East Nusa Tenggara Province. The performance of models is compared based on R^2.

MATERIALS AND METHODS

Study Area and Data

Indonesia is an archipelagic country that stretches astronomically from 95° east longitude to 141° east longitude and is located between 6° north latitude to 11° south latitude. Besides, geographically, Indonesia is located between two continents and two oceans. This location makes Indonesia has three main climates, namely hot climate, monsoon climate and sea climate and has two seasons, namely summer and rainy season.

East Nusa Tenggara is one of the provinces in the drought emergency category. Almost every year this island-based province has to go through a long dry season. This research is focused on one of the stations in the province of East Nusa Tenggara, namely the Komodo meteorological station, West Manggarai. The National Center for Environmental Prediction–NOAA website, http://www.esrl.noaa.gov, provided monthly rainfall data for the years 1999 to 2015, which served as the secondary data source for this study. The data to be used are the monthly rainfall data (X_1), the SPI value (X_2), and the nature of the weather (Y). Table 1 shows the classes of SPI.

Table 1. SPI drought severity classes [17]

SPI Value	Class
SPI≥2.00	Extreme wet
1.5≤SPI≤1.99	Very wet
1≤SPI≤1.49	Moderate wet
-0.99≤SPI≤0.99	Near normal
-1.49≤SPI≤-1	Moderate dry
-1.99≤SPI≤-1.5	Severe dry
SPI≤-2	Extreme dry

Wavelet

Local analysis, which is capable of showing signal information that other signal analyzers do not have, such as trends, breakpoints, and similarities, is the main thing that can be done by wavelet analysis. Because of its ability to view data from different sides without showing quality deterioration, wavelets are able to simplify and reduce noise. Wavelet transformations are split into two types, namely Continuous Wavelet Transformation, which is useful in all real numbers for defined time-series data, and Discrete Wavelet Transformation, which is useful in the integer range for time series data[12].

The way Continuous Wavelet Transformation (CWT) works is to calculate the convolution of the signal with a modulation window on each desired scale at any time. The scale is related to frequency in this transformation scale and translation, while translation is the location of the modulation window as it is shifted along the time-related signal. CWT can be defined, mathematically, as follows:\

$$W_n(s) = \sum_{n'=0}^{N-1} x_{n'} \psi * \left[\frac{(n'-n)\delta t}{s}\right], \tag{1}$$

with
$W_n(s)$: wavelet transform
s : scale
δt : time spacing
n : $1, 2, \cdots, N$
n' : $0, 1, \cdots, N-1$
x : time series data before wavelet transfformation
$\psi *$: wavelet transformation, where * denotes complex conjugate
N : number of data

The primary distinction from the Fourier transform, which just employs the sinus function as the modulation window, is that the basic wavelet function can be tailored as needed to achieve the optimal transformation outcomes. Based on the information gathered during the decomposition process, the mother wavelet is employed in the wavelet transformation. This study made use of the Morlet mother wavelet.

$$\psi(\eta) = \pi^{-1/4} e^{i\omega_0 \eta} e^{-\eta^2/2}, \tag{2}$$

with
$\psi(\eta)$: wavelet function
ω_0 : for morlet 6
η : $\frac{t-1}{s}$
i : $1, 2, \cdots, N$

The next step after the decomposition is performed is the wavelet reconstruction with the function:

$$x_n = \frac{\delta j \delta t^{1/2}}{C_\delta \psi_0(0)} \sum_{j=0}^{J} \frac{\Re\{W_n(s_j)\}}{s_j^{1/2}}, \tag{3}$$

with
x_n : reconstructed data
C_δ : reconstruction actor (for morlet 0,776)
δj : factor for scale averaging (for morlet 0,6)
$\psi_0(0)$: for morlet $\pi - 1/4$
$W_n(s)$: wavelet transform
J : many scales
j : $0, 1, 2, \cdots, J$

Fuzzy Logic

Fuzzy logic was first developed by Lotfi Zadeh in 1965. Its core ideas were convexity, complement, union, intersection, inclusion, and connection. With fuzzy logic, human knowledge systems can be easily and effectively implemented in machine language. To describe its behavior, Fuzzy logic utilizes a set of rules. The Fuzzy set has a characteristic value function which includes actual numbers in the interval [13].

Small modifications can cause important differences in the use of the classical set, so the use of the classical set is considered unfair. To solve this problem, fuzzy sets are used. In a fuzzy set, based on how large their existence is in the set that can be seen from the membership value.

Support Vector Machine

SVM was initially used in 1995 to generate predictions for regression and classification models [14][18]. The SVM approach consistently finds the same answer while searching for the best possible global solution. SVM maps a high-dimensional space using training data in order to function [16][18]. In a high dimensional space, a classifier that can maximize the margin between two data classes will be looked for. By using this method, it is possible to determine which separator function among infinite functions is the best at separating two data sets from distinct classes (-1, +1) or at being the best hyperplane.

At its core, SVM is a linear classifier, or more specifically, classification cases that are separable linearly. Within the linear classification, SVM is divided into two categories: separable and non-separable categories. Given a set of $\boldsymbol{X} = \{x_1, x_2, \cdots, x_n\}$, with $\boldsymbol{x_i} \in \boldsymbol{R^n}$, $i = 1, \cdots, n$. With $y_i = +1$, else $y_i = -1$, so that the data pair becomes $(x_1,\ y_1)$, $(x_2,\ y_2)$, …, $(x_n,\ y_n)$ a training vector set of two groups to be categorized by SVM [15][16][18],

$$(\boldsymbol{x_i},\ y_i), \boldsymbol{x_i} \in \boldsymbol{R^n}, \qquad y_i \in \{-1, 1\}, \qquad i = 1, \cdots, n \tag{4}$$

A hyperplane separator is called $\boldsymbol{w}$, and *a* parameter specifies the relative plane location to a central coordinate *b* and shows the following:

$$(\boldsymbol{w}^T . \boldsymbol{x}) + \ b = 0 \tag{5}$$

It must satisfy this limitation when defining canonical-shaped hyperplane separation [15][16][8].

$$y_i\ [(\boldsymbol{w}^T . \boldsymbol{x_i}) + \ b] \geq 1, \ \ i = 1, 2, \cdots, n \tag{6}$$

when obtaining the ideal hyperplane by maximizing the $\frac{2}{\|w\|}$ margin or minimizing the following feature:

$$\boldsymbol{\Phi}(\boldsymbol{w}) \ = \ \frac{1}{2}\ \|\boldsymbol{w}\|^2 \tag{7}$$

The Langrange function, where is the Lagrange multiplier, is used to optimize. Primal space is Lagrange's function, which leads to the following outcomes.

$$\hat{\alpha} = \arg \begin{matrix} \min \\ \alpha \end{matrix}\ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (\boldsymbol{x_i^T} . \boldsymbol{x}_j) - \sum_{i=1}^{l} \alpha_i \tag{8}$$

with,

$$\alpha_i \geq 0, i = 1, 2, \cdots, n \ \text{ and } \ \sum_{i=1}^{n} \alpha_i y_i = 0 \tag{9}$$

thus the classification using the following formula [20],

$$f(x) = h\big(\hat{\boldsymbol{w}}^T . \boldsymbol{x} + \ \hat{b}\big) \tag{10}$$

where,

$$h(x) = \begin{cases} -1, & x < \ -1 \\ x, -1 \leq x \leq 1 \\ 1, & x > 1 \end{cases} \tag{11}$$

$$\boldsymbol{w} = \sum_{i=1}^{n} \hat{\alpha_i} y_i \boldsymbol{x}_i \qquad dan \qquad \hat{b} \ = \ -\frac{1}{2}\ \boldsymbol{w}(\boldsymbol{x}_r + \ \boldsymbol{x}_s) \tag{12}$$

For nonlinear data case, SVM classifies by solving the following optimization problems [16]:

$$\hat{\alpha} \ = \ \arg \begin{matrix} \min \\ \alpha \end{matrix}\ \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j\ \boldsymbol{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) - \sum_{i=1}^{l} \alpha_i \tag{13}$$

$\boldsymbol{K}\big(\boldsymbol{x_i}, \boldsymbol{x_j}\big)$ It is a kernel function and the classification equation for new data is:

$$f(\boldsymbol{x}) = sign\left(\sum_{SVs} \hat{\alpha_i}\ y_i \boldsymbol{K}(\boldsymbol{x_i}, \boldsymbol{x_j}) + \ \hat{b} \right) \tag{14}$$

where

$$\widehat{\boldsymbol{w}}.\boldsymbol{x} = \sum_{SVs} \widehat{\alpha_i}\, y_i \boldsymbol{K}(\boldsymbol{x_i}, \boldsymbol{x_j})$$
$$\hat{b} = -\frac{1}{2} \sum_{SVs} \widehat{\alpha_i}\, y_i [\boldsymbol{K}(\boldsymbol{x_r}, \boldsymbol{x_i}) + \boldsymbol{K}(\boldsymbol{x_s}, \boldsymbol{x_i})] \tag{15}$$

RESULTS AND DISCUSSION

Figure 1 shows the rainfall data utilized in the time series display. Figure 1 shows a pattern generated from monthly rainfall data plots over a 16-year period. Neither too dry nor excessively wet, Indonesia's typical weather has been close to normal for the past sixteen years. Based on Figure 1, Indonesia had through three periods of severe drought, however not as severe as one in August 2000, August 2004, and August 2006.

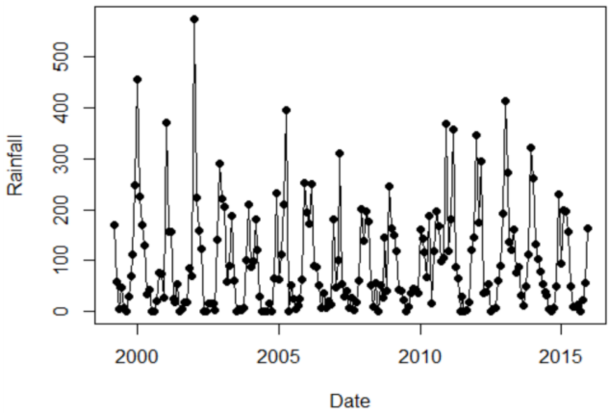
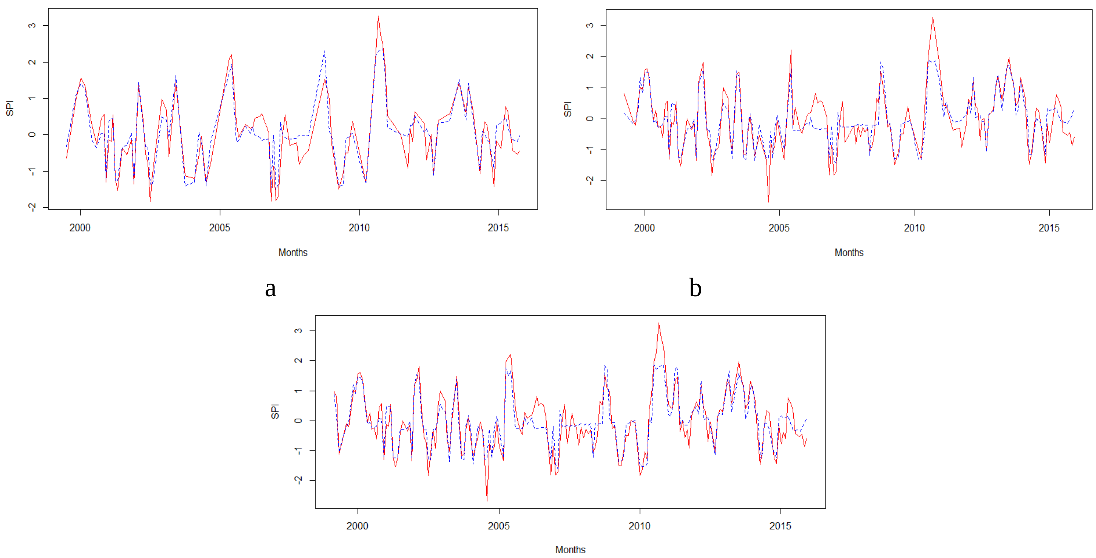


Figure 1. The plot of time series data.

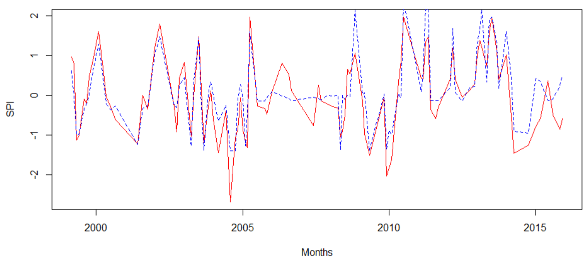
The data used in this study will be partitioned into several training data and test data with a ratio of 1. 60:40, 2. 75:25, and 3. 90:10. It can be seen in Figure 2 which shows the results of forecasting using the SVM method for training data and Figure 3 which shows the results of forecasting using the SVM method for testing data. In each figure, a shows the 60:40 data partition, b is the 75:25 data partition and c is the 90:10 data partition. In each figure, the actual data represented by the red line is compared with the predicted data represented by the blue line.



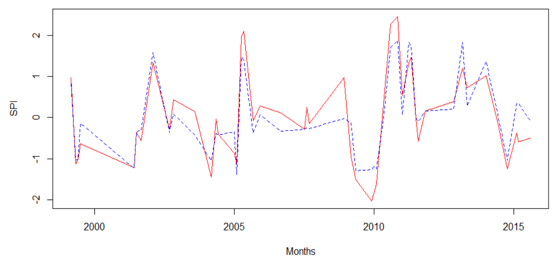
a b

c

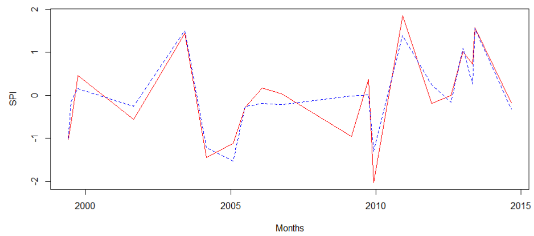
Figure 2. Forecast results using the SVM method for training data



a



b



c

Figure 3. Forecast results using the SVM method for testing data

Table 2. Test results.

The models	**Num. Of data (%)**	**Training Data**			**Testing Data**		
		MSE	RMSE	R^2	MSE	RMSE	R^2
SVM	60:40	0.1617	0.4020	0.8048	0.2728	0.5223	0.6507
	75:25	0.1997	0.4469	0.6937	0.2011	0.4484	0.7654
	90:10	0.1876	0.4331	0.7348	0.1519	0.3898	0.7957
WF	-	-	-	0.1130	-	-	0.5360

From Table 2 it can be seen that the R2 value of SVM with 60:40 data partition is 65.07%, for data partition 75:25 the R2 value is 76.54% and for data partition 90:10 the R2 value is 79.57%. Meanwhile, the R2 value of wavelet fuzzy logic is 53.6%. So it can be concluded that forecasting using SVM is better than wavelet fuzzy logic because the R2 value for data partition 90:10 is 79.57%.

CONCLUSIONS

The results of the research conducted showed lower errors and more predictive accuracy of the two methods being compared. The SVM method shows that flexibility and higher accuracy compared to other methods can be used in simulation and forecasting. This is indicated by a higher R^2 value using the 90:10 data partition. The results showed that R^2 of wavelet fuzzy logic is smaller than one of SVM. It can be concluded that SVM is better than the wavelet fuzzy logic for forecasting SPI value of drought in Indonesia. These results indicate that machine learning can be successful by using more training data, so that further research can be carried out using even larger data. Future studies should try to examine the wavelet fuzzy and SVM method and use these new techniques to explore SPI forecasts in other regions with different characteristics. Future research should also try to measure time-shift errors as it is part of forecasting issues with the regression model.

REFERENCES

Belayneh, A., & Adamowski, J. (2012). *Standard Precipitation Index Drought Forecasting Using Neural Networks, Wavelet Neural Networks, and Support Vector Regression. Applied Computational Intelligence and Soft Computing*, Volume 2012, Article ID 794061. DOI: 10.1155/2012/794061.

Cancelliere, A., Di Mauro, G., Bonaccorso, B., & Rossi, G. (2005). *Stochastic forecasting of Standardized Precipitation Index*. Proc. 31st IAHR Congress, Seoul, Korea, IAHR, 3252–3260.

Dai, A. (2013). *Increasing Drought Under Global Warming In Observations And Models.* Nat Clim Change, 3(1), 52–58.

Elkharrim, M., & Bahi, L. (2014). *Using statistical downscaling of GCM simulations to assess climate change impacts on drought conditions in the northwest of Morocco*. Modern Applied Science, 9(2).

Gunn, S. (1998). *Support Vector Machines for Classification and Regression*. ISIS Technical Report, Image Speech and Intelligent System Group, Southampton: University of Southampton.

Kusumadewi, S., Hartati, S., Harjoko, A., & Wardoyo, R. (2006). *Fuzzy Multi-Attribute Decision Making (Fuzzy MADM)*. Yogyakarta: Graha Ilmu.

Maca, P., & Pech, P. (2015). *Forecasting SPEI And SPI Drought Indices using The Integrated Artificial Neural Networks*. Comput Intell Neurosci, 17. Article ID 3868519.

McKee, T. B., et al. (1993). *The Relationship Of Drought Frequency And Duration To Time Scales. In: Proceedings of 8th Conference on Applied Climatology*, California, pp. 17–22.

Mishra, A. K., & Desai, V. R. (2005). *Drought forecasting using stochastic models*. Stochastic Environ. Res. Risk Assess., 19, 326–339.

Mishra, A. K., & Singh, V. P. (2010). *A review of drought concepts*. J. Hydrol., 391 (1–2), 202–216.

Mishra, A. K., & Singh, V. P. (2011). *Drought modeling—A review*. J. Hydrol., 403 (1–2), 157–175. DOI: 10.1016/j.jhydrol.2011.03.049.

Mokhtarzad, M., Eskandari, F., Vanjani, N.J., & Arabasadi, A. (2017). *Drought forecasting by ANN, ANFIS, and SVM and comparison of the models*. Environ Earth Sci, 76:72. DOI: 10.1007/s12665-017-7064-0.

Percival, D. B., & Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge University Press. New York.

Sain, H., Kuswanto, H., Purnami, S., & Rahaya, S. (2023). *Fuzzy support vector machine for classification of time series data: A simulation study*. Decision Science Letters, 12(3), 487-498.

Sain, H., & Purnami, S. W. (2015). *Combine sampling support vector machine for imbalanced data classification*. In Procedia Computer Science, 72, 59-66.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag, NY.

Wambua, R. M., Mutua, B. M., & Raude, J. M. (2016). *Prediction of missing hydro-meteorological data series using artificial neural networks (ANN) for Upper Tana River Basin*. Kenya Am J Water Resourc, 4(2), 35–43.

Zhang, Y., Yang, H., Cui, H., & Chen, Q. (2019). *Comparison of the Ability of ARIMA, WNN and SVM Models for Drought Forecasting in the Sanjiang Plain, China*. Natural Resources Research.