



Forecasting Solar Activities based on Sunspot Number Using Support Vector Regression (SVR)

Received
17 July 2020

Revised
27 July 2020

Accepted for Publication
27 July 2020

Published
27 August 2020

Suwanto* and D C R Novitasari

Department of Mathematics, Faculty of Science and Technology, Universitas Islam Negeri Sunan Ampel, Surabaya, 60237, Indonesia.

*E-mail: h72216070@uinsby.ac.id



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)

Abstract

The most significant progress on the 4.0 industrial revolution is the entire computer and robot connected to the internet connection. Satellites, as one of the internet network transmitters, have the threat of destruction if a solar storm occurs. The size of the activity that is on the sun can be known by observing sunspots. Solar activity in the future is known by forecasting sunspot numbers. This research forecasts sunspot numbers using support vector regression (SVR) to minimize adverse effects on the earth as an outcome of solar storms. The best SVR results on forecast sunspot numbers are on annual sunspot numbers obtained using RBF kernel. Measurement results from MSE, RMSE, and MAAPE are 35.32, 5.94, and 0.12, respectively. Forecasting concluded accurately based on MAAPE value, on 2020 and 2021 indicated potentially flare because the result of forecasting sunspot numbers is more than twenty.

Keywords: Forecasting, sunspot, SVR.

1. Introduction

The 4.0 Industry era enables most of the industrial activities to be carried out by computers, robots, and automated technology. The concept of 4.0 Industry revolution is formulated to realize smarter factories [1]. The most significant advancement during the 4.0 Industry revolution is the connected robots and computers through the internet connection. One of the internet network access is satellite. The recent generation of satellite technology can reach the highest performance, even if its first performance was poor. The new generation internet satellite service is sufficiently excellent to compete with other network access [2].

The satellite functions as a signal transmitter from waves in the earth that are threatened by outer space activities, such as sun storms appear due to the substantial activities at the sun. The level of that solar activity can be identified through observation on the sunspot. That is because the complex configuration of sunspot easily provokes an unstable magnetic field that carries flare and coronal mass ejection (CME) [3]. The sun storm may destroy the satellites that orbit around the earth that cripples all activities relying on internet satellite [4]. Therefore, the number of solar activities should be identified for the required preparation and management. The sunspot number represents the number of black spots appear on the sun's surface. The enormous solar activity is represented by a considerable amount of black spots on the sun surface [5]. The number of sunspots can be empirically estimated [6]. Once the configured sunspots become more complex, the magnetic field on the sun surface becomes unstable, which causes a flare, as well as a coronal mass injection (CME). The sequence of sunspot, according to modified Zurich sunspot classification, is divided into A, B, C, D, E, F, and H class [7].

A projection of the sunspot number can be carried out to identify the intensity of solar activities in the future. Some previous studies have been carried out, such as the prediction of sunspot number to identify flare using fuzzy time series Markov chain model [8], long-term sunspot number prediction based on empirical mode decomposition (EMD) analysis and *auto-regressive* (AR) model using sunspot sample data from 1848 to 1992 [9], as well as prediction of time series sunspot based on variational mode decomposition (VMD) and backpropagation neural network-firefly algorithm [10].

In addition to those methods, the prediction can also be carried out using support vector regression (SVR). The usage of SVR as the prediction method has been completed in some previous studies, such as in the forecast of rainfall using some weather parameter [11], prediction of foodstuffs availability in facing demographic bonus [12], and prediction of inflation rate [13]. Another prediction performed using SVR is the forecast of link load on the network [14] using support vector regression (SVR), auto-regressive (AR), and moving average (MA) methods. The most exceptional result of the study on link load prediction is obtained from SVR with a lower error rate than the other two methods. Besides, SVR can also summarize documents based on the sentence position [14]. Therefore, this study forecasts solar activities based on the number of sunspots using SVR to minimize the effect of the solar storm on earth.

2. Method

The prediction of solar activity based on the sunspot number was completed using the sunspot number data from the Sunspot Index and Long-term Solar Observations (SILSO). The data of the sunspot number was constructed to be time-series data. The data was then divided into two and used for the training and testing process.

A technique that predicts a value in the future by considering the current and past data is forecasting [15], [16]. Forecasting has been widely used in various sectors, from the forecast of water weather [17] to the forecast of energy consumption on the natural resource sector [18]. This forecasting is also applied in the economic area, such as the forecast of economic growth [19], forecast of inflation rate [13], and so forth. Forecasting requires special consideration on its level of accuracy since it is followed by a discrepancy between the actual and forecast values.

The tests were started by the independent variable determination test. Once the independent variable was obtained, the time series data (that was divided into two) was formulated. The first part of the data was used for the training to get the prediction model. At the same time, the secondary data was used to validate or test the developed prediction system. The prediction method used in this study is the SVR method.

SVR is a regression model from the machine learning technique with a support vector machine (SVM) method. The primary concept of SVM is to create an isolating line, known as a hyperplane. As suggested by the name, the SVM hyperplane aims to segregate an object into some object, but in SVR, the hyperplane is the function of the regression line [20]– [22]. The regression function of the SVR method is presented below.

$$f(x) = (\alpha_i - \alpha_i^*) \langle x, x_i \rangle + b \tag{1}$$

The measurement of predictor performance was completed by finding the error rate. A little error rate indicates that the system can be used for a forecast. Various estimation can be applied to ensure the quality of a prediction system. This study used some error rate estimation, consisting of mean square error (MSE), root means square error (RMSE), and mean arctangent absolute percentage error (MAAPE) [23], [24]. MSE represents the different average of the prediction value and the actual squared value. If the MSE value is added by root operation, then RMSE value is discovered. On the other hand, MAAPE is the arcus tangent average of the percentage of absolute error. The MSE, RMSE, and MAAPE can be calculated using formulas below.

$$MSE = \sum_{t=1}^n \frac{(A_t - F_t)^2}{n} \tag{2}$$

$$RMSE = \sqrt{\sum_{t=1}^n \frac{(A_t - F_t)^2}{n}} \tag{2}$$

$$MAAPE = \frac{1}{n} \sum_{t=1}^n \left(\arctan \left| \frac{A_t - F_t}{n} \right| \right) \tag{3}$$

A_t is the actual value on the period of t , F_t represents the forecasting value in the period of t , and n is the total observation data.

3. Results and Discussion

The initial step in a study of solar activity forecast is data analysis. The input data, in the form of sunspot number, was sorted into time series. The sample data of sunspot numbers are presented in Table 1; meanwhile, the time series sunspot number is shown in Table 2. The time-series data in Table 2 is formulated by the past sunspot number (x_{t-1}) and the current sunspot number (x_t). The data was used in the forecasting process with x_{t-1} as the input data and x_t as the forecast target.

The forecast of solar activity using support vector regression (SVR) in this study consists of various processes. The first process was training or the formulation of the forecasting model. The second process was the model testing to identify its accuracy level. The SVR model formulation was completed using several types of the kernel, such as linear, gaussian, radial basis function (RBF), also polynomial.

The error rate measurement was carried out by the 2, 3, and 4 formula. The error value of each data comparison and kernel are presented in Table 3. According to the obtained error rates, the RBF kernel usage brings the most excellent results, compare to MSE, RMSE, and MAAPE values of 47.44, 21.78, and 0.29, respectively.

After the error minimum was obtained, the system was used to forecast sunspot numbers from 2013 to 2018, and the results are displayed in Table 4. The estimated RMSE value is 21.78. Based on the smallest to the greatest data range of 4.2–203.3, the forecast or prediction is classified as accurate.

Both small and substantial sunspot numbers illustrate the activities in the sun since sunspot is the pioneer of flare or CME. Based on the previous studies [5], the appearance of flare is related to the sunspot number due to the results of the curve fitting between flare index and the sunspot number has a correlation coefficient of 0.878. Thus, flare appears once the sunspot number reaches its peak—the appearance of flare or CME with tremendous explosion threat technology and climate on the earth.

Table 1. Sample of sunspot number.

Year	Sunspot Number
1990	191.8
1991	203.3
1992	133.0
1993	76.1
1994	44.9

Table 2. Sample of time series data.

x_{t-1}	x_t (Target)
191,8	203,3
203,3	133,0
133,0	76,1
76,1	44,9

Table 3. The error rate of kernel usage.

Kernel	MSE	RMSE	MAAPE
Linear	93.52	30.58	0.41
Gaussian	77.22	27.78	0.35
RBF	47.44	21.78	0.29
Polinomial	98.77	31.42	0.41

Table 4. Results of forecasting.

Year	Forecasting Results
2020	30.51
2021	75.12

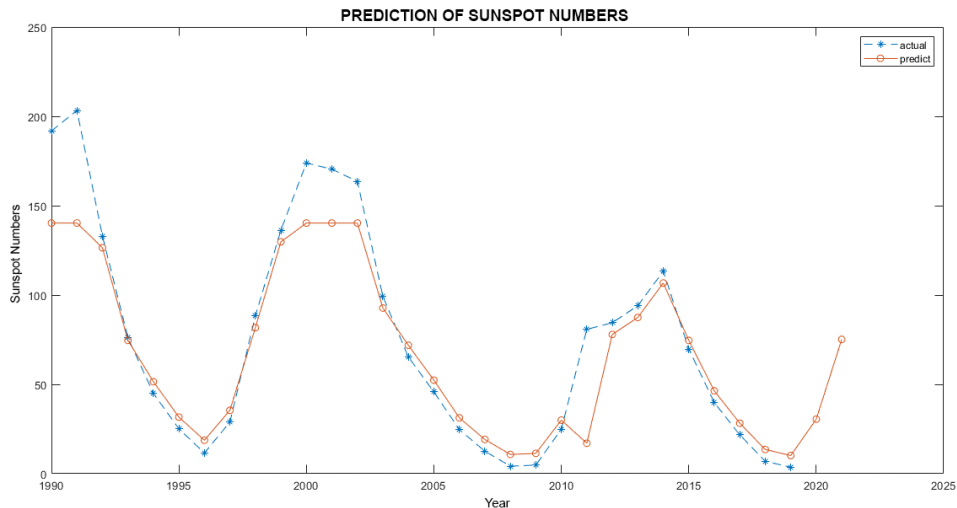


Figure 1. Forecasting results.

Figure 1 illustrates the results of forecasting using kernel RBF. The significant differences materialize in the year points of 1990–1991, 2000–2002, and 2011. Those points carry substantial contributions to the obtained error rate. The application of kernel RBF has the lowest error value than the other kernels; thus, it is the most excellent result.

The forecast for sunspot number in 2020 and 2021 are 30.51 and 75.12, respectively. These results show an improvement from the sunspot number in 2019. The results are in accordance with the intensity of future solar activity. Besides, the increase of solar activities in 2020 and 2021 accelerates the possibility of flare or CME explosion. The lower limit of the sunspot number potentially causes a flare, CME, and sun storm of 20 [24]. Consequently, this situation requires alertness due to the 2020 and 2021 forecast shows more than 20 sunspot number.

4. Conclusion

The forecast of solar activity based on sunspot numbers using SVR with kernel RBF produces the most excellent results. The results using kernel RBF has the lowest MSE, RMSE, and MAPEE than the other kernels. The obtained MSE, RMSE, and MAPEE values are 47.44, 21.78, and 0.29, respectively. Those error rates are derived from the coherence between actual and forecast data, primarily on 1990–1991, 2000–2002, and 2011 with significant differences. The RMSE value of 21.78 is categorized as excellent results. Thus, the forecast is classified as accurate. The results of the sunspot number forecast in 2020 and 2021 are 30.51 and 75.12, respectively. That number potentially causes a flare, CME, and sun storm since the sunspot numbers are above 20.

References

- [1] M. A. K. Bahrin *et al.*, "Industry 4.0: A Review on Industrial Automation and Robotic," *Jurnal Teknologi*, vol. 78, no. 6–13, pp. 137–143, 2016.
- [2] A. Botta and A. Pescapè, "New Generation Satellite Broadband Internet Services: Should ADSL and 3G Worry?," in *The 5th IEEE International Traffic Monitoring and Analysis Workshop (TMA 2013)*, Turin: Università Degli Studi di Milano, 2013.
- [3] G. Chattopadhyay and S. Chattopadhyay, "Monthly Sunspot Number Time Series Analysis and Its Modeling Through Autoregressive Artificial Neural Network," *The Europ. Phys. J. Plus*, vol. 127, no. 43, pp. 1–8, 2012.
- [4] A. Gkana and L. Zachilas, "Sunspot Numbers: Data Analysis, Predictions, and Economic Impacts," *J. Eng. Sci. Technol. Rev (JESTR)*, vol. 8, no. 1, pp. 79–85, 2015.
- [5] Zulkarnain and A. Putriani, "Korelasi Bilangan Sunspot sebagai Parameter Produktivitas Flare pada Siklus Matahari Ke-24," in *Pros. Sem. Nas. Fis.*, Riau: Universitas Riau, 2017.
- [6] D. H. Hathaway, "The Solar Cycle," *Living Rev. in Solar Phys.*, vol. 12, no. 4, pp. 12–13, 2015.
- [7] P. S. McIntosh, "The Classification of Sunspot Groups," *Solar Phys.*, vol. 125, no.2, pp. 251–267, 1990.

- [8] D. C. R. Novitasari, N. Ardhiyah, and N. Widodo, "Flare Identification by Forecasting Sunspot Number using Fuzzy Time Series Markov Chain Model," in *Int. Sem. Intell. Technol. and Its Appl. (ISITIA)*, Surabaya: Institut Teknologi Sepuluh Nopember Surabaya, 2019.
- [9] T. Xu *et al.*, "Long-Term Sunspot Number Prediction based on EMD Analysis and AR Model," *Chinese J. Astron. Astrophys.*, vol. 8, no. 3, pp. 337–342, 2008.
- [10] G. Li, X. Ma, and H. Yang, "A Hybrid Model for Forecasting Sunspots Time Series based on Variational Mode Decomposition and Backpropagation Neural Network Improved by Firefly Algorithm," *Hindawi Computat. Intell. Neurosci.*, vol. 2018, pp. 1–9, 2018.
- [11] D. C. R. Novitasari *et al.*, "Weather Parameters Forecasting as Variables for Rainfall Prediction using Adaptive Neuro Fuzzy Inference System (ANFIS) and Support Vector Regression (SVR)," *J. Phys.: Conf. Ser.*, vol. 1501, p. 012012, 2020.
- [12] S. D. Agustina *et al.*, "Support Vector Regression Algorithm Modeling to Predict the Availability of Foodstuff in Indonesia to Face the Demographic Bonus," *J. Phys.: Conf. Ser.*, vol. 1028, p. 012240, 2018.
- [13] E. M. Priliani, A. T. Putra, and M. A. Muslim, "Forecasting Inflation Rate using Support Vector Regression (SVR) based Weight Attribute Particle Swarm Optimization (WAPSO)," *Scient. J. Inform.*, vol. 5, no. 2, pp. 118–127, 2018.
- [14] A. Fanani *et al.*, "Regression Model Focused on Query for Multi Documents Summarization based on Significance of the Sentence Position," *TELKOMNIKA*, vol. 17, no. 6, pp. 3050–3056, 2019.
- [15] S. R. Binaiya, M. N. Hayati, and I. Purnamasari, "Peramalan Menggunakan Time Invariant Fuzzy Time Series (Studi Kasus: Indeks Harga Konsumen Provinsi Kalimantan Timur)," *EKSPONENSIAL*, vol. 10, no. 2, pp. 175–182, 2019.
- [16] N. Rukhansah, M. A. Muslim, and R. Arifudin, "Fuzzy Time Series Markov Chain dalam Meramalkan Harga Saham," in *Sem. Nas. Ilm. Komp. (SNIK)*, Surakarta: Universitas Sebelas Maret, 2015.
- [17] D. A. Adyanti *et al.*, "Forecasts Marine Weather on Java Sea using Hybrid Methods: TS-ANFIS," in *Proc. 4th Int. Conf. Electric. Eng. Comp. Sci. Inform. (EECSI)*, Yogyakarta: Institute of Advanced Engineering and Science, 2017.
- [18] A. Mehedintu, M. Sterpu, and G. Soava, "Estimation and Forecasts for the Share of Renewable Energy Consumption in Final Energy Consumption by 2020 in the European Union," *Sustainability*, vol. 10, no. 5, pp. 1–22, 2018.
- [19] T. Wang, "Forecast of Economic Growth by Time Series and Scenario Planning Method: A Case Study of Shenzhen," *Scient. Res. Publ.*, vol. 7, no. 2, pp. 212–222, 2016.
- [20] B. Santosa, *Tutorial Support Vector Machines*, Surabaya: Institut Teknologi Sepuluh Nopember Surabaya, 2015.
- [21] A. J. Smola and B. Scholkopf, "A Tutorial on Support Vector Regression," *Stat. Comp.*, vol. 14, no. 3, pp. 199–222, 2004.
- [22] S. Saigal and D. Mehrotra, "Performance Comparison of Time Series Data using Predictive Data Mining Techniques," *Adv. Inform. Min.*, vol. 4, no. 1, pp. 57–66, 2012.
- [23] S. Kim and H. Kim, "A New Metric of Absolute Percentage Error for Intermittent Demand Forecasts," *Int. J. Forecasting*, vol. 32, no. 3, pp. 669–679, 2016.
- [24] B. Setiahadhi, "Limit Bawah Bilangan Sunspot untuk Sunspot yang Berpotensi Melepaskan Flare: Metode Peringatan Dini Antariksa," in *Sem. Nas. Inov. Apl. Teknol. Indus.*, Malang: Institut Teknologi Nasional, 2017.