.

Artificial Neural Network Algorithm Evaluation, Multiple Regression Analysis And Support Vector Machine On Stock Market Prediction

Mohamat Setiawan[1], Ade Johar Maturidi[2]
[1]Manajemen Informatika, Politeknik LP3I Cirebon
[2] Manajemen Informatika, Politeknik LP3I Cirebon

ABSTRACT

Financial time series is one of the most challenging applications of modern time series forecasting. Financial time series is closely related to noise (disturbance signals), non-stationary and deterministic chaotic. The characteristics indicate that no exhaustive information can be obtained from the past behavior of financial markets to capture the full dependency between future prices and that of the past. The purpose of this study is to determine the correct algorithm for Artificial Neural Networks, Multiple Regression Analysis and Support Vector Machines so that the highest accuracy value for prediction trend accuracy and the smallest value for Root Mean Square Error can be obtained. The research method used is 1. Neural network, 2. SVM, 3. Linear regression. From the data that has been tested using the Rapidminer tool with parameters that refer to several literatures [5], data collection is taken from the Stock Market Online Application "MetaTrader version 4" of the type "daily/Daily" with a time span of "03/09/2001 up to 25/07/2012", a total of 2052 data", with the attribute "Date, Open, High, Low, Close, Volume" with the Main attribute "Close" using the Support vector machine algorithm, artificial neural network and multiple linear regression, then the conclusion drawn is that the value close to the series value is the value tested on the support vector machine algorithm, with parameters For RMSE values close to the value "0" obtained from the measurement results in the SVM Algorithm on the RBF (radial basis function) kernel with a value "gamma" $\gamma = 100$ with RMSE = 0.000, and SE = 0.000. with prediction accuracy error = 0.976.

INTRODUCTION

The market influence that easily affects the value of the rupiah currency makes it rare and difficult for researchers to study the rupiah currency market. The economic crisis in Indonesia in 1997 also showed the relationship between macroeconomic conditions and stock performance, where the weakening of the rupiah exchange rate had a major impact on the Capital Market in Indonesia. With the case example above and with the increasing

growth of the world economy, the development of the capital market in Indonesia is very interesting to study (Collins et al., 2000)

Another influence of the global financial crisis on the macro economy is from the interest rate side. With the rise and fall of the dollar exchange rate, interest rates will rise because Bank Indonesia will hold the rupiah so that inflation will increase. Second, the combination of the influence of the high dollar exchange rate and high interest rates will have an impact on the investment sector and the real sector, where investment in the real sector such as property and small and medium enterprises (SMEs) in semester terms will be greatly disrupted. Its influence on investment in the capital market, this global crisis will make people no longer choose the capital market as an attractive place to invest because the macro conditions are less supportive. The value that is close to the series value is the value with the test on the support vector machine algorithm, with the parameter For the RMSE value that is close to the value "0" obtained from the measurement results on the SVM Algorithm on the RBF kernel (radial basis function) with the value of "gamma" $\gamma = 100$ with the value of RMSE = 0.000, and SE = 0.000. with prediction accuracy error = 0.976

METHOD

The research methods used are 1. Neural network is an approximation (=approach) function and universal that maps every nonlinear function without a priori assumptions. (=A priori assumes before knowing). SVM is a promising method for financial time series prediction because it uses a risk function consisting of empirical errors (=a source of knowledge obtained from observation or experimentation) and regularized (=fixed settings) derived from the principle of minimizing structural risk. 3. Linear regression is an approach to modeling the relationship between the dependent variable y scalar and one or more explanatory variables denoted X. The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, it is called multiple linear regression. (This in the term should be distinguished from the multiple linear regression model, where several correlated dependent variables are estimated, [citation needed] Rather than a single scalar variable.). The design of this research is the strategy, plan, and structure of conducting project research. The sequence of processes to be carried out is depicted in Figure 1.

Figure 1. Process Sequence

RESULTS AND DISCUSSION

From the results of the observation of the dataset test using data from metatrader version 4 as many as 2052 transaction data from September 03, 2004 (2004.09.03 in Date/Times English (United States) format) to July 25, 2012 (2012.07.25 in Date/Times English (United States) format), with the attributes DATE, OPEN, HIGH, LOW, CLOSE and VOLUME with the Set Role "CLOSE", as follows:
1. RMSE
The results of the comparison of 3 Algorithms by taking from each study/research with the lowest value (min), then the results of the lowest value (min) for the RMSE value are in the measurement using 3 Algorithms, namely the SVM Algorithm value, especially in the RBF Kernel with parameter values, namely the value of "C" = 0 and the value of "Gamma" = 100 with the value of RMSE = 0.000 and at the value of "C" = 10 and the value of "Gamma" = 100 with RMSE value = 0.000.
2. Square Error
From the research results by taking the lowest value (min) from each algorithm that was researched, then for each Square error value obtained, both the SVM Algorithm, MLR Algorithm and ANN algorithm, all have SE values = 0.000
3. Prediction Trend Acuracy
From the research that has been done, the Prediction Trend Acuracy (PTA) value, which is taken from each highest value (max) from the results of the algorithm research, both SVM, ANN and MLR, then for the highest value, namely in the SVM research on the RBF kernel, the parameter value is the value "C" = 0 and the value "Gamma" = 100 with a PTA value = 0.976 and at the value "C" = 10 and the value "Gamma" = 100 with a PTA value = 0.976

CONCLUSION

From the data that has been tested using the Rapidminer tool with parameters referring to several literatures (Haushofer, 2003), the data collection was taken from the Online Stock Market Application "MetaTrader version 4" with the type "daily" with a time span from "03/09/2001 to 25/07/2012", as many as 2052 data ", with the attributes" Date, Open, High, Low, Close, Volume "with the main attribute" Close "using the Support vector machine algorithm, artificial neural network and multiple linear regression, then the conclusion taken is that the value that approaches the series value is the value with the test on the support vector machine algorithm, with the parameters For the RMSE value that approaches the value "0" obtained from the measurement results on the SVM Algorithm on the RBF kernel (radial basis function) with the value of "gamma" $\gamma = 100$ with the value of RMSE = 0.000, and SE = 0.000. with prediction accuracy error = 0.976.

REFERENCES

Bryson, A. E., Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1974). Backpropagation. en.wikipedia.org/Backpropagation neural network.

Categories of forecasting methods. (n.d.).en.wikipedia.org/forecasting.

Chen, R. (2007). Using SVM with Financial Statement Analysis for Prediction of Stocks, 7(4), 63–72.

Collins, M., Avenue, P., Room, A., Park, F., & Schapire, R. E. (2000). Logistic Regression , AdaBoost and Bregman Distances. Proceedings of the Thirteenth Annual Conference on Computational Learning Theory, 13, 1–26.

Cuaresma, J. C. (2010). Modeling and Predicting the EUR / USD Exchange Rate : The Role of Nonlinear Adjustments to Purchasing Power Parity, 64–76.

Elkan, C. (2010). Predictive analytics and data mining. Predictive analytics and data mining (pp. 1–107).

Falinouss, P. (2007). MASTER ' S THESIS Stock Trend Prediction Using News Articles.

Farrell, M. T., & Correa, A. (2007). Gaussian Process Regression Models for Predicting Stock Trends. Mit. Media Edu, 1–9.

Fitting a trend : least-squares. (n.d.).en.wikipedia.org/Trend Estimation.

Fox, J. (2010). Appendices to Applied Regression Analysis , Generalized Linear Models , and Related Methods , Second Edition.

From, F. (n.d.). Polynomial kernel, 2–4.

Galushkin, A. I. (2007). Neural network Theory (pp. 1–391).

Hanias, M. P. (2008). Time Series Prediction of Dollar \ Euro Exchange Rate Index. International Research Journal of Finance and Economics, 15(15), 232–239.

Hardiningsih, P. (2001). PENGARUH FAKTOR FUNDAMENTAL DAN RISIKO EKONOMI TERHADAP RETURN SAHAM PERUSAHAAN DI SURSA EFEK JAKARTA. Universitas Diponegoro.

Harianto, Farid, 1998, Perangkat dan Teknik Analisis Investasi di Pasar Modal Indonesia, PT BEJ, Jakarta.

Haushofer, H. (2003). Fundamental and Nonfundamental Factors in the Euro / U . S . Dollar Market in 2002 and 2003, 58–76.

Hsieh, C., Chang, K., Ringgaard, M., & Tsuda, K. (n.d.). Radial basis function kernel. en.wikipedia.org/radial basis function kernel.

Kara, Y., Acar, M., & Kaan, Ö. (2011). Expert Systems with Applications Predicting direction of stock price index movement using artificial neural networks and support vector machines : The sample of the Istanbul Stock Exchange. Expert Systems With Applications, 38(5), 5311–5319. doi:10.1016/j.eswa.2010.10.027

KHOKHAR, D. M. N. M. S. D. H. S. D. S. M. H. S. R. H., & AWAN, Z. B. C. M. @ M. S. A. M. (2005). Design and Development Of Intelligent Knowlegde Discovery Sistem Stock Exchange Database.

Kim, K. (2003). Financial time series forecasting using support vector machines. NeuroComputing, 55, 307–319. doi:10.1016/S0925-2312(03)00372-2

Linear prediction. (n.d.).en.wikipedia.org/linear prediction.

Machine Learning Model and feature selection Model selection issues. (n.d.).

Minsky, M. (1969). Feedforward neural network. en.wikipedia.org/feedforward neural network.

Multinomial, P., Multilevel, P., Semiparametric, N., Quantile, R., Principal, I., & Errorsvariables, L. S. (n.d.). Linear regression. en.wikipedia.org/Linear regression Regression.

Paula Fernandes, J. T. (joaopt@ipb. pt). (n.d.). APPLYING THE ARTIFICIAL NEURAL NETWORK METHODOLOGY TO TOURISM TIME SERIES FORECASTING.

Radial basis function network. (n.d.).en.wikipedia.org/radial basis function network.

Sap, M. N. M., & Awan, A. M. (2005). Stock Market Prediction Using Support Vector Machines. Jurnal Teknologi Maklumat, Jilid 17,, 27–35.

Setiadi, Darmawan (2010), PENGARUH VARIABEL FUNDAMENTAL TERHADAP HARGA SAHAM PDA PERUSAHAAN MANUFAKTUR DI BURSA EFEK INDONESIA, Universitas Dian nuswantoro.

Suparti, A. M. dan A. R. (2007). ESTIMASI REGRESI WAVELET THRESHOLDING DENGAN METODE BOOTSTRAP. Vol, Jurnal Matematika, 10(2), 43–50.

Tandelilin, Eduardus, 2000. "Pasar Modal Indonesia: Problem dan Prospek", Wahana vol.3 no.2,

Tartglia, A., & Fisica, D. Di. (2008). A Comparison of Neural Network, Arima Model and Multiple Regression Analysis in Modeling Rice Yields. INTERNATIONAL JOURNAL OF SOFT COMPUTING APPLICATIONS (IJSCA), (3), 113–127.

THOBARRY, A. A. (2009). ANALISIS PENGARUH NILAI TUKAR, SUKU BUNGA, LAJU INFLASI DAN PERTUMBUHAN GDP TERHADAP INDEKS HARGA SAHAM SEKTOR PROPERTI (KAJIAN EMPIRIS PADA BURSA EFEK INDONESIA PERIODE PENGAMATAN TAHUN 2000-2008). Universitas Diponegoro.

Tomasouw, B. P. (2012). MULTICLASS TWIN BOUNDED SUPPORT VECTOR MACHINE UNTUK. Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA,, 2(2004)