

## INTISARI

### OTOMATISASI PEMODELAN *NONLINEAR AUTOREGRESSIVE* *NEURAL NETWORK WITH EXOGENOUS INPUTS* UNTUK PERAMALAN DATA RUNTUN WAKTU

Oleh

HERMANSAH

17/420358/SPA/00627

Penelitian disertasi ini mengusulkan tiga prosedur baru dalam pembentukan model *nonlinear autoregressive neural network with exogenous inputs* (NARX) dengan arsitektur *series-parallel* untuk peramalan data runtun waktu. Pertama, prosedur pembentukan model NARX menggunakan pendekatan dengan pemodelan secara otomatis. Kedua, prosedur otomatisasi pemodelan NARX dikembangkan menggunakan operator ensemble. Tujuan operator ensemble digunakan adalah untuk mengatasi masalah overfitting serta kelemahan dari otomatisasi pemodelan NARX individu/tunggal. Operator ensemble yang digunakan, yaitu mean, median, dan modus. Ketiga, prosedur otomatisasi pemodelan NARX dikembangkan menggunakan metode *stepwise* atau *least absolute shrinkage and selection operator* (LASSO). Metode *stepwise* atau LASSO ini digunakan untuk mendapatkan neuron input optimal di lapisan input. Berikut adalah tahapan pada otomatisasi pemodelan NARX. Proses awal dimulai dengan mengidentifikasi frekuensi dan lag autoregressive data. Kemudian mengidentifikasi data memiliki komponen *trend* dan musiman. Jika komponen *trend* dan musiman teridentifikasi, maka differencing pertama diambil dan *dummy* musiman deterministik dibuat. Kemudian data diskalakan ke nilai  $[-1, 1]$ . Kemudian neuron input awal ditentukan berdasarkan lag autoregressive data dan neuron input optimal diseleksi menggunakan metode *stepwise* atau LASSO. Selanjutnya, jumlah neuron optimal di lapisan tersembunyi ditentukan berdasarkan nilai *mean squared error* (MSE) terkecil dari satu sampai total neuron di lapisan input. Selanjutnya, arsitektur model NARX diimplementasikan 50 kali menggunakan bobot dan bias awal yang berbeda secara acak. Hasil dari 50 model NARX tersebut kemudian dikombinasikan dengan operator ensemble untuk menghasilkan peramalan akhir.

Pada penelitian disertasi ini, hasil studi empiris menunjukkan bahwa akurasi peramalan model NARX secara otomatis lebih akurat daripada model-model

yang lain. Dengan demikian dapat disimpulkan bahwa prosedur pembentukan model NARX secara otomatis merupakan cara yang efektif untuk meningkatkan akurasi peramalan. Selain itu, otomatisasi pemodelan NARX juga dilakukan perbandingan untuk setiap operator ensemble. Hasil studi kasus untuk data yang memiliki pola *trend* dan musiman menunjukkan bahwa operator ensemble median dan modus yang jarang digunakan lebih akurat daripada operator ensemble mean. Operator ensemble median membutuhkan waktu komputasi rata-rata paling sedikit, diikuti operator ensemble modus. Sedangkan pada pola data *trend* dan non-musiman diperoleh operator ensemble mean paling akurat. Pada hasil studi kasus juga menunjukkan bahwa variabel input eksternal dengan variabel *dummy* biner paling akurat untuk pola data *trend* dan musiman dengan sifat multiplikatif. Sedangkan pada pola data *trend* dan musiman dengan sifat aditif diperoleh variabel input eksternal dengan variabel *dummy* sinus kosinus lebih akurat. Di sisi lain, akurasi peramalan didapatkan lebih baik pada data dengan melalui proses differencing pertama.

Akurasi peramalan pada otomatisasi pemodelan NARX juga dilakukan perbandingan antara metode *stepwise* dan LASSO untuk mendapatkan neuron input optimal di lapisan input. Hasil studi kasus pada data training dan testing diperoleh otomatisasi pemodelan NARX menggunakan metode *stepwise* (model NARX 1) memberikan akurasi yang lebih baik. Sedangkan otomatisasi pemodelan NARX menggunakan metode LASSO (model NARX 2) menempati urutan kedua dengan perbedaan yang signifikan. Selain itu, kombinasi prosedur pembentukan model NARX secara otomatis diuji pada tiga bagian, yaitu operator ensemble, algoritma pembelajaran, dan fungsi aktivasi. Algoritma pembelajaran yang digunakan, yaitu *backpropagation*, *resilient backpropagation with and without weight backtracking* (RPROP), dan RPROP dengan *smallest absolute gradient* atau *smallest learning rate* (GRPROP). Sedangkan fungsi aktivasi yang digunakan, yaitu fungsi logistik dan tangen hiperbolik. Hasil pada studi kasus menunjukkan bahwa operator ensemble sangat berguna karena memberikan akurasi yang lebih baik. Sedangkan fungsi aktivasi menggunakan tangen hiperbolik memberikan hasil terbaik secara konsisten. Sementara algoritma pembelajaran RPROP menunjukkan paling efektif dan menjanjikan sehubungan dengan waktu konvergensi.

**Kata-kata kunci:** Data runtun waktu, Peramalan, *Neural network*, NARX, Otomatisasi, *Stepwise*, LASSO, Operator ensemble

## ABSTRACT

### **AUTOMATIC NONLINEAR AUTOREGRESSIVE NEURAL NETWORK WITH EXOGENOUS INPUTS MODELING FOR TIME SERIES DATA FORECASTING**

By

HERMANSAH

17/420358/SPA/00627

This dissertation proposes three new procedures for forming a nonlinear autoregressive neural network with exogenous inputs (NARX) model with a series-parallel architecture for forecasting time series data. First, the NARX modeling procedure uses an automated modeling approach. Second, the NARX modeling automation procedure was developed using the ensemble operator. The purpose of the operator ensemble used is to overcome the problem of overfitting and the drawbacks of automating individual/single NARX modeling. The ensemble operators used are mean, median, and mode. Third, the NARX modeling automation procedure was developed using the stepwise or the least absolute shrinkage and selection operator (LASSO) method. This stepwise or LASSO method obtains optimal input neurons in the input layer. The following are the steps in the automation of NARX modeling. The initial process begins by identifying the frequency and lag of autoregressive data. Then identify data that has trend and seasonal components. If trend and seasonal components are identified, the first differencing is taken, and a deterministic seasonal dummy is created. Then the data is scaled to the value  $[-1, 1]$ . Then the initial input neuron is determined based on the lag autoregressive data, and the optimal input neuron is selected using the stepwise or LASSO method. Next, the optimal number of neurons in the hidden layer is determined based on the smallest mean squared error (MSE) from one to the total number of neurons in the input layer. Next, the architecture of the NARX model is implemented 50 times using randomly different initial weights and biases. The results of the 50 NARX models are then combined with the ensemble operator to produce the final forecast.

In this dissertation research, empirical studies show that the forecasting accuracy of the NARX model is automatically more accurate than other models. Thus it can be concluded that the NARX model formation procedure automatically is an effective way to improve forecasting accuracy. In addition, the

NARX modeling automation is also compared for each ensemble operator. The case study results for data that have trend and seasonal patterns show that the median ensemble operator and the rarely used mode are more accurate than the mean ensemble operator. The median ensemble operator requires the least average computation time, followed by the mode ensemble operator. Meanwhile, in trend and non-seasonal data patterns, the most accurate ensemble mean operator is obtained. The case study results also show that external input variables with binary dummy variables are the most accurate for trend and seasonal data patterns with multiplicative properties. Meanwhile, external input variables with dummy sine cosine variables are more accurate in trend and seasonal data patterns with additive properties. On the other hand, forecasting accuracy is obtained better on the data by going through the first differencing process.

Forecasting accuracy in NARX modeling automation is also carried out by comparing stepwise and LASSO methods to obtain optimal input neurons in the input layer. The results of the case study on the training and testing data obtained that the automation of NARX modeling using the stepwise method (model NARX 1) provides better accuracy. While the automation of NARX modeling using the LASSO method (model NARX 2) ranks second with a significant difference. In addition, the combination of NARX model formation procedures is also automatically tested in three parts: ensemble operators, learning algorithms, and activation functions. The learning algorithms used are backpropagation, resilient backpropagation with and without weight backtracking (RPROP), and RPROP with the smallest absolute gradient or smallest learning rate (GRPROP). While the activation functions used are logistic function and tangent hyperbolic. The case study results show that the ensemble operator is very useful because it provides better accuracy. While the activation function using a hyperbolic tangent gives the best results consistently. At the same time, the RPROP learning algorithm shows the most effectiveness and promise concerning the convergence time.

**Keywords:** Time series data, Forecasting, Neural network, NARX, Automation, Stepwise, LASSO, Operator ensemble