

PREDIKSI INDEKS HARGA SAHAM MENGGUNAKAN SELF ORGANIZING MAP DAN GENETIC PROGRAMMING

STOCK PRICE INDEX PREDICTION USING SELF ORGANIZING MAP AND GENETIC PROGRAMMING

Lintong Aldiron Sihombing¹

Rian Febrian Umbara²

Izzatul Ummah³

^{1,2,3}Prodi S1 Ilmu Komputasi, Fakultas Informatika, Universitas Telkom

¹aldiron4@gmail.com

²rianum123@gmail.com

³izzatul.ummah@gmail.com

Abstrak

Investasi dalam bidang keuangan adalah penting untuk semua orang. Salah satu investasi keuangan adalah saham. Indeks harga saham merupakan indikator yang menggambarkan pergerakan harga saham dalam suatu periode. Prediksi indeks harga saham dapat dilakukan menggunakan metode gabungan Self Organizing Map (SOM) dan Genetic Programming (GP) yang ditemukan oleh Chih-Ming-Hsu. Self Organizing Map adalah salah satu metode yang bertujuan untuk mengelompokkan data, kemudian metode Genetic Programming fokus pada salah satu tujuan utama ilmu komputer, yaitu *automatic programming*. Keluaran dari metode ini adalah sebuah formula aritmatika yang direpresentasikan dalam bentuk pohon (*tree*).

Metode Self Organizing Map dan Genetic Programming mempresentasikan formula-formula prediksi di setiap cluster. Hasil prediksi dari gabungan metode Self Organizing Map dan Genetic Programming ini dapat diukur keakuratan nya dengan menghitung *Root Mean Square Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE). Error yang didapatkan untuk prediksi Indeks Harga Saham Gabungan (IHSG) dengan menggunakan metode Self Organizing Map dan Genetic Programming untuk data training adalah RMSE = 0.0807 dan MAPE = 13.7536%, data testing RMSE = 0.0682 dan MAPE = 0.12.9248%, data validasi RMSE = 0.0517 dan MAPE = 9.5406%. Sedangkan menggunakan Genetic Programming saja, didapatkan hasil untuk data training, RMSE = 0.4397 MAPE = 95.0631%, data testing RMSE = 0.4454 MAPE = 95.0264%, dan data validasi RMSE = 0.4702 MAPE = 97.4519%. Dengan nilai ini disimpulkan bahwa metode Self Organizing Map dan Genetic Programming tersebut terbukti mampu memprediksi harga penutupan Indeks Harga Saham Gabungan (IHSG) dengan nilai yang cukup akurat.

Kata kunci : Indeks harga saham, Self Organizing Map, Genetic Programming.

Abstract

Investment in finance nowadays is important for all people. One of investments in finance is stock. Stock price index is an indicator that shows the movement of stock price in one period. Stock price index prediction can be done by using the hybrid of Self Organizing Map (SOM) and Genetic Programming (GP) proposed by Chi-Ming-Hsu. Self Organizing Map is a method that aimed to cluster data, and Genetic Programming is aimed to make automatic programming. The output of this method is a arithmetic formula in tree.

Self Organizing Map and Genetic Programming method present prediction formulas in every cluster. The accuracy prediction result of hybrid method SOM and GP can be measured by Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The errors of Composite Stock Price Index prediction by using hybrid method SOM – GP are, for training data RMSE = 0.0807 and MAPE = 13.7536%, testing data RMSE = 0.0682 and MAPE = 0.12.9248%, and for validation data RMSE = 0.0517 and MAPE = 9.5406%. While only using Genetic Programming the errors of training data are RMSE = 0.4397 and MAPE = 95.0631%, for testing data are RMSE = 0.4454 and MAPE = 95.0264%, and for validation data are RMSE = 0.4702 and MAPE = 97.4519%. From these results, it can be concluded that hybrid method SOM – GP can predict the closing price of Composite Stock Price Index accurately.

Keywords : Stock price index, Self Organizing Map, Genetic Programming.

1. Pendahuluan

Saham adalah satuan nilai atau pembukuan dalam berbagai instrumen finansial yang mengacu pada bagian kepemilikan sebuah perusahaan. Pasar yang menyediakan tempat jual beli saham dinamakan bursa

saham. Akan sangat membantu apabila kita dapat memprediksi indeks harga saham untuk masa mendatang. Indeks harga saham merupakan indikator yang menggambarkan pergerakan harga saham dalam suatu periode. Dengan membaca indeks ini, kita dapat mengetahui tren yang sedang terjadi di pasar, apakah sedang naik, turun, atau stabil sehingga investor dapat menentukan kapan untuk menjual, menahan atau membeli saham

Chih-Ming Hsu yang pertama sekali mengenalkan bagaimana memprediksi harga saham menggunakan gabungan Self Organizing Map (SOM) dan Genetic Programming (GP)[4]. Menurut Chih-Ming Hsu, proses gabungan dengan Self Organizing Map dan Genetic Programming dibagi kedalam 3 tahap. Pada tahap pertama, adalah proses *clustering* data menggunakan Self Organizing Map. Hasil dari tahap ini adalah beberapa *cluster* (kelompok) data. Pada tahap yang kedua, adalah proses membangun model prediksi menggunakan algoritma Genetic Programming (GP). Pada tahap ini hasil dari *cluster* data pada tahap pertama dibagi lagi menjadi data *training*, data *testing* dan data *validation*. Kemudian dari data *training* dan data *testing* tersebut akan dihasilkan model-model Genetic Programming (GP). Pada tahap ketiga, adalah proses memilih model yang terbaik dari tiap model-model Genetic Programming yang dihasilkan sebelumnya.

2. Landasan Teori

2.1 Saham

Saham adalah surat berharga yang merupakan tanda kepemilikan seseorang atau badan terhadap suatu perusahaan. Wujud **saham** adalah selembar kertas yang menerangkan bahwa pemilik kertas itu adalah pemilik perusahaan yang menerbitkan kertas tersebut[1].

2.2 Indeks Harga Saham

Indeks harga saham merupakan indikator yang menggambarkan pergerakan harga saham dalam suatu periode. Dengan membaca indeks ini, kita dapat mengetahui tren yang sedang terjadi di pasar, apakah sedang naik, turun, atau stabil sehingga investor dapat menentukan kapan untuk menjual, menahan atau membeli saham.

2.3 16 Indikator Teknis

Enam belas (16) indikator teknis adalah variable input untuk memprediksi indeks harga saham, sesuai dengan penelitian oleh Kim and Han (2000), Kim and Lee (2004), Tsang et al. (2007), Chang and Liu (2008), Ince and Trafalis (2008), Huang and Tsai (2009), dan Lai et al. (2009). Enam belas (16) indikator teknis tersebut adalah *10-day moving/average* (MA_10), *20 day bias* (BIAS_20), *Moving average convergence/divergence* (MACD), *9-day stochastic indicator* (K_9), *9-day stochastic indikator* (D_9), *9-day Williams overbought/oversold indeks* (WMS%R_9), *14-day plus directional indikator* (+DI_14), *14-day minus directional indikator* (-DI_14), *10-day momentum* (MTM_10), *10-day rate of change* (ROC_10), *5-day relative strength indeks* (RSI_5), *24-day commodity channel index* (CCI_24), *26-day buying/selling momentum indikator* (AR_26), *26-day buying/selling willingness indikator* (BR_26), *26-day volume ratio* (VR_26), *13-day psychological line* (PSY_13),

Enam belas (16) indikator teknis ini dihitung untuk data IHSG setiap harinya. Setelah perhitungan 16 indikator teknis ini, kemudian adalah menormalisasi 16 indikator teknis inputan tersebut ke dalam range [-1 1], yang kemudian akan dilakukan proses pengelompokan data menggunakan metode *Self Organizing Map* (SOM).

2.4 Self Organizing Map (SOM)

Self Organizing Map (SOM) pertama sekali dikenalkan oleh Kohonen (1989)[3]. Self Organizing Map (SOM) adalah salah satu metode yang bertujuan untuk mengelompokkan data. *Clustering* adalah aktifitas pengelompokkan berita tanpa supervisi ke dalam *cluster*. Proses pengelompokan data ini dilakukan berdasarkan nilai bobot setiap data. SOM mencari neuron dengan nilai bobot yang sesuai dengan nilai masukan, kemudian mengubah nilai bobot tersebut dengan jarak terdekat. Bobot yang paling dekat dengan nilai masukan akan disebut sebagai *winner* yang dihitung dengan *Euclidean distance*. Kemudian tetangga dari neuron pemenang ini juga akan diupdate nilainya sama seperti neuron pemenang. Jarak tetangga biasanya telah ditetapkan di awal.

Langkah-langkah proses kerja Self Organizing Map adalah sebagai berikut :

Langkah 0 : Menginisialisasi berat W_j dan fungsi ketetanggaan h_{ij} , kemudian menentukan radius topologi ketetanggaan R , dan Learning Rate (α).

Langkah 1 : Apabila kriteria berhenti belum terpenuhi, lakukan langkah 2 – 8.

Langkah 2 : Untuk tiap vektor masukan $\bar{X}$, lakukan langkah 3-5.

Langkah 3 : Untuk setiap *cluster* neuron j untuk menghitung jarak vektor, lakukan

$$D(j) = \|\bar{X} - \bar{W}_j\|.$$

Langkah 4 : Cari indeks c dimana nilai $D(c)$ adalah minimum.

Langkah 5 : Untuk semua neuron j , dan setiap tetangga dengan radius R dari neuron c , lakukan :

$$\bar{W}_j(t+1) = \bar{W}_j(t) + \alpha(t) h_{cj}(t) |\bar{X} - \bar{W}_j(t)|$$

Dimana t adalah kordinat waktu diskrit.

$$h_{cj}(t) = \begin{cases} 1, & \text{untuk neuron } j \text{ dalam radius } R \text{ neuron winner } c \\ 0, & \text{untuk lainnya.} \end{cases}$$

Langkah 6 : Update Learning Rate α dan fungsi ketetanggaan h_j .

Langkah 7 : Kurangi radius atau jarak topologi ketetanggaan R pada langkah selanjutnya.

Langkah 8 : Periksa kriteria untuk berhenti.

Jadi hasil akhir dari proses Self Organizing Map ini adalah kelompok-kelompok data (*cluster*).

2.5 Genetic Programming (GP)

Genetic Programming fokus pada salah satu tujuan utama ilmu komputer, yaitu *automatic programming*. Tujuan dari automatic programming adalah membuat suatu program secara otomatis yang membuat komputer sanggup menyelesaikan masalah[5]. Genetic Programming adalah suatu algoritma berbasis Evolutionary Computation atau seperti teori evolusi. GP diperkenalkan pertama kali oleh J. Koza pada era 1990-an di Amerika Serikat[2]. Biasanya GP digunakan untuk machine learning: prediksi, klasifikasi, dan sebagainya. Ciri khusus GP adalah membutuhkan ukuran populasi yang besar hingga ribuan individu. Representasi kromosom GP adalah pohon (*tree*).

Secara umum, tujuan Genetic Programming adalah menghasilkan fungsi matematika pada setiap *cluster*. Kemudian fungsi tersebut adalah sebuah representasi prediksi indeks harga saham dalam bentuk persamaan atau formula untuk memperoleh indeks harga saham yang sebenarnya .

2.6 Gabungan Self Organizing Map dan Genetic Programming (SOM-GP)

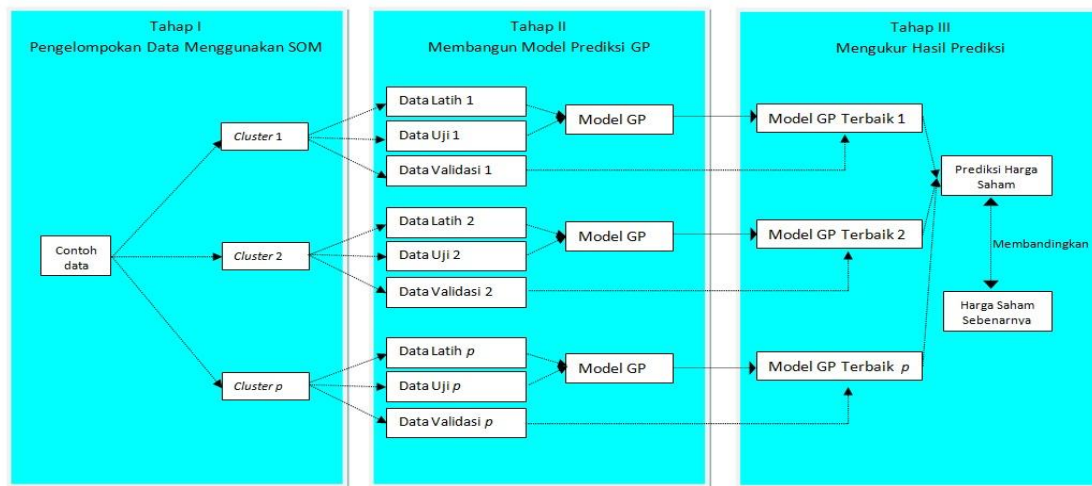
Gabungan Self Organizing Map dan Genetic Programming (SOM-GP) merupakan proses yang digunakan untuk memprediksi indeks harga saham. Prosedur SOM-GP ini dibagi kedalam tiga tahapan pengerjaan. Pada tahap pertama dilakukan pengumpulan data penjualan saham seperti harga pembukaan, harga penutupan, harga tertinggi, harga terendah. Kemudian indikator *10-day moving/average* (MA_10), *20 day bias* (BIAS_20), *Moving average convergence/divergence* (MACD), dan 13 indikator lainnya juga dihitung untuk mengumpulkan data yang akan digunakan untuk menjadi data masukan $\mathbf{X} = (X_1, X_2, X_3, \dots, X_{16})$. Dan kumpulan hasil penghitungan tersebut digunakan sebagai masukan pada *layer input*. Kemudian dari setiap hasil pengelompokan data ini akan dibagi menjadi data *testing* dan data *training*. Kemudian menormalisasi 16 indikator teknis ke dalam range $[-1 \ 1]$.

Pada tahap yang kedua, prediksi indeks harga saham pada hari berikutnya yang dinotasikan dengan y' terlebih dahulu dinormalisasikan dalam range -1 dan 1 . Kemudian adalah mengelompokkan data 16 indikator teknis yang telah dinormalisasi dan harga penutupan prediksi yang dinormalisasi menggunakan metode SOM. Algoritma Genetic Programming akan diterapkan untuk data *training* dan data *testing* yang kemudian akan menghasilkan model Genetic Programming, dimana hasil dari Genetic Programming ini adalah sebuah fungsi relasi antara harga saham perhari dengan harga saham pada hari selanjutnya (y') yang dipresentasikan dalam bentuk pohon (*tree*).

Pada tahap ketiga, dilakukan proses SOM untuk tiap indikator normalisasi $\mathbf{X}' = (X'_1, X'_2, X'_3, \dots, X'_{16})$ dan mengelompokkan $\mathbf{X}'$ kedalam cluster C . Lalu dari model-model Genetic Programming yang dihasilkan pada tahap yang kedua akan dipilih model GP yang terbaik dan kemudian akan dibandingkan dengan indeks harga saham yang sebenarnya.

3. Perancangan Sistem

Model umum sistem dalam penyelesaian masalah untuk memprediksi indeks harga saham dilakukan sesuai dengan diagram alir (*flowchart*) dibawah ini.



Gambar 1.1. Flowchart penyelesaian masalah.

3.1 Tahap I

Pengelompokan contoh data menggunakan *Self Organizing Map (SOM)*.

Langkah-langkah pengelompokan data menggunakan SOM adalah sebagai berikut:

- Mengumpulkan data historis harga saham kemudian menghitung 16 indikator-indikator teknis.

Pengumpulan data historis adalah langkah awal dalam memulai prediksi. Pada penelitian ini penulis mengumpulkan data historis dari Indeks Harga Saham Gabungan (IHSG) dalam kurun waktu 5 tahun, yaitu mulai 3 Januari 2011 - 19 Januari 2015 yang diambil dari alamat web *Yahoo Finance*. Terhitung jumlah data yang dikumpulkan sebanyak 1027 data. Data perhari dalam Indeks Harga Saham Gabungan ini meliputi *Opening Price*, *Closing Price*, *Low Price*, *High Price*, *Volume*, dan *Adj Close*.

- Menormalisasi indikator-indikator teknis $X = (X_1, X_2, \dots, X_{16})$ ke dalam rentang $[-1, 1]$, yang direpresentasikan dengan $X' = (X'_1, X'_2, \dots, X'_{16})$.

Normalisasi data adalah proses penskalaan nilai atribut ke dalam range tertentu. Data atribut yang dinormalisasi pada tahap ini adalah 16 indikator teknis, yang dinormalisasi ke dalam range $[-1, 1]$. Normalisasi data 16 indikator teknis ini dilakukan menggunakan rumus :

$$X' = (1 - e^{(-X)}) / (1 + e^{(-X)})$$

dimana :

$x = (\text{data mean}) / \text{std}$.

$e = \text{bilangan natural } (2.718281828)$.

Data mean = rata-rata per indikator.

Std = standart deviasi per indikator.

- Menerapkan metode SOM untuk membagi data inputan $X' = (X'_1, X'_2, \dots, X'_{16})$ kedalam beberapa kelompok data (*clusters*), kemudian mengukur keefesienannya berdasarkan *clustering efficiency (CE)*.

3.2 Tahap II

Membangun model-model prediksi menggunakan *Genetic Programming (GP)*.

Langkah membangun model-model menggunakan GP tersebut adalah :

- Menormalisasi harga penutupan hari berikut yang akan diprediksi kedalam range antara -1 dan 1, yang dinotasikan dengan y' .

Pada tahap ini rumus untuk menormalisasi menggunakan rumus Newdata seperti pada tahap I.

- Membagi data masukan $X' = (X'_1, X'_2, \dots, X'_{16})$ yang telah dinormalisasi dan y' ke dalam data *training*, data *testing* dan data *validation* sesuai dengan proporsi yang telah ditetapkan.

Pembagian data masukan $X' = (X'_1, X'_2, \dots, X'_{16})$ yang telah dinormalisasi dan y' dilakukan sesuai proporsi perbandingan 4 (*training*) : 1 (*testing*) : 1 (*validasi*).

- Gunakan algoritma *Genetic Programming* untuk membangun beberapa model prediksi dari data *training* dan data *testing* yang telah dikelompokkan dan dibagi menggunakan metode SOM pada tahap 1. Proses *Genetic Programming* dilakukan per *cluster*. Setiap *cluster* data akan menghasilkan formula-formula prediksi. Untuk sekali *running* untuk proses ini akan menghasilkan 1 formula solusi. Semakin

banyak melakukan *running*, maka akan memperbesar kemungkinan mendapatkan formula prediksi baru.

- o Memilih model-model *Genetic Programming* yang paling optimal pada tiap *cluster* untuk memperkirakan hubungan fungsional antara indikator teknis yang telah dinormalisasi dan prediksi harga saham esok hari (y'), berdasarkan *Mean Square Error (MSE)* terkecil dari data *training* dan data *testing*.

Pemilihan model terbaik prediksi untuk setiap cluster ini dilakukan dengan menghitung error terkecil untuk setiap formula. Adapun rumus yang dipakai untuk menghitung error tersebut adalah

$$RMSE \text{ (Root Mean Square Error)} = \sqrt{\sum_{i=1}^{no} \left(\frac{(y_i - y'_i)^2}{no} \right)},$$

$$MAPE \text{ (Mean Absolute Percentage Error)} = \sum_{i=1}^{no} \left(\frac{|y_i - y'_i| / y_i}{no} \right) \times 100\%.$$

y_i = closing price data sebenarnya.

y'_i = closing price data prediksi.

No = jumlah data di tiap data training, testing, dan validasi.

3.3 Tahap III

Mengukur hasil prediksi.

Langkah-langkah yang dilakukan pada tahap ketiga ini adalah :

- o Masukkan indikator-indikator teknis yang dinormalisasi $X' = (X'_1, X'_2, \dots, X'_{16})$ setiap harinya pada data validasi ke SOM seperti pada tahap pertama.
- o Masukkan indikator-indikator teknis yang dinormalisasi $X' = (X'_1, X'_2, \dots, X'_{16})$ setiap harinya pada data validasi ke dalam model *Genetic Programming (GP)* terbaik sesuai dengan *cluster* datanya, hingga didapatkan formula prediksi terbaik dan hasil prediksi harga penutupan esok hari.
- o Denormalisasi $(X'_1, X'_2, \dots, X'_{16})$ dan y' untuk memperoleh hasil prediksi *Closing Price* yang sesungguhnya untuk esok hari (y').
- o Evaluasi keakuratan hasil prediksi menggunakan SOM-GP melalui RMSE, MAE, dan MAPE.

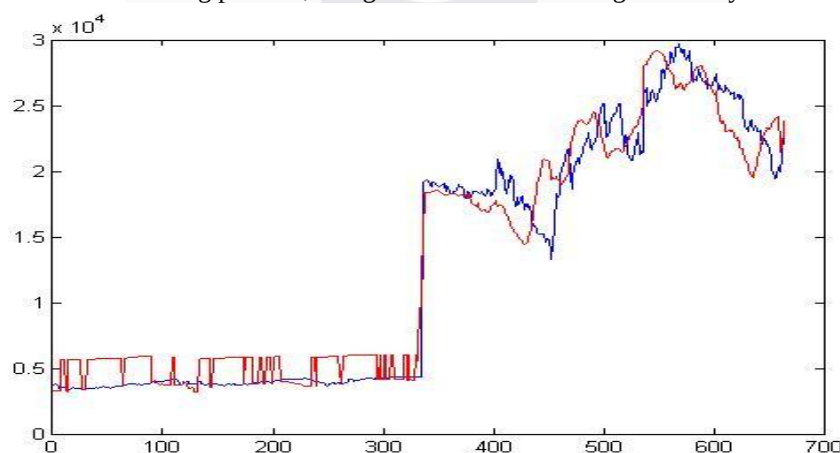
4. Implementasi dan Analisis

4.1. Implementasi

Pada tahap ini, data indeks harga saham IHSG yang telah dihitung 16 indikator inputannya dan dinormalisasi dalam range [-1 1] dikelompokkan ke dalam 5 kelompok menggunakan metode *Self Organizing Map*.

Plot perbandingan data training prediksi dan data training sebenarnya:

Garis merah = data training prediksi, dan garis biru = data training sebenarnya.

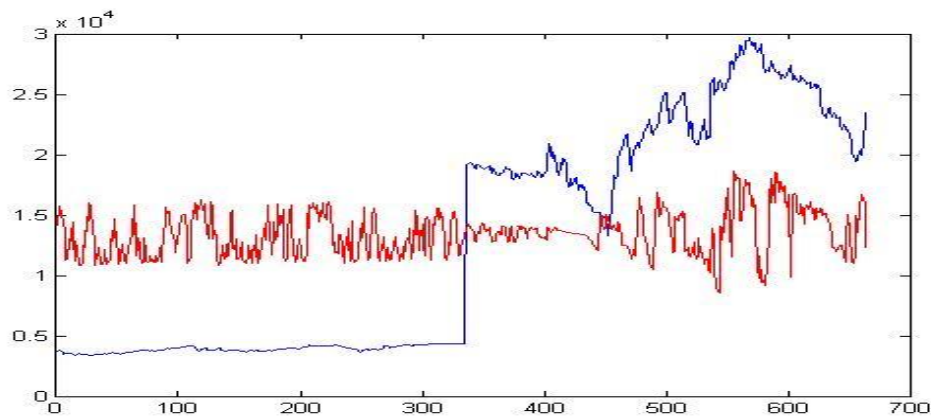


Gambar 1.2 Perbandingan harga prediksi data training indeks harga saham IHSG menggunakan metode SOM-GP sebanyak 664 hari.

Hasil error dari grafik diatas adalah RMSE = 0.0807 dan MAPE = 13.7536%.

Plot perbandingan data training prediksi dan data training sebenarnya:

Garis merah = data training prediksi, dan garis biru = data training sebenarnya.

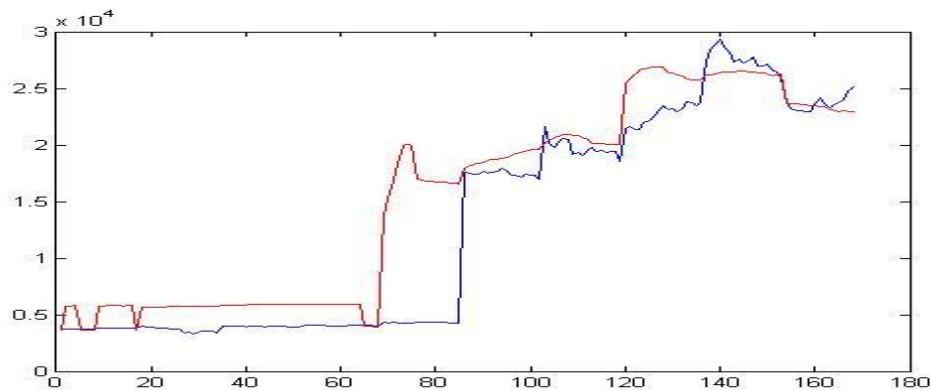


Gambar 1.3 Perbandingan harga prediksi data training indeks harga saham IHSG menggunakan metode GP sebanyak 664 hari.

Hasil error dari grafik diatas adalah RMSE = 0.4397 dan MAPE = 95.0631%.

Plot perbandingan data testing prediksi dan data testing sebenarnya:

Garis merah = data testing prediksi, dan garis biru = data testing sebenarnya.

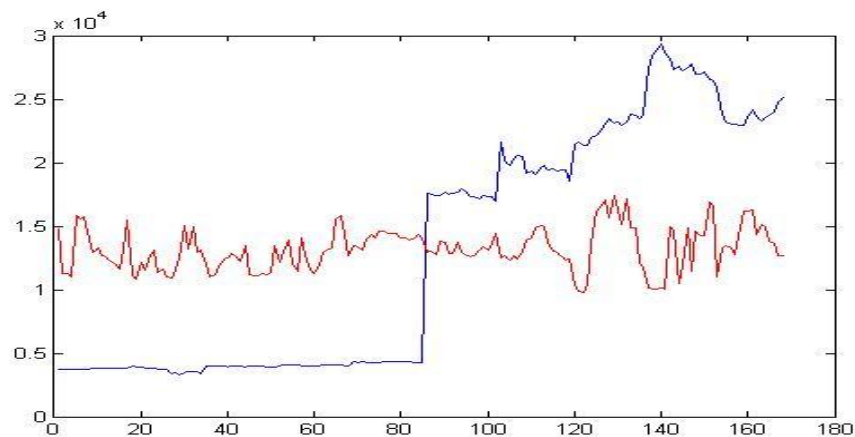


Gambar 1.4 Perbandingan harga prediksi data testing indeks harga saham IHSG menggunakan metode SOM-GP sebanyak 168 hari.

Hasil error dari grafik diatas adalah RMSE = 0.0682 dan MAPE = 12.9248%.

Plot perbandingan data testing prediksi dan data testing sebenarnya:

Garis merah = data testing prediksi, dan garis biru = data testing sebenarnya.

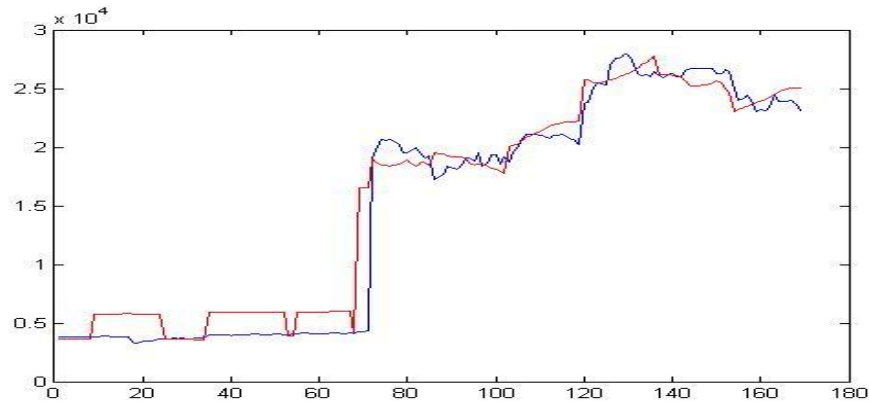


Gambar 1.5 Perbandingan harga prediksi data testing indeks harga saham IHSG menggunakan metode GP sebanyak 168 hari.

Hasil error dari grafik diatas adalah RMSE = 0.4454 dan MAPE = 95.0264%.

Plot perbandingan data validasi prediksi dan data validasi sebenarnya:

Garis merah = data validasi prediksi, dan garis biru = data validasi sebenarnya.

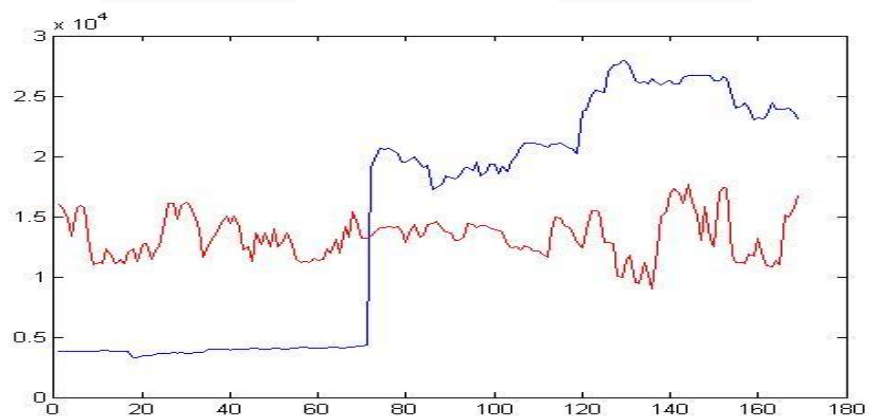


Gambar 1.6 Perbandingan harga prediksi data validasi indeks harga saham IHSG menggunakan metode SOM-GP sebanyak 169 hari

Hasil error dari grafik diatas adalah RMSE = 0.0517 dan MAPE = 9.5406%.

Plot perbandingan data validasi prediksi dan data validasi sebenarnya:

Garis merah = data validasi prediksi, dan garis biru = data validasi sebenarnya.



Gambar 1.7 Perbandingan harga prediksi data validasi indeks harga saham IHSG menggunakan metode GP sebanyak 169 hari.

Hasil error dari grafik diatas adalah RMSE = 0.4702 dan MAPE = 97.4519%.

4.2 Analisis

Berdasarkan hasil dan pengujian yang telah dilakukan maka dapat dianalisis bahwa hasil tersebut sudah cukup akurat untuk memprediksi harga penutupan indeks harga saham Indeks Harga Saham Gabungan dengan masing-masing error. Perbandingan error antara metode SOM-GP dan GP saja dapat dilihat berbeda cukup jauh, yang menyatakan bahwa hasil penggabungan metode Self Organizing Map dan Genetic Programming lebih baik, karena nilai error prediksi yang dihasilkan lebih kecil dibanding dengan metode Genetic Programming saja.

5. Kesimpulan

- i. Hasil prediksi harga penutupan indeks harga saham Indeks Harga Saham Gabungan (IHSG) disimpulkan akurat menggunakan metode gabungan Self Organizing Map dan Genetic Programming (SOM-GP), dengan hasil error :
 - data training: RMSE = 0.0807, MAPE = 13.7536%.
 - data testing: RMSE = 0.0682, MAPE = 12.9248%.
 - data validasi: RMSE = 0.0517, MAPE = 9.5406%.

- ii. Untuk pengujian hanya menggunakan metode Genetic Programming saja tanpa metode SOM untuk indeks harga saham Indeks Harga Saham Gabungan (IHSG), didapatkan hasil error prediksi yang lebih besar dibandingkan menggunakan gabungan metode SOM-GP. Hasil error prediksi Indeks Harga Saham Gabungan dengan menggunakan metode Genetic Programming saja dihasilkan:
data training : RMSE = 0.4397, MAPE = 95.0631%.
data testing : RMSE = 0.4454, MAPE = 95.0264%.
data validasi: RMSE = 0.4702 , dan RMSE = 97.4519%.
- iii. Dengan hasil dan analisis yang telah dilakukan dapat disimpulkan bahwa prediksi indeks harga saham Indeks Harga Saham Gabungan (IHSG) menggunakan metode gabungan Self Organizing Map – Genetic Programming (SOM-GP) menghasilkan harga prediksi yang lebih akurat dibanding dengan metode Genetic Programming saja.

Daftar pustaka

- [1] Basir, Saleh & Hendy M. Fakhruddin. (2005). *Aksi Korporasi, Strategi Untuk Meningkatkan Nilai Saham Melalui Aksi Korporasi*. Jakarta : Salemba Empat.
- [4] Hsu, M. C. (2011). *A hybrid procedure for stock price prediction by integrating self-organizing map and genetic programming*. Taiwan, ROC.
- [3] Kohonen, T. (1989). *Self-organization and associative memory*. Berlin: Springer-Verlag.
- [2] Koza, J. R. (1992) book on genetic programming entitled *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press.
- [5] Suyanto. (2008). *Evolutionary computation – komputasi berbasis “evolusi” dan “genetika”*. Bandung, Informatika.

