ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 4, Nomor 2, Tahun 2015, Halaman 315 - 321
Online di: http://ejournal-s1.undip.ac.id/index.php/gaussian

PREDIKSI INDEKS HARGA SAHAM GABUNGAN MENGGUNAKAN *SUPPORT VECTOR REGRESSION* (SVR) DENGAN ALGORITMA *GRID SEARCH*

Lutfia Septiningrum[1], Hasbi Yasin[2], Sugito[3]
[1]Mahasiswa Jurusan Statistika FSM UNDIP
[2,3]Staf Pengajar Jurusan Statistika FSM UNDIP

ABSTRACT

The existence of capital market Indonesia is one of the important factors in the development of the national economy, proved to have many industries and companies that use these institutions as a medium to absorb investment and media to strengthen its financial position. Capital market Indonesia is an emerging market development is very vulnerable to global economic conditions and capital markets of the world. Prediction JCI (Jakarta Composite Index) is necessary to know the great value that will occur in the future so as investors can take the right policy. To predict in this study used a Support Vector Regression (SVR) method to find the hyperplane in the best regression function to predict the closing price of the JCI using a linear kernel function with output in the form of continuous data. Parameter selection cost and epsilon using a grid search algorithm combined with cross validation and obtained best cost 1 and best epsilon 0.1. While the criteria to measure the goodness of the model is MAPE (Mean Absolute Percentage Error) and R^2 (Coefficient Determination). The results of this study showed that SVR with linear kernel function provides excellent accuracy in the prediction of JCI with R^2 results on training data 98.4% with a MAPE 0.873% while the testing of data R^2 90.9% with a MAPE 0.613%.

Keywords: JCI, Support Vector Regression (SVR), Hyperplane, Kernel Linear, Grid Search Algorithm, Cross Validation, Accuracy

1. PENDAHULUAN

Pasar modal yang ada di Indonesia merupakan pasar yang sedang berkembang yang dalam perkembangannya sangat rentan terhadap kondisi ekonomi global dan pasar modal dunia. Sehingga prediksi Indeks Harga Saham Gabungan (IHSG) perlu dilakukan untuk mengetahui besar nilai yang akan terjadi diwaktu mendatang agar sebagai investor atau penanam modal dapat mengambil kebijakan yang tepat kedepannya kelak.

Indikator utama yang mencerminkan kinerja pasar modal apakah sedang mengalami peningkatan (*bullish*) ataukah sedang mengalami penurunan (*bearish*) yaitu IHSG. Karena IHSG ini mencatat pergerakan harga saham dari semua sekuritas yang tercatat di Bursa Efek Indonesia. Sehingga pergerakan IHSG menjadi perhatian bagi semua investor di Bursa Efek Indonesia, sebab pergerakan IHSG ini akan mempengaruhi sikap para investor apakah akan membeli, menahan ataukah menjual sahamnya. Selain itu kenaikan dan penurunan IHSG merupakan sebuah ukuran atas persepsi pasar di luar kenaikan dan penurunan nilai tukar valuta asing terhadap rupiah (Manurung, 2004).

Support Vector Machine (SVM) digunakan untuk mengatasi malah asumsi. SVM yang digunakan untuk kasus regresi dinamakan *Support Vector Regression* (SVR). Konsep *loss function* pada SVR dapat digunakan untuk kasus regresi. Beberapa jenis *loss function* adalah *ε-insensitive, quadratic, Huber dan Lapace*. SVR digunakan untuk meramalkan pergerakan

Indeks Harga Saham Gabungan (IHSG) dengan kriteria keakuratan yang digunakan adalah MAPE (*Mean Absolute Percentage Error*) dan R^2 (Koefisien Determinasi), sedangkan untuk pemilihan parameter terbaik pada penelitian ini digunakan algoritma *grid search* dengan tipe *cross validation Leave One Out*.

2. TINJAUAN PUSTAKA

2.1. Pengertian Indeks Harga Saham Gabungan (IHSG)

IHSG adalah angka yang menunjukkan situasi pasar modal atau peristiwa yang terjadi di pasar modal dalam situasi yang berbeda dari waktu ke waktu, yang dapat digunakan sebagai alat analisis. Kata gabungan itu sendiri dari IHSG adalah kinerja saham yang di dalamnya terdapat lebih dari satu 1 (satu) saham, bahkan semua saham yang tercatat di bursa tersebut.

2.2. Konsep *Support Vector Regression* (SVR)

Support Vector Regression (SVR) merupakan pengembangan SVM untuk kasus regresi. Tujuan dari SVR adalah untuk menemukan sebuah fungsi $f(x)$ sebagai suatu *hyperplane* (garis pemisah) berupa fungsi regresi yang mana sesuai dengan semua input data dengan sebuah *error* ε dan membuat ε setipis mungkin (Scholkopt and Smola, 2002). Misalkan ada l data training, $(\boldsymbol{x_i}, y_i)$, $i = 1,..,l$ dimana $\boldsymbol{x}_i$ merupakan vektor input $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots \boldsymbol{x}_n\} \subseteq \boldsymbol{\Re}^n$ dan output skalar $y = \{y_i, \dots y_l\} \subseteq \Re$ dan l adalah banyaknya data *training*. Dengan SVR, ingin ditentukan suatu fungsi $f(x)$ yang mempunyai deviasi paling besar ε dari target aktual y_i, untuk semua data *training*. Jika nilai ε sama dengan 0 maka diperoleh suatu persamaan regresi yang sempurna (Santosa, 2007).

Tujuan dari SVR ini adalah untuk memetakan *vector* input ke dalam dimensi yang lebih tinggi Abe (2005). Misalkan sebuah fungsi berikut adalah garis regresi sebagai *optimal hyperplane* :

$$f(\boldsymbol{x}) = \boldsymbol{w}^T \varphi(\boldsymbol{x}) + b$$

Dengan :

$\boldsymbol{w}$ = vektor bobot berdimensi *l*

$\varphi(\boldsymbol{x})$ = fungsi yang memetakan *x* pada ruang dengan *l* dimensi

b = bias

2.3. Optimasi Persamaan *Hyperplane* SVR

1. Quadratic Programming

Optimasi penyelesaian masalah dengan bentuk *Quadratic Programming* :

$$min \frac{1}{2} \parallel \boldsymbol{w} \parallel^2 \quad (1)$$

dengan kendala:

$$y_i - w^T \varphi(\boldsymbol{x_i}) - b \le \varepsilon \text{ untuk } i = 1, \dots, l \quad (2)$$

$$w^T \varphi(\boldsymbol{x_i}) - y_i + b \le \varepsilon \text{ untuk } i = 1, \dots, l$$

Faktor $||\boldsymbol{w}||^2$ dinamakan *regulasi*. Meminimalkan $||\boldsymbol{w}||^2$ akan membuat suatu fungsi setipis *(flat)* mungkin, sehingga bisa mengontrol kapasitas fungsi *(function capacity)*.

Pada Persamaan (1) diasumsikan bahwa semua titik ada dalam rentang $f(\boldsymbol{x}) \pm \varepsilon$ (*feasible*), dalam hal ketidaklayakan *(infiesibility)*, dimana ada beberapa titik yang mungkin keluar dari rentang $f(\boldsymbol{x}) \pm \varepsilon$ maka ditambahkan variabel *slack* ξdan ξ^* untuk mengatasi masalah pembatasan yang tidak layak (*infeasible constraints)* dalam problem optimasi (Santosa, 2007).

semua titik diluar margin akan dikenai pinalti. Selanjutnya problem optimisasi di atas bisa diformulasikan sebagai berikut :

$$min\frac{1}{2} \parallel \boldsymbol{w} \parallel^2 + \mathrm{C} \sum_{i=1}^{\ell} (\xi_i + \xi_i^*) \quad (3)$$

Dengan kendala:

$$y_i - \boldsymbol{w}^T \varphi(\boldsymbol{x_i}) - b - \xi_i \leq \varepsilon, \quad i = 1, .., l \quad (4)$$

$$\boldsymbol{w}^T \varphi(\boldsymbol{x_i}) - y_i + b - \xi_i^* \leq \varepsilon, \quad i = 1, .., l$$

$$\xi_i, \xi_i^* \geq 0$$

2. *Lagrange Multiplier*

Solusi optimasi untuk persamaan (3) dengan batas bawah persamaan (4) adalah dengan fungsi *Lagrange* berikut :

$$Q(\boldsymbol{w}, b, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*) = L$$

$$L = \frac{1}{2} \|\boldsymbol{w}\|^2 + C \sum_{i=1}^{l} (\xi_i + \xi_i^*) - \sum_{i=l}^{l} \alpha_i (\varepsilon + \xi_i - y_i + \boldsymbol{w}^T \varphi(\boldsymbol{x}) + b) - \sum_{i=1}^{l} \alpha_i^* (\varepsilon + \xi_i^* + \quad y_i - \boldsymbol{w}^T \varphi(\boldsymbol{x}) - b) - \sum_{i=1}^{l} (\eta_i \xi_i + \eta_i^* \xi_i^*) \quad (5)$$

L dinamakan *Lagrangian*, η_i, η_i^*, α_i, α_i^* adalah *Lagrange Multiplier*. *Lagrange Multiplier* digunakan untuk menyelesaikan maslah optimasi dengan kendala*(constrined optimization)*. Di mana kendalanya akan dikonversi menjadi masalah optimasi tanpa kendala*(unconstrained optimization)*

3. Menentukan Nilai w dan b

Untuk mencari nilai parameter persamaan *optimal hyperplane* dapat ditentukan dari turunan parsial pengali lagrange*(Lagrange Multiplier)* terhadap $\boldsymbol{w}, b, \xi, \xi^*$

$$\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) = 0 \quad (6)$$

$$\frac{\partial L}{\partial w} = \boldsymbol{w} - \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) \varphi(\boldsymbol{x_i}) = 0 \quad (7)$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0 \quad (8)$$

$$\frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0 \quad (9)$$

Dari Persamaan (7) $\boldsymbol{w}$ dapat ditulis dengan :

$$\boldsymbol{w} = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) \varphi(\boldsymbol{x}_i) \quad (10)$$

w dapat dideskripsikan secara lengkap yaitu sebuah kombinasi linier dari pola data *training*. Solusi dual untuk persamaan (5) diperoleh dengan mensubstitusikan Persamaan (6), (7), (8), (9) ke dalam persamaan (5) dengan memaksimalkan solusi dual $Q(\alpha, \alpha^*)$.

Untuk mencari solusi optimal parameter b dengan menggunakan syarat KKT (*Karush Kuhn Tucker*) sehingga diperoleh solusi optimal parameter b adalah

$$b = y_i - \boldsymbol{w}^T \varphi(\boldsymbol{x_i}) - \varepsilon \text{ untuk } 0 < \alpha_i < C \quad (11)$$

$$b = y_i - \boldsymbol{w}^{\boldsymbol{T}} \varphi(\boldsymbol{x}_i) + \varepsilon \text{ untuk } 0 < \alpha_i^* < C \quad (12)$$

2.4. Fungsi Kernel

Banyak teknik data mining atau *machine learning* yang dikembangkan dengan asumsi kelinieran, sehingga algoritma yang dihasilkan terbatas untuk kasus-kasus yang linier. Dengan metode kernel (K*ernel Trick*) suatu data x di *input space* dipetakan ke *feature space* dengan dimensi yang lebih tinggi melalui φ (Santosa,2007).

- Linier : $\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}.\boldsymbol{y}$
- *Polynomial* : $\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}.\boldsymbol{y} + \boldsymbol{c})^{\boldsymbol{d}}$
- *Radial Basis Function* (RBF) : $\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{y}) = exp(-\gamma \|\boldsymbol{x} - \boldsymbol{y}\|^2)$, dengan $\gamma = \frac{1}{2\sigma^2}$
- *Tangent hyperbolic* (sigmoid): $\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{y}) = tanh(\sigma(\boldsymbol{x}.\boldsymbol{y}) + c)$

$\boldsymbol{x}$ dan $\boldsymbol{y}$ adalah pasangan dua data dari semua bagian data latih. Parameter $\sigma, c, d > 0$, merupakan konstanta. Menurut Vapnik (1995) dan Haykin (1999) dalam Santosa (2007), fungsi kernel yang legitimate diberikan oleh Teori Mercer di mana fungsi tersebut harus memenuhi syarat kontinu dan positif definite.

2.5. Seleksi Parameter

1 *Cross-Validation*

Menurut Leidiyana (2013), *cross-validation* adalah pengujian standar yang dilakukan untuk memprediksi *error rate*. Data *training* dibagi secara random ke dalam beberapa bagian dengan perbandingan yang sama kemudian *error rate* dihitung bagian demi bagian, selanjutnya hitung rata-rata seluruh *error rate* untuk mendapatkan *error rate* secara keseluruhan. Laju error dapat dihitung dengan formulasi berikut:

Dalam *cross-validation*, dikenal validasi *leave-one-out* (LOO). Dalam LOO, data dibagi ke dalam 2 subset, subset 1 berisi N-1 data untuk *training* dan satu data sisanya untuk *testing* (Santosa, 2007).

2 *Grid Search*

Salah satu algoritma untuk menentukan parameter optimal pada model SVR adalah menggunakan algoritma *grid search*. Algoritma ini membagi jangkauan parameter yang akan dioptimalkan kedalam grid dan melintasi semua titik untuk mendapatkan parameter yang optimal . Dalam aplikasinya, algoritma *grid search* harus dipandu oleh beberapa metrik kinerja, biasanya diukur dengan *cross-validation* pada data *training*. Oleh karena itu, disarankan untuk mencoba beberapa variasi pasangan parameter pada *hyperplane* SVR (Yasin.*et al*, 2014)

2.6. Pengukuran Kinerja Prediksi

Dalam kasus regresi ada beberapa ukuran error yang sering dipakai untuk menilai suatu perfomansi fungsi prediksi. Ukuran error yang digunakan dalam penelitian ini adalah nilai MAPE (Mean Absolute Percentage Error), sehingga formula dari MAPE dapat dinyatakan sebagai berikut:

$$\text{MAPE}=\frac{\sum_{i=1}^{m} APE}{m}$$

dimana :

$$\text{APE}=\frac{\sum_{i=1}^{m}|yi-\widehat{yi}|}{yi}\ x\ 100$$

Sedangkan untuk tingkat akurasi suatu model prediksi digunakan Koefisien determinasi (R^2)

$$R^2= 1-\frac{JKE}{JKT}$$

3. METODOLOGI PENELITIAN

Data yang digunakan dalam penelitian ini merupakan data sekunder, yaitu data historis yang diambil melalui situs resmi (www.idx.co.id). Data tersebut adalah data Indeks Harga Saham Gabungan (IHSG) yang berupa data *time series* mulai tanggal 3 Januari 2011 sampai dengan tanggal 30 November 2014 pada hari kerja.

Langkah – langkah yang dilakukan pada penelitian ini adalah sebagai berikut :

1. Identifikasi permasalahan, yaitu dalam penelitian ini adalah IHSG.
2. Analisis Deskriptif
3. Menentukan variabel penelitian, yaitu Indeks Harga Saham Gabungan (IHSG).
4. Mengumpulkan data IHSG periode 3 Januari 2011 sampai dengan 30 November 2014.
5. Menentukan variabel independen menggunakan plot PACF.
6. Melakukan analisis SVR untuk menghasilkan pemodelan menggunakan softwere R 3.1.1.
   Melakukan analisis data yang terdiri :
   a. Membagi data menjadi data *training* dan data *testing.*
   b. Memilih parameter terbaik menggunakan algoritma *grid search*
   c. Melakukan analisis SVR dengan fungsi *ε-insensitive Loss Function* dan Fungsi *Kernel Linier*.
   d. Mengukur Model terbaik dengan menggunakan MAPE dan R^2
   e. Melakukan prediksi menggunakan model terbaik yang dihasilkan pada analisis SVR.

4. ANALISIS DAN PEMBAHASAN

4.1. Penentuan Parameter Terbaik Untuk *Hyperplane*

Tabel 1 *Grid Search – Cross Validation*

	cost	epsilon	error
1	0.1	0.1	2689.083
2	**1**	0.1	**2652.666**
3	10	0.1	2659.174
4	100	0.1	2661.315

Dari Tabel 1 diperoleh nilai C dan epsilon terbaik dengan melihat nilai error terkecil dari proses *Grid Search cross- validation*. Cross-Validation yang digunakan pada *default* program R merupakan LOO atau *Leave One Out* sehingga diperoleh nilai parameter terbaik yaitu C=1 dan epsilon=0.1. sedangkan nilai w dan bias dihitung menggunakan bantuan paket program R 3.1.2, sehingga hasil yang diperoleh sebagai berikut:

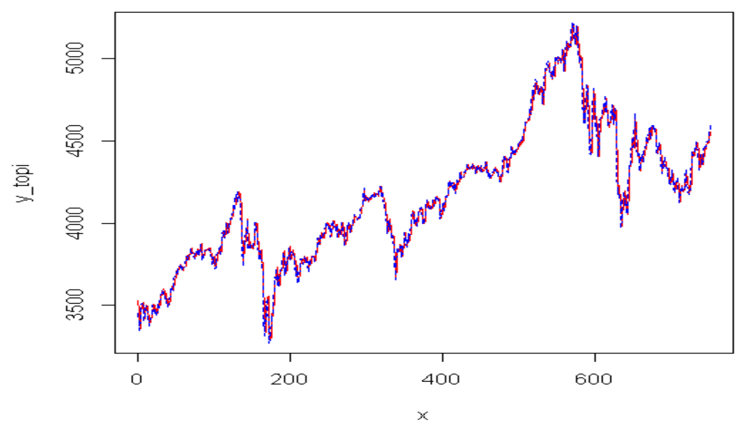
Tabel 2 Nilai w Dan Bias

variabel	w	b
x1	0,96434	
x4	0,01753	0,00030
x13	0,00973	

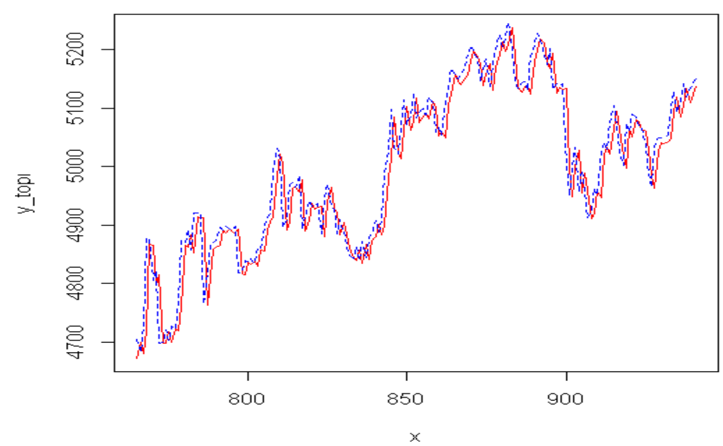
Dari Tabel 2 dapat dilihat bahwa nilai beta dan w akan digunakan dalam mencari *support vector*, *support vector* yang dihasilkan nantinya akan digunakan untuk memprediksi menggunakan persamaan SVR yang terbentuk.

4.2. Akurasi Prediksi

Gambar 1. Grafik Prediksi Data *Training*



Gambar2. Grafik Prediksi Data *Testing*



Perhitungan akurasi dan kesalahan prediksi dengan menggunakan data *training* menghasilkan R^2 sebesar 0,98431 atau 98,431% sedangkan untuk nilai MAPE sebesar 0.873%. Sedangkan pada data *testing* menghasilkan R^2 sebesar 0.90981 atau 90.981% sedangkan untuk nilai MAPE sebesar 0.613%. Hal ini menunjukkan untuk data *testing* menggunakan fungsi pemetaan kernel linier mempunyai hasil akurasi atau ketepatan prediksi yang cukup besar, untuk nilai errornya atau kesalahan yang terjadi pada proses prediksi relatif kecil

4.3. Hasil Prediksi 10 Periode Kedepan

Tabel 3. Hasil Prediksi

No	Prediksi
1	5136,605
2	5140,778
3	5128,223
4	5132,466
5	5120,492
6	5124,816
7	5113,597
8	5117,213
9	5106,655
10	5110,216

Hasil Prediksi yang diperoleh menunjukkan harga penutupan IHSG dari periode ke-1 sampai dengan periode ke-10 yang mengalami mengalami fluktuasi. Hasil ini dapat menggambarkan keadaan pasar modal yang tidak stabil. Dengan menggunakan metode SVR untuk memprediksi nilai penutupan IHSG dapat mempermudah investor dalam melihat pergerakan pasar modal

5. KESIMPULAN

Prediksi Indeks Harga Saham Gabungan menggunakan metode *Support Vector Regression* (SVR) dengan parameter Cost dan epsilon terbaik berturut-turut adalah 1 dan 0.1 dengan data training yang digunakan sebanyak 751 data menghasilkan akurasi yang cukup baik yaitu dihasilkan akurasi sebesar 98.431% dengan error menggunakan MAPE sebesar 0.873%. Pengujian dengan data testing sebanyak 177 data, akurasi prediksi yang didapat pada fungsi kernel linier yaitu 90.981% dengan error menggunakan MAPE sebesar 0.613%.

DAFTAR PUSTAKA

Gunn, S. R. 1998. *Support Vector Machines for Classification and Regression*. Technical Report. University of Southampton.

Leidiyana, H.2013. *Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bemotor*.Jurnal Penelitian Ilmu Komputer, System Embedded & Logic 1(1) : 65-76

Santosa, B.2007.*Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis.*Yogyakarta:Graha Ilmu

Yasin. H., Prahutama.A.,dan Utami,T.A. 2014. *Prediksi Harga Saham Menggunakan Support Vector Regresion Dengan Algoritmaa Grid Search*. Media Statistika,Vol 7, No 1:29-35