

## **PREDIKSI HARGA SAHAM MENGGUNAKAN *SUPPORT VECTOR REGRESSION* DENGAN ALGORITMA *GRID SEARCH***

**Hasbi Yasin<sup>1</sup>, Alan Prahutama<sup>2</sup>, Tiani Wahyu Utami<sup>3</sup>**

<sup>1,2</sup>Dosen Jurusan Statistika Undip

<sup>3</sup>Dosen Program Studi Statistika Unimus

Email: [hasbiyasin@undip.ac.id](mailto:hasbiyasin@undip.ac.id)

### **Abstract**

The stock market has become a popular investment channel in recent years because of the low return rates of other investment. The stock price prediction is in the interest of both private and institution investors. Accurate forecasting of stock prices is an appealing yet difficult activity in the business world. Therefore, stock prices forecasting is regarded as one of the most challenging topics in business. The forecasting techniques used in the literature can be classified into two categories: linear models and non linear models. One of forecasting techniques in nonlinear models is support vector regression (SVR). Basically, SVR adopts the structural risk minimization principle to estimate a function by minimizing an upper bound of the generalization. The optimal parameters of SVR can be use Grid Search Algorithm method. Concept of this method is using cross validation (CV). In this paper, the SVR model use linear kernel function. The accurate prediction of stock price, in telecommunication, is 92.47% for training data and 83.39% for testing data.

Keywords: Stock price, SVR, Grid Search, Linear kernel function.

### **1. Pendahuluan**

Pasar Modal merupakan sistem keuangan yang terorganisasi terdiri dari bank komersial, lembaga perantara di bidang keuangan dan seluruh surat-surat berharga yang beredar. Salah satu manfaat dari pasar modal adalah menciptakan kesempatan kepada masyarakat untuk berpartisipasi dalam kegiatan perekonomian khususnya dalam berinvestasi<sup>[1]</sup>. Salah satu asset untuk investasi adalah saham. Saham merupakan surat berharga yang dikeluarkan oleh sebuah perusahaan. Pendapatan yang didapat dari pemegang saham, tergantung dari perusahaan yang menerbitkan saham (emiten). Jika emiten mampu menghasilkan keuntungan yang besar maka keuntungan yang didapatkan oleh pemegang saham juga akan besar. Semakin tinggi keuntungan yang ditawarkan, semakin tinggi risiko yang akan dihadapi dalam berinvestasi<sup>[11]</sup>. Oleh karena itu perlu diprediksi harga saham sekarang yang didasarkan atas dasar harga saham kemarin.

Prediksi untuk data time series dibedakan menjadi dua yaitu dengan pendekatan linier dan non linier. Pendekatan linier salah satunya dengan menggunakan metode *Autoregressive Integreted Moving Average* (ARIMA). Pendekatan non linier sudah banyak dikembangkan antara lain *Neural Network*, *Fuzzy*, *Adaptive Neuro Fuzzy Inference Sistem* (ANFIS) dan *Support Vector Machine* (SVM). SVM merupakan pendekatan non linier yang berbasis *machine learning*. Salah satu modifikasi SVM yang digunakan untuk pendekatan regresi adalah *Support Vector Regression* (SVR). Konsep dari SVR adalah memaksimalkan *hyperplane* untuk mendapatkan data-data yang menjadi *support vector*. Keunggulan SVR salah satunya adalah mampu mengatasi *overfitting*. Permasalahan yang sering dialami ketika menggunakan SVR adalah pada saat penentuan parameter model yang optimal. Salah satu cara yang dapat digunakan adalah dengan metode *grid search*.

Metode ini mampu mendapatkan parameter SVR yang optimal<sup>[16]</sup>. Penelitian-penelitian mengenai SVM telah banyak dilakukan antara lain tentang peramalan curah hujan menggunakan teknik *machine learning* oleh Hong<sup>[6]</sup>. Sementara itu, Chen<sup>[3]</sup> meneliti tentang Aplikasi SVR menggunakan algoritma GASA pada data permintaan handphone 3G penduduk Taiwan. Goyaletal.<sup>[4]</sup> melakukan pemodelan iklim di subtropics menggunakan LS-SVR. Trapsilasiwi<sup>[13]</sup> meneliti tentang peramalan beban listrik di pulau Jawa – Bali menggunakan *Support Vector Machine* (SVM). Sedangkan penelitian mengenai prediksi harga saham sudah banyak dilakukan antara lain oleh Kao et al.<sup>[8]</sup> tentang penggabungan analisis komponen independen dengan SVR untuk peramalan harga saham. Wang et al.<sup>[15]</sup> melakukan peramalan harga saham dengan pendekatan model hybrid. Kazem et al.<sup>[9]</sup> meramalkan harga saham menggunakan SVR dengan algoritma Firefly. Oleh karena itu ingin dikaji tentang prediksi harga saham menggunakan SVR dengan algoritma *grid search*.

## 2. Tinjauan Pustaka

### 2.1. Support Vector Regression (SVR)

SVR merupakan bagian dari SVM yang diperkenalkan oleh Vapnik (1995). SVM adalah sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur (*feature space*) berdimensi tinggi, dilatih dengan algoritma pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan learning bias. Konsep SVM menggunakan konsep  $\varepsilon$ -insensitive loss function. SVM dapat digeneralisasi untuk melakukan pendekatan fungsi yang dikenal dengan SVR<sup>[5]</sup>. Konsep SVM menggunakan *hyperplane* tunggal pada ruang berdimensi banyak yang pada akhirnya partisi-partisi tersebut dapat diselesaikan secara non linier. *Hyperplane* yang optimum dapat diselesaikan dengan metode *quadratic programming*<sup>[2]</sup>. Konsep SVR didasarkan pada *risk minimization*, yaitu untuk mengestimasi suatu fungsi dengan cara meminimalkan batasatas dari *generalization error*, sehingga SVR mampu mengatasi *overfitting*. Fungsi regresi dari metode SVR adalah sebagai berikut:

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

dengan  $w$  merupakan vector pembobot,  $\varphi(x)$  merupakan fungsi yang memetakan  $x$  dalam suatu dimensi dan  $b$  merupakan bias. Untuk memaksimalkan *hyperplane* dengan meminimalkan nilai *Loss function*

$$R(f(x)) = \frac{1}{2} \|w\|^2 + \frac{C}{n} \sum_{i=1}^n L_{\varepsilon}(y_i, f(x_i))$$

dimana

$$L_{\varepsilon}(y_i, f(x_i)) = \begin{cases} 0 & ; |y_i - f(x_i)| \leq \varepsilon \\ |y_i - f(x_i)| - \varepsilon; & \text{lainnya} \end{cases}$$

dengan  $L_{\varepsilon}$  merupakan  $\varepsilon$ -insensitive loss function,  $C$  dan  $\varepsilon$  merupakan parameter. Konsep dari kuadrat loss function adalah meminimumkan nilai sebagai berikut:

$$R(w, \xi, \xi^*) = \frac{1}{2} \|w\|^2 + C \left( \sum_{i=1}^n (\xi_i + \xi_i^*) \right)$$

dengan batasan  $w\varphi(x_i) + b - y_i \leq \varepsilon + \xi_i^*$ ;  $y_i - w\varphi(x_i) - b \leq \varepsilon + \xi_i$  and  $\xi_i, \xi_i^* \geq 0$ . Dengan menggunakan penyelesaian lagrange dalam bentuk:

$$L(w, b, \xi, \xi^*, \alpha_i, \alpha_i^*, \beta_i, \beta_i^*) = \frac{1}{2} \|w\|^2 + C \left( \sum_{i=1}^n (\xi_i + \xi_i^*) \right) - \sum_{i=1}^n \alpha_i [w\varphi(x_i) + b - y_i + \varepsilon + \xi_i]$$

$$-\sum_{i=1}^n \alpha_i^* [y_i - w \varphi(x_i) - b_i + \varepsilon + \xi_i^*] - \sum_{i=1}^n (\beta_i \xi_i + \beta_i^* \xi_i^*)$$

Dengan menggunakan pendekatan Karush-Kuhn-Tuck didapatkan sebagai berikut:

$$Q(\alpha, \alpha^*) = -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*)$$

dengan batasan  $\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0$ ;  $0 \leq \alpha_i \leq C$ ;  $0 \leq \alpha_i^* \leq C$  dimana  $K(x_i, x_j)$  merupakan fungsi kernel. Penentuan nilai parameter C dan fungsi kernel sangat penting dalam menentukan tingkat keakuratan dari prediksi.

## 2.2. Fungsi Kernel

Beberapa metode dalam analisis data mining banyak menggunakan fungsi linier. Banyak kasus di dunia nyata merupakan kasus yang nonlinier, sehingga untuk mengatasinya dengan cara mentransformasi data kedalam dimensi ruang yang lebih tinggi. SVM dapat digunakan pada data nonlinier dengan menggunakan pendekatan kernel sehingga dapat dipisahkan secara linier pada *feature space* yang baru. Fungsi kernel yang digunakan pada metode SVR adalah sebagai berikut Haykin:

1. Linier:  $\mathbf{x}^T \mathbf{x}$
2. Polinomial:  $(\mathbf{x}^T \mathbf{x} + 1)^n$
3. Radial Basis Function (RBF):  $\exp\left(-\frac{1}{2\sigma^2} \|x - x_i\|^2\right)$

Pemilihan fungsi kernel yang tepat merupakan hal yang penting, karena akan menentukan *feature space*.

## 2.3. Algoritma 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<sup>[16]</sup>. 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<sup>[7]</sup>. Pasangan parameter yang menghasilkan akurasi terbaik yang didapatkan dari uji *cross-validation* merupakan parameter yang optimal. Parameter optimal tersebut yang selanjutnya digunakan untuk model SVR terbaik. Setelah itu, model SVR tersebut digunakan untuk memprediksi data *testing* untuk mendapatkan generalisasi tingkat akurasi model.

Menurut Leidiyana<sup>[10]</sup>, *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. Dalam *cross-validation*, dikenal validasi *leave-one-out* (LOO). Dalam LOO, data dibagi kedalam 2 subset, subset 1 berisi N-1 data untuk *training* dan satu data sisanya untuk *testing*<sup>[12]</sup>.

$$CV = \sum_{i=1}^n (y_i - \hat{y}_{\neq i})^2$$

dengan:  $\hat{y}_{\neq i}$  : Nilai penaksir  $y_i$  (*fitting value*) dimana pengamatan ke  $i$  dihilangkan dari proses penaksiran  
 $y_i$  : Nilai aktual  $y$  pada pengamatan ke  $i$

### 3. Metodologi Penelitian

Data yang digunakan dalam tulisan ini adalah data sekunder, yaitu data saham harian PT XL Axiata Tbk periode 1 Januari 2013 sampai dengan 30 September 2014 yang terdiri dari 455 data saham harian yang tercatat. Data akan dibagi menjadi data *training* sebanyak 400 data, sedangkan sisanya sebagai data *testing*. Pada penelitian ini, harga saham hari ini diasumsikan hanya dipengaruhi oleh harga saham sehari sebelumnya.

Langkah-langkah analisis:

1. Mempersiapkan data saham harian PT XL Axiata Tbk yang terdiri dari variabel independen adalah  $y_{t-1}$  dan variabel dependen adalah  $y_t$ .
2. Membagi data tersebut menjadi data *training* dan data *testing* dengan persentase proporsi tertentu.
3. Melakukan pemodelan harga saham menggunakan metode SVR dengan penentuan parameter terbaik menggunakan algoritma *grid search*, dengan tahapan sebagai berikut:
  - a. Menentukan fungsi kernel, nilai-nilai parameter kernel dan parameter *cost* dan parameter *epsilon* untuk optimasi *hyperplane* pada data *training*.
  - b. Memilih nilai parameter terbaik menggunakan metode *grid-search* pada tiap *hyperplane* dengan fungsi kernel *linear* dengan bantuan paket program R 3.0.2, sebagai berikut:
    - 1) Membagi data *training* sebanyak baris data.
    - 2) Gunakan tiap bagian sebagai data *testing* dan bagian lainnya sebagai data *training*.
    - 3) Optimasi parameter  $\alpha$  dan  $b$  untuk setiap *hyperplane* dengan metode SMO.
    - 4) Menggunakan *hyperplane-hyperplane* tersebut untuk tiap model regresi.
    - 5) Hitung error tiap model regresi.
    - 6) Hitung rata-rata error hasil regresi untuk setiap *hyperplane*.
    - 7) Menentukan *hyperplane* terbaik.
  - c. Menggunakan *hyperplane* dengan parameter terbaik yang diperoleh pada regresi data *testing*.
  - d. Evaluasi model regresi pada *testing* menggunakan koefisien determinasi ( $R^2$ ).

### 4. Hasil dan Pembahasan

#### 4.1. Penentuan Fungsi Kernel dan Parameter Terbaik untuk *Hyperplane*

Pada penelitian ini hanya digunakan fungsi kernel linier pada fungsi pemisah (*hyperplane*) SVR. Parameter terbaik pada fungsi kernel tersebut ditentukan dengan mencobakan beberapa nilai pada rentang tertentu untuk membangun *hyperplane*. Parameter yang dioptimalkan adalah nilai  $C$  dan nilai epsilon. Untuk menentukan parameter terbaik pada pemodelan harga saham ini digunakan metode *grid search* yang dipadukan dengan metode *leave-one-outcross-validation* untuk menghitung nilai error model yang dicobakan pada data *training*. Parameter yang terbaik untuk *hyperplane* ditentukan dengan nilai error terkecil. Dari seleksi parameter yang dilakukan, didapatkan parameter terbaik untuk *hyperplane* dengan fungsi kernel linier adalah  $C = 0,1$  dan epsilon =  $0,1$ .

**Tabel 1.** Error Model SVR untuk Penentuan Parameter Optimal dengan Metode *Grid Search*

Parameter <i>hyperplane</i>		Error	Dispersi
C	Epsilon		
0,1	0,1	<b>12823,80</b>	23025,09
0,1	0,01	12880,26	23026,63
0,1	0,001	12880,32	23030,99
0,1	0,0001	13104,30	23100,03
1	0,1	12909,85	22832,47
1	0,01	12895,54	22810,81
1	0,001	13167,85	23050,95
1	0,0001	13017,25	22828,78
10	0,1	13018,68	22832,54
10	0,01	13286,11	23156,85
10	0,001	13031,08	22834,80
10	0,0001	13031,01	22835,00

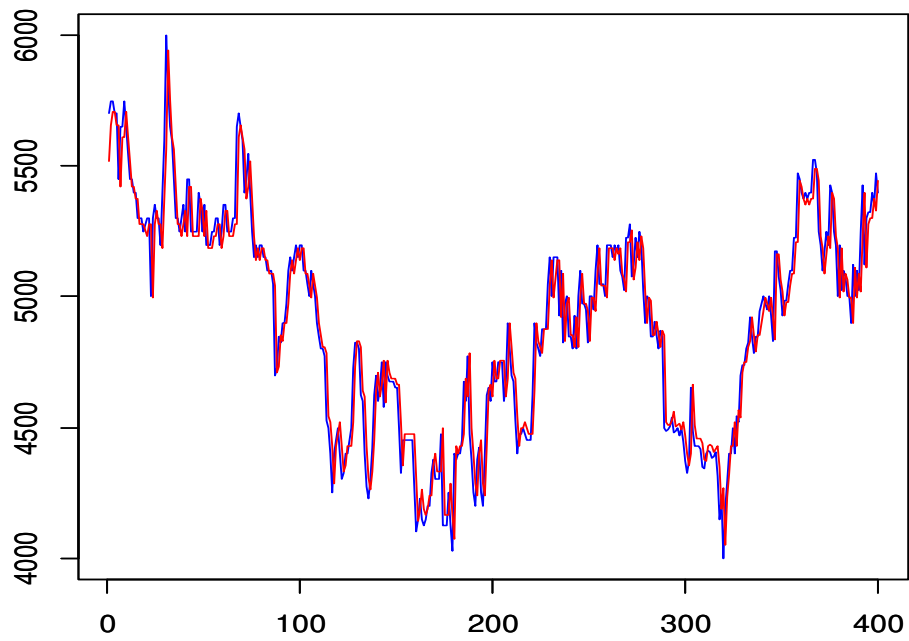
#### 4.2. Validasi Model SVR

Parameter terbaik yang diperoleh dari penerapan metode *grid search* pada data *training* selanjutnya diterapkan pada data *testing* sebagai evaluasi terhadap hasil klasifikasi yang ditunjukkan dengan nilai akurasi. Perbandingan akurasi model untuk data *training* dan data *testing* dengan parameter terbaik yang telah ditentukan sebelumnya ditunjukkan pada Tabel 2.

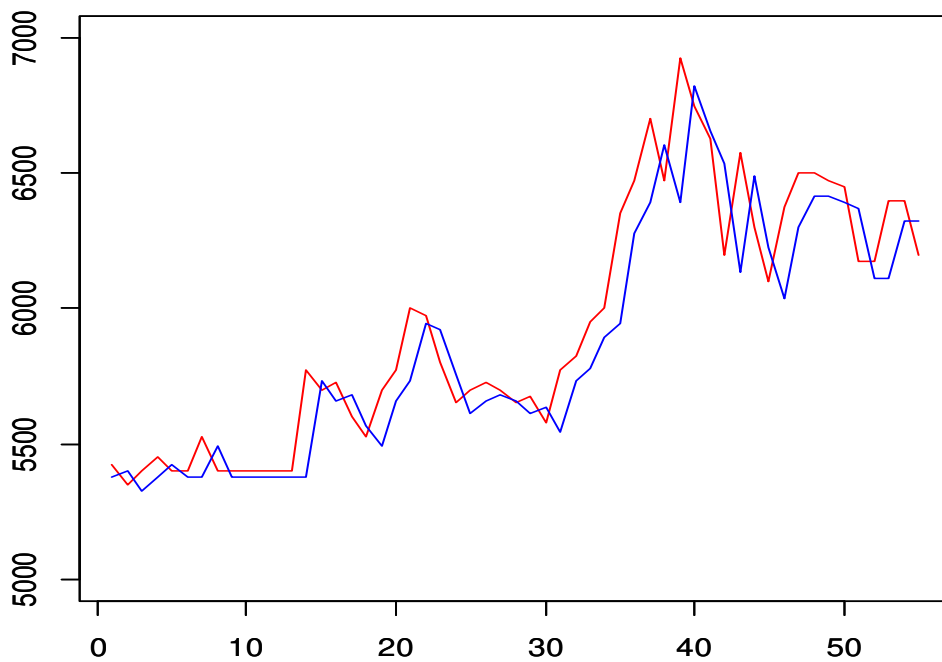
**Tabel 2.** Akurasi Model SVR untuk Data Harga Saham

Data	Koefisien Determinasi
Training	0,9247005
Testing	0,8339279

Secara visual hasil prediksi data training dan data testing dapat dilihat pada Gambar 1 dan Gambar 2. Dalam gambar tersebut terlihat bahwa pola data sudah mengikutipola yang sama sehingga model SVR yang diperoleh layak digunakan untuk prediksi harga saham PT. XL Axiata Tbk.



Gambar 1. Plot Hasil Prediksi dan Aktual dari Data *Training*



Gambar 2. Plot Hasil Prediksi dan Aktual dari Data *Testing*

#### 4. Kesimpulan

Berdasarkan analisis yang telah dilakukan, diperoleh kesimpulan bahwa model SVR terbaik dengan fungsi kernel linier menggunakan parameter  $C = 0,1$  dan nilai epsilon = 0,1. Model tersebut sudah layak untuk memprediksi harga saham PT. XL Axiata Tbk karena mempunyai tingkat akurasi 92,47% untuk data training dan 83,39% untuk data testing.

## DAFTAR PUSTAKA

1. Anoraga, P dan Pakarti, P., *Pengantar Pasar Modal Edisi Revisi*, Rineka Cipta, Jakarta, 2001.
2. Bertsimas, D. dan Shioda, R., Classification and Regression Via Integer Optimization, *Operation Research*, 2007, Vol. 55, No. 2: 252-271.
3. Chen, L.Y., Application of SVR with Chaotic GASA Algorithm to Forecasting Taiwanese 3G Mobile Phone Demand., *Journal of Neurocomputing*, 2014, Vol. 127: 206-213.
4. Goyal, et. al., Modeling of Daily Pan Evaporation in Sub Tropical Climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS, *Journal international of Expert Sistem with Applications*, 2014, Vol.41:5267-5276.
5. Gunn, S., *Support Vector Machines for Classification*, University of Southampton, 1998.
6. Hong, W. C., Rainfall Forecasting by Technological Learning Models, *International Journal of Applied Mathematics and Computation*, 2008, Vol. 200: 41-57.
7. Hsu, C.W., et al., *A Practical Guide to Support Vektor Classification*, Department of Computer Science and Information Engineering, National Taiwan University, 2004.
8. Kao, et al., Integration on Nonlinear Independent Component Analysis and Support Vector Regression for Stock Price Forecasting, *Neurocomputing*, 2013, Vol. 99:534-542.
9. Kazem, et. al., Support Vector Regression with Chaos Based Firefly Algorithm for Stock Market Price Forecasting, *Applied Soft Computing*, 2013, Vol. 13, No. 2: 947-958.
10. Leidiyana, H., Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Risiko Kredit Kepemilikan Kendaraan Bermotor, *Jurnal Penelitian Ilmu Komputer, Sistem Embedded & Logic*, 2013, Vol. 1, No. 1: 65-76.
11. Pai, P.F and Lin, C.S., A Hybrid ARIMA and Support Vector Machine Model in Stock Proce Forecasting, *Omega*, 2005: 497-505.
12. Santosa, B., *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Graha Ilmu: Yogyakarta, 2007.
13. Trapsilasiwi dan Abdillah, Peramalan Beban Listrik Jangka Pendek pada PT. PLN area Jawa Timur – Bali Menggunakan Support Vector Machine. *Jurnal Statistika*, Institut Teknologi Sepuluh Nopember, 2010, Vo. 3, No. 1.
14. Wang et al., Stock Index Forecasting Based on a Hybrid Model, *Omega*, 2012, Vol. 40, No. 6: 758-766.
15. Yao, Y, et al., An Improved Grid Search Algorithm and Its Application in PCA and SVM Based Face Recognition, *Journal of Computational Information Sistems*, 2014, Vol. 10, No. 3:1219–1229.