

IMPLEMENTASI SUPPORT VECTOR MACHINE PADA PREDIKSI HARGA SAHAM GABUNGAN (IHSG)

Eka Patriya

*Fakultas Ekonomi Universitas Gunadarma
Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat
ekapatriya@staff.gunadarma.ac.id*

Abstrak

Saham adalah instrumen pasar keuangan yang banyak dipilih oleh investor sebagai alternatif sumber keuangan, akan tetapi saham yang diperjual belikan di pasar keuangan sering mengalami fluktuasi harga (naik dan turun) yang tinggi. Para investor berpeluang tidak hanya mendapat keuntungan, tetapi juga dapat mengalami kerugian di masa mendatang. Salah satu indikator yang perlu diperhatikan oleh investor dalam berinvestasi saham adalah pergerakan Indeks Harga Saham Gabungan (IHSG). Tindakan dalam menganalisa IHSG merupakan hal yang penting dilakukan oleh investor dengan tujuan untuk menemukan suatu trend atau pola yang mungkin berulang dari pergerakan harga saham masa lalu, sehingga dapat digunakan untuk memprediksi pergerakan harga saham di masa mendatang. Salah satu metode yang dapat digunakan untuk memprediksi pergerakan harga saham secara akurat adalah machine learning. Pada penelitian ini dibuat sebuah model prediksi harga penutupan IHSG menggunakan algoritma Support Vector Regression (SVR) yang menghasilkan kemampuan prediksi dan generalisasi yang baik dengan nilai RMSE training dan testing sebesar 14.334 dan 20.281, serta MAPE training dan testing sebesar 0.211% dan 0.251%. Hasil penelitian ini diharapkan dapat membantu para investor dalam mengambil keputusan untuk menyusun strategi investasi saham.

Kata Kunci: *Indeks Harga Saham Gabungan, Machine Learning, Prediksi, Support Vector Regression*

Abstract

Stocks are financial market instruments that have been chosen by investors as an alternative financial source, but shares traded on financial markets often experience high (up and down) price fluctuations. Investors have the opportunity to not only make a profit but also suffer losses in the future. One indicator that investors need to consider in investing in stocks is the movement of the Indonesia Composite Index (IDX Composite). The act of analyzing the IDX Composite is an important thing to do by investors in order to find trends or patterns that might repeat from past stock prices, so they can be used to predict future stock prices. One method that can be used to predict price movements is machine learning. In this research, the IDX Composite closing price prediction model was made using the Support Vector Regression (SVR) algorithm which produces good prediction and generalization capabilities with RMSE training and testing values of 14.334 and 20.281, and MAPE training and testing of 0.211% and 0.251%. The results of this research are expected to help investors in making decisions to develop stock investment strategies.

Keywords: *Indonesia Composite Index, Machine Learning, Prediction, Support Vector Regression*

PENDAHULUAN

Salah satu indikator yang perlu diperhatikan oleh investor dalam berinvestasi saham adalah pergerakan Indeks Harga Saham Gabungan (IHSG). Pergerakan IHSG mencerminkan kinerja pasar keuangan apakah sedang mengalami peningkatan atau penurunan, apabila kondisi ekonomi negara sedang baik, maka IHSG menunjukkan adanya *trend* yang meningkat [1], tetapi jika kondisi ekonomi negara dalam keadaan turun, maka akan berpengaruh juga terhadap IHSG. Tindakan dalam menganalisa IHSG merupakan hal yang penting dilakukan oleh investor dengan tujuan untuk menemukan suatu *trend* atau pola yang mungkin berulang dari pergerakan harga saham masa lalu, sehingga dapat digunakan untuk memprediksi pergerakan harga saham di masa mendatang[2]. Hasil prediksi yang baik dapat membantu investor menyusun rencana kapan harus membeli atau menjual saham. Salah satu indikator yang digunakan untuk menggambarkan pergerakan saham yaitu indeks harga saham. Terdapat beberapa jenis indeks harga saham, salah satunya ialah Indeks Harga Saham Gabungan (IHSG). IHSG merupakan indeks yang melibatkan semua saham yang tercatat sebagai komponen perhitungan indeks harga [3]. Pergerakan indeks saham bila mengalami penurunan maka dapat dikatakan kebanyakan harga saham mengalami penurunan dan begitu pula sebaliknya. Investor harus mampu mengkalkulasi dampak positif maupun negatif

terhadap kinerja perusahaan beberapa tahun kedepan, kemudian mengambil keputusan saat ini untuk membeli atau menjual saham yang bersangkutan [4].

Salah satu metode yang dapat digunakan untuk memprediksi pergerakan harga saham secara akurat adalah *machine learning*. *Machine learning* merupakan metode yang membuat bagaimana komputer/program mempunyai suatu kecerdasan dengan menggunakan data seperti angka, teks, gambar, video, atau audio sebagai bahan belajar untuk memperoleh pengetahuan dari data tersebut dan kemudian dapat membantu manusia dalam mengambil keputusan atau menyelesaikan suatu permasalahan. Berikut beberapa penelitian terkait implementasi metode *machine learning* untuk memprediksi pergerakan harga saham telah dilakukan peneliti terdahulu.

Penelitian [5] membuat 2 model prediksi harga penutupan saham yaitu model Jaringan Saraf Tiruan (JST) dengan algoritma *Backpropagation* (BP) dan model JST-BP dengan algoritma *Particle Swarm Optimization* (PSO). Data yang digunakan adalah data IHSG pada Maret 2016 sampai Februari 2017. Hasil penelitian menunjukkan bahwa model JST-BP dengan algoritma PSO memberikan hasil yang paling optimal dengan *Mean Squared Error* (MSE) sebesar 0,0062 dan *Mean Absolute Percentage Error* (MAPE) sebesar 1,88%.

Penelitian [6] melakukan studi perbandingan terhadap algoritma *Multilayer Perceptron* (MLP) dan *Support Vector*

Machine (SVM) untuk memprediksi indeks saham harian LQ45 khusus sektor perbankan yaitu Bank Central Asia Tbk (BBCA.JK).

Peneliti menggunakan data sekunder harga penutupan saham harian periode Juli 2009 sampai Agustus 2013. Hasil penelitian yang diperoleh adalah model SVM *lag* 10 memberikan hasil yang paling akurat dengan *Root Mean Squared Error* (RMSE) sebesar 317,89.

Penelitian [7] melakukan prediksi nilai harga penutupan IHSG menggunakan *Support Vector Regression* (SVR), *Random Forest* (RF), dan SVR-RF. Peneliti mengambil data selama lima tahun yaitu 2011-2015. Berdasarkan hasil penelitian, diperoleh bahwa model SVR menghasilkan prediksi jangka pendek terbaik untuk hari ke (t+1) dan (t+5) dengan MAPE secara urut sebesar 1,9119% dan 4,5691%, sementara model SVR-RF menghasilkan prediksi jangka panjang terbaik untuk hari ke (t+10), (t+20), dan (t+30) MAPE secara urut sebesar 4,1173%, 8,6391%, dan 9,1708%.

Penelitian [8] membuat sebuah sistem prediksi terhadap harga penutupan empat saham *blue chip* yaitu AMD, IBM, NIKE, dan NVIDIA menggunakan SVR dan *Firefly Algorithm* (FA). Data yang digunakan adalah data harga saham harian, mingguan, dan bulanan dengan periode 2010-2014. Hasil penelitian menunjukkan sistem prediksi menghasilkan akurasi yang cukup bagus terhadap keempat harga saham harian, mingguan, dan bulanan *blue chip* dengan MAPE rata-rata

kurang dari 5%.

Penelitian [9] menghasilkan sebuah model menggunakan SVR dan algoritma *Grid Search* untuk memprediksi data harga saham harian PT. XL Axiata Tbk. Peneliti menggunakan data saham periode 1 Januari 2013 sampai 30 September 2014. Hasil penelitian yang diperoleh adalah model SVR-*Grid Search* menghasilkan koefisien determinasi atau tingkat akurasi sebesar 92,47% untuk data *training* dan 83,39% untuk data *testing*.

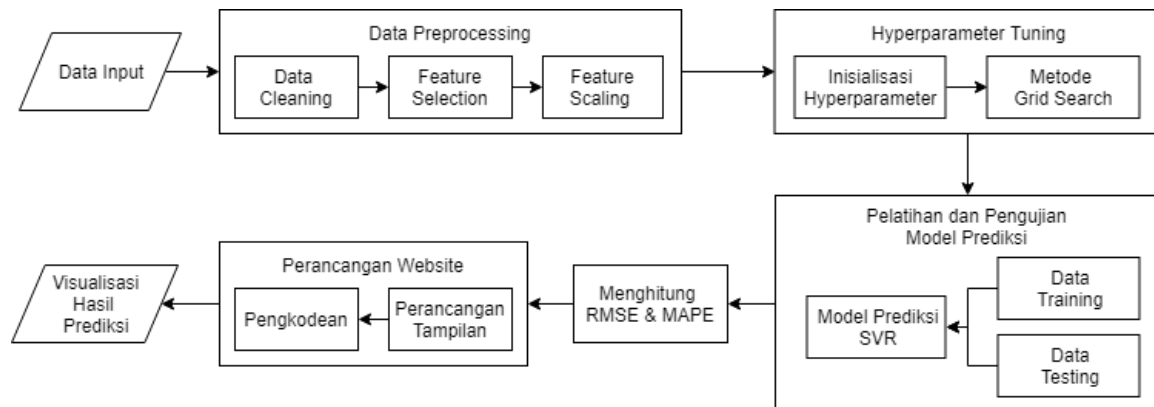
Pada penelitian ini, peneliti akan membuat model prediksi harga penutupan Indeks Harga Saham Gabungan (IHSG) menggunakan algoritma *Support Vector Regression* (SVR). Peneliti akan melakukan *hyperparameter tuning* menggunakan metode *Grid Search* untuk menentukan *hyperparameter* yang optimal pada SVR. Tingkat kesalahan pada hasil prediksi dihitung menggunakan nilai *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE). Hasil penelitian ini diharapkan dapat membantu para investor dalam mengambil keputusan untuk menyusun strategi investasi saham.

METODE PENELITIAN

Tahapan proses pada penelitian ini dimulai dapat dilihat pada Gambar 1. Tahapan dimulai dari data *input*, melakukan data *preprocessing* yang terdiri dari data *cleaning*, *feature selection* dan *feature scaling*, melakukan *hyperparameter tuning*

dengan menentukan *hyperparameter* dan melakukan metode *grid search*, melatih dan menguji model prediksi SVR menggunakan data *training* dan *testing*, menghitung *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE) terhadap

tingkat kesalahan hasil prediksi, merancang *website* yang akan ditanami model prediksi (*embedded model*), dan tahapan akhir adalah visualisasi hasil prediksi terhadap data aktual berdasarkan periode yang diujikan terhadap model prediksi.



Gambar 1. Metode Penelitian

Tabel 1. Sampel Data Saham IHSG Januari 2014

| Date | Open | High | Low | Close | Adj. Close | Volume |
|------------|-----------|-----------|-----------|-----------|------------|-----------|
| 2014-01-02 | 4294.4951 | 4327.2651 | 4287.8081 | 4327.2651 | 4327.1367 | 231289550 |
| | 17 | 37 | 05 | 37 | 19 | 0 |
| 2014-01-03 | 4297.7148 | 4298.2309 | 4247.9868 | 4257.6630 | 4257.5366 | 218691150 |
| | 44 | 57 | 16 | 86 | 21 | 0 |
| 2014-01-06 | 4259.5830 | 4263.6230 | 4188.375 | 4202.8090 | 4202.6840 | 197230780 |
| | 08 | 47 | | 82 | 82 | 0 |
| 2014-01-07 | 4206.2968 | 4212.3168 | 4175.8061 | 4175.8061 | 4175.6821 | 238471990 |
| | 75 | 95 | 52 | 52 | 29 | 0 |
| 2014-01-08 | 4183.5498 | 4204.3017 | 4161.1928 | 4200.5927 | 4200.4677 | 230021660 |
| | 05 | 58 | 71 | 73 | 73 | 0 |
| 2014-01-09 | 4199.9677 | 4216.8217 | 4189.1210 | 4201.2177 | 4201.0927 | 275824190 |
| | 73 | 77 | 94 | 73 | 73 | 0 |
| 2014-01-10 | 4196.6889 | 4270.1591 | 4190.5629 | 4254.9711 | 4254.8452 | 296373780 |
| | 65 | 8 | 88 | 91 | 15 | 0 |
| 2014-01-13 | 4294.1938 | 4393.3198 | 4292.3339 | 4390.7709 | 4390.6406 | 490466570 |
| | 48 | 24 | 84 | 96 | 25 | 0 |
| 2014-01-15 | 4402.3579 | 4459.4770 | 4398.1171 | 4441.5942 | 4441.4624 | 456294940 |
| | 1 | 51 | 88 | 38 | 02 | 0 |
| 2014-01-16 | 4455.5449 | 4457.7641 | 4411.9057 | 4412.4887 | 4412.3579 | 321094710 |
| | 22 | 6 | 62 | 7 | 1 | 0 |

Data Input

Penelitian ini akan menggunakan data saham harian Indeks Harga Saham Gabungan (IHSG) yang diperoleh dari *website* Yahoo! Finance. Data saham terdiri dari kolom *Date*, *Open*, *High*, *Low*, *Close*, *Adj. Close*, dan *Volume* yang memiliki periode Januari 2014 – April 2019 atau sebanyak 1336 sampel. *Date* menunjukkan tanggal yang tercatat pada waktu harga saham terjadi dalam satu hari. *Open* adalah harga pembukaan saham yang pertama kali dikeluarkan saat kegiatan transaksi saham dibuka dalam satu hari. *High* dan *Low* adalah harga tertinggi dan terendah saham yang terjadi dalam satu hari saat kegiatan transaksi saham berlangsung. *Close* adalah harga penutupan saham yang terakhir kali terjadi saat kegiatan transaksi saham ditutup dalam satu hari. *Adj. Close* adalah harga penutupan saham yang telah disesuaikan terhadap harga *Close* ketika terjadi aksi korporasi perusahaan seperti *dividen* dan *stock split*. *Volume* merupakan jumlah total transaksi saham yang terjadi dalam satu hari. Sebagai contoh dapat dilihat pada Tabel 1 merupakan data saham IHSG 10 hari pertama pada bulan Januari 2014.

Seperti dapat dilihat pada Tabel 1, kegiatan transaksi saham berlangsung dalam periode bisnis harian (Senin – Jumat). Pada keempat baris yang diberi warna kuning menunjukkan adanya tanggal yang terlewat karena pada tanggal tersebut tidak ada kegiatan transaksi saham atau tidak termasuk dalam periode bisnis harian, sehingga tidak tercatat ke dalam data saham harian.

Data Preprocessing

Tahapan data *preprocessing* dilakukan untuk membentuk data saham yang masih “mentah” menjadi bentuk data *input* yang dapat diterima dan dipelajari polanya (*pattern*) oleh model prediksi. Tahap ini terdiri dari tiga proses yaitu data *cleaning*, *feature selection* dan *feature scaling*. Hasil dari tahap ini yaitu data *input* yang terbagi menjadi data *training* dan *testing*, untuk selanjutnya digunakan dalam tahap *hyperparameter tuning*, pelatihan, dan pengujian model prediksi. Data saham yang diperoleh dari *website* Yahoo! Finance mempunyai beberapa sampel *null* atau “NaN”, hal ini dikarenakan setiap tahun terdapat hari libur bursa seperti hari raya tahun baru, libur lebaran, perayaan natal, dan lain-lain, sehingga tidak ada perdagangan saham yang terjadi pada hari tersebut. Peneliti melakukan data *cleaning* untuk membersihkan sampel tersebut karena model prediksi tidak bisa menerima *input* yang mempunyai nilai *null* atau “NaN”. Data saham yang telah dilakukan data *cleaning* mempunyai 1292 sampel. *Feature selection* bertujuan untuk memilih fitur atau variabel independen yang digunakan sebagai *input* untuk dipelajari oleh model prediksi. Fitur yang digunakan sangat mempengaruhi performa model, karena pemilihan fitur yang baik atau mempunyai relasi yang tinggi dengan variabel target (*output*) akan menghasilkan model prediksi yang baik, dan begitu juga sebaliknya. Dalam penelitian ini akan dilakukan *feature selection*

terhadap kolom harga *Open*, *High*, dan *Low* yang digunakan sebagai input untuk mengestimasi harga penutupan (*Close*) pada Indeks Harga Saham Gabungan (IHSG).

Feature scaling digunakan untuk membuat variabel independen memiliki bobot atau rentang nilai yang sama agar model prediksi dapat belajar dengan baik dari setiap variabel independen yang diberikan. Hal ini dikarenakan jika salah satu variabel independen memiliki bobot atau rentang nilai

yang tinggi, maka berakibat dapat mendominasi *decision function* dan membuat model tidak dapat belajar sesuai yang diharapkan dari variabel independen lain. Penelitian ini menggunakan *library Standard Scaler* yang disediakan oleh *scikit-learn* untuk mempercepat proses komputasi *feature scaling* pada data *input*. *Library Standard Scaler* berfungsi untuk mengubah variabel independen menjadi bentuk dasar (*standard*) dengan Persamaan (1) berikut :

$$Z_i = \frac{(X_i - U)}{S} \quad (1)$$

Dimana X_i adalah sampel data ke I, U merupakan rata-rata dari sampel data dan S adalah simpangan baku dari sampel data. Proses *standard scaling* dilakukan untuk mencari nilai rata-rata (*mean*) dan simpangan baku (*standard deviation*) pada setiap variabel independen, lalu kedua nilai tersebut digunakan dalam Persamaan (1) untuk diimplementasikan kepada semua data *input* pada setiap variabel independen, sehingga menghasilkan data *input* yang mempunyai atribut nilai *mean* sebesar 0 dan *standard deviation* 1. Perlu diketahui bahwa nilai *mean* dan *standard deviation* yang dihasilkan pada setiap variabel independen berbeda-beda, maka proses *standard scaling* terjadi secara independen terhadap semua data *input* pada setiap variabel independen. Sebagai tambahan, proses perhitungan untuk mencari nilai *mean*

dan *standard deviation* hanya dilakukan pada data *training*, dikarenakan untuk mencegah kebocoran (*leaking*) nilai statistik pada data *testing* terhadap model prediksi jika dilakukan perhitungan pada semua dataset. Hal ini dilakukan karena model prediksi tidak boleh melihat (*peeking*) data *testing* sebelum dilakukan pelatihan sehingga mempunyai kemampuan generalisasi yang dapat diandalkan (*reliable*). Setelah mendapatkan nilai *mean* dan *standard deviation* pada data *training*, selanjutnya digunakan Persamaan (1) untuk mengubah data *training* dan *testing* ke dalam bentuk dasar (*standard*). Berikut merupakan hasil perhitungan nilai *mean* dan *standard deviation* yang diperoleh pada data *training* dari setiap variabel independen yang digunakan dalam penelitian, dapat dilihat pada Tabel 2.

Tabel 2. Hasil Perhitungan *Mean* dan *Standard Deviation*

| Variabel Independen | Mean | Standard Deviation |
|---------------------|-------------------|--------------------|
| Open | 5222.12177 397 | 549.67569141 |
| High | 5244.67440 955 | 548.51505795 |
| Low | 5198.27217 26 | 548.56202697 |

Berdasarkan hasil perhitungan pada Tabel 2 selanjutnya dapat diimplementasikan untuk semua data *training* dan *testing* pada setiap variabel independen menggunakan Persamaan (1) sehingga menghasilkan bentuk dasar (*standard*) untuk digunakan di dalam *hyperparameter tuning*, pelatihan, dan pengujian model prediksi.

Hyperparameter Tuning

Tahapan *hyperparameter tuning* dilakukan untuk menentukan *hyperparameter* yang optimal atau terbaik untuk data *input* dan model prediksi, sehingga dapat menghasilkan model prediksi dengan kemampuan generalisasi yang baik. Pada dasarnya setiap data *input* yang digunakan dalam melatih sebuah model prediksi membutuhkan *hyperparameter* yang sesuai dengan karakteristik data *input* tersebut, dengan harapan dapat memberikan hasil prediksi yang akurat dan dipercaya. Oleh karena itu, tahapan ini dilakukan untuk menentukan *hyperparameter* yang sesuai dengan data *training* dan menghasilkan generalisasi yang baik pada data *testing*. Pada penelitian ini, tahapan *hyperparameter tuning* diawali dengan menginisialisasi *hyperparameter* yang akan

dioptimalkan, lalu dilakukan metode *Grid Search* untuk menentukan *hyperparameter* optimal, dan hasilnya akan digunakan dalam pelatihan dan pengujian model prediksi.

Dalam menentukan *hyperparameter* yang optimal dan sesuai dengan data *input*, terlebih dahulu peneliti menginisialisasi *hyperparameter* yang digunakan dalam model prediksi *Support Vector Regression* (SVR). *Hyperparameter* yang dapat dioptimalkan dalam SVR yaitu konstanta *C*, *epsilon*, dan fungsi *kernel*. Konstanta *C* berfungsi sebagai *trade-off* antara ketipisan fungsi dan batas toleransi kesalahan. *Epsilon* digunakan untuk menentukan seberapa besar batas rentang toleransi kesalahan yang disebut zona *ε-insensitive*. Fungsi *kernel* dipakai untuk mentransformasikan ruang *input* ke dalam dimensi yang lebih tinggi, sehingga bisa dilakukan pencarian *hyperplane* secara linear pada data non-linear. Penelitian ini menggunakan fungsi *kernel Radial Basis Function* (RBF) dengan mengoptimalkan *gamma* yang digunakan sebagai *hyperparameter* pada *kernel* RBF. *Gamma* menentukan seberapa jauh pengaruh dari satu sampel data *training* terhadap perhitungan *decision function*.

Metode Grid Search

Setelah *range hyperparameter* diinisialisasi, peneliti mengimplementasikan metode *grid search* untuk melatih beberapa model prediksi yang dibentuk melalui setiap kombinasi *range hyperparameter*, lalu model prediksi yang menghasilkan *score Mean Squared Error* (MSE) terkecil akan dipilih dan atribut *hyperparameter* model tersebut ditetapkan sebagai *hyperparameter* yang optimal. Dalam melatih beberapa model, *grid search* menggunakan teknik yang disebut *cross-validation*. *Cross-validation* akan membagi data *training* yang diperoleh dari tahapan *data preprocessing*, menjadi beberapa lipatan (*fold*) yang terdiri dari data *training* dan *validation*. Pada setiap model dari kombinasi *hyperparameter* yang terbentuk, dilakukan *cross-validation* untuk menghitung *score MSE* rata-rata untuk semua lipatan (*fold*) yang diperoleh selama pelatihan model tersebut. Sebagai tambahan, semakin banyak

komposisi *hyperparameter* yang diinisialisasi, maka semakin banyak model yang terbentuk dan dilakukan *cross-validation*, sehingga metode *grid search* membutuhkan waktu komputasional yang lama dalam proses menentukan *hyper-parameter* yang optimal. Peneliti menggunakan *library Time Series Split* dan *Grid Search CV* yang disediakan oleh *scikit-learn* untuk mengimplementasikan metode *grid search* dengan teknik *time series 5-fold cross-validation*, karena data *input* pada penelitian ini merupakan data *time series*, maka teknik *cross-validation* biasa tidak dapat digunakan karena dapat merusak urutan periode waktu yang seharusnya penting diperhatikan di dalam data *time series*. Adapun, perhitungan pembagian data *training* menjadi beberapa lipatan (*fold*) pada *time series cross-validation* menggunakan Persamaan (2) dan Persamaan 3.

$$\text{Training Size} = \frac{i \cdot n_{\text{samples}}}{n_{\text{fold}} + 1} + n_{\text{samples}} \% (n_{\text{fold}} + 1) \quad (2)$$

$$\text{Validation Size} = \frac{n_{\text{samples}}}{n_{\text{fold}} + 1} \quad (3)$$

Training size pada persamaan 2 dihitung untuk menentukan ukuran pelatihan ukuran *i* pelatihan data set yang digunakan, dimana *i* adalah indeks pada *fold* ke-*i*, *n_{samples}* merupakan jumlah sampel pada data *training*, *n_{fold}* adalah jumlah *fold* yang digunakan. Validation Size menunjukkan validasi dari

data set yang merupakan perbandingan antara sampel data ke-*n* dengan penjumlahan satu dan jumlah *fold* yang digunakan seperti dapat dilihat pada Persamaan 3.

Tabel 3 merupakan hasil perhitungan dalam pembagian data *training* sebanyak 1033 sampel menggunakan 5-*fold*.

Tabel 3. Hasil Perhitungan 5-Fold Cross Validation

| Indeks Fold | Training Size (Sampel) | Validation Size (Sampel) |
|-------------|------------------------|--------------------------|
| Fold ke 1 | 173 | 172 |
| Fold ke 2 | 345 | 172 |
| Fold ke 3 | 517 | 172 |
| Fold ke 4 | 689 | 172 |
| Fold ke 5 | 861 | 172 |

Berdasarkan Tabel 3 diketahui bahwa sampel data *training* selalu meningkat. Hal ini dikarenakan data *training* pada indeks *fold* sebelumnya merupakan *subset* dari data *training* pada indeks *fold* selanjutnya. Sampel data *training* pada indeks *fold* selanjutnya merupakan hasil penjumlahan data *training* dan *validation* pada indeks *fold* sebelumnya. Sehingga data *training* semakin bertambah sementara data *validation* selalu konstan dan berpindah sesuai urutan periode waktu pada setiap *fold*.

Pelatihan dan Pengujian Model Prediksi

Tahapan ini dilaksanakan setelah tahap data *preprocessing* dan *hyperparameter tuning* selesai dilakukan. Pelatihan dan pengujian pada model prediksi menggunakan data *training* dan data *testing* yang telah dihasilkan pada tahap data *preprocessing*. Model prediksi *Support Vector Regression* (SVR) dibentuk menggunakan *hyperparameter* optimal yang diperoleh pada tahap *hyperparameter tuning*. Setelah dilakukan pelatihan dan pengujian, maka model akan menghasilkan prediksi yang akan dihitung

tingkat kesalahannya menggunakan nilai *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE).

Data Training dan Testing

Pada penelitian ini, data *input* dibagi sebesar 80% untuk data *training* dan 20% untuk data *testing*, dengan memperhatikan urutan periode waktu. Data *training* mempunyai sampel sebanyak 1033 sampel pada periode Januari 2014 – Maret 2018 yang digunakan untuk melatih model prediksi menghasilkan kemampuan generalisasi yang dapat dipercaya dan tingkat kesalahan yang kecil, sedangkan data *testing* memiliki sampel sebanyak 259 sampel pada periode April 2018 – April 2019 yang berguna untuk menguji kinerja model dalam memprediksi nilai pada data yang belum pernah dilihat (*unseen data*). Data *training* dan *testing* juga dipakai untuk membandingkan nilai aktual dan hasil prediksi dalam perhitungan tingkat kesalahan.

Model Prediksi Support Vector Regression

Model prediksi menggunakan algoritma *Support Vector Regression* (SVR)

dalam metode pembelajarannya untuk mengenal pola (*pattern*) pada data *training*. *Hyperparameter* yang optimal pada SVR yang diperoleh dari tahap *hyperparameter tuning* digunakan untuk membantu pelatihan model prediksi sehingga menghasilkan kinerja model yang presisi. Peneliti membuat model prediksi yang bertujuan untuk memprediksi harga penutupan saham (*Close*) terhadap Indeks Harga Saham Gabungan (IHSG) dinotasikan sebagai (Y_t), menggunakan 3 variabel independen yaitu harga pembukaan (*Open*), tertinggi (*High*), dan terendah (*Low*). Dalam melatih model SVR, peneliti menggunakan data *training* dan dibantu oleh *hyperparameter* optimal untuk menghasilkan model yang mempunyai kemampuan untuk memprediksi harga penutupan saham (*Close*). Setelah itu dilakukan pengujian model SVR dengan memasukan data *testing* sebagai periode yang diujikan, sehingga model dapat menghasilkan prediksi yang diharapkan.

Selanjutnya hasil prediksi akan di-

hitung seberapa besar kesalahannya terhadap data aktual menggunakan perhitungan *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE) sebagai berikut :

1. Perhitungan RMSE

Hasil prediksi yang diperoleh selanjutnya dapat dihitung seberapa besar kesalahan yang dihasilkan dengan membandingkannya dengan data aktual. Pada penelitian ini digunakan perhitungan *Root Mean Squared Error* (RMSE) untuk mengukur seberapa besar rata-rata kesalahan (*residual*) antara hasil prediksi dan nilai aktualnya. Sebagai tambahan RMSE sensitif terhadap perubahan skala pada variabel target, sehingga nilai RMSE dapat berubah-ubah tergantung ukuran variansi variabel target, dalam penelitian ini variabel target merupakan harga penutupan saham (*Close*). RMSE dihitung menggunakan persamaan 4.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - Y'_t)^2} \quad (4)$$

Dimana Y_t adalah nilai aktual pada data ke- t , Y'_t merupakan nilai prediksi pada data ke- t dan n adalah jumlah data.

2. Perhitungan *Mean Absolute Percentage Error* (MAPE) dilakukan untuk menemukan tingkat kesalahan dalam bentuk persentase. Formula untuk menghitung MAPE diberikan pada persamaan 5.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - Y'_t|}{Y_t} \times 100\% \quad (5)$$

Dimana Y_t adalah nilai aktual pada data ke- t , Y'_t merupakan nilai prediksi pada data ke- t dan n adalah jumlah data.

Secara garis besar semakin kecil kedua nilai maka semakin kecil tingkat kesalahan yang dihasilkan model prediksi. Oleh karena itu, peneliti akan membandingkan nilai RMSE dan MAPE pada data *training* dan *testing* untuk mengetahui seberapa besar perbedaan kedua nilai yang diperoleh pada kedua data tersebut. Jika perbedaannya sangat signifikan maka dapat disimpulkan bahwa model prediksi memiliki kemampuan generalisasi yang buruk (*overfit*), sebaliknya jika kedua nilai mempunyai perbedaan yang kecil maka model mampu memprediksi dengan baik terhadap data yang belum pernah dilihat (*unseen data*). Peneliti menggunakan *library numpy* dan *scikit-learn* dalam membantu perhitungan RMSE dan MAPE terhadap data saham (aktual) dan hasil prediksi sebanyak 1292 sampel.

HASIL DAN PEMBAHASAN

Penelitian ini dibuat secara bertahap sesuai dengan tahapannya yaitu dari tahap *preprocessing* sampai tahap visualisasi. Berikut ini adalah penjelasan dari masing-

masing tahapan tersebut.

Hasil Preprocessing

Hasil dari tahapan data *preprocessing* merupakan data saham berupa variabel independen dan target yang sudah dipisahkan menggunakan rasio 80% *training* dan 20% *testing*, yang selanjutnya dilakukan *feature scaling* pada kedua data tersebut.

Hasil Hyperparameter Tuning

Hasil dari tahapan ini merupakan perolehan ketiga *hyperparameter* optimal yaitu C , *epsilon*, dan *gamma* dengan *score Mean Squared Error* (MSE) yang terkecil. Proses dalam tahapan ini menghasilkan 343 model prediksi berdasarkan *range hyperparameter* yang telah diinisialisasikan sebelumnya. Berdasarkan hasil penelitian dari 343 model prediksi diperoleh bahwa *hyperparameter tuning* menghasilkan *hyperparameter* optimal $C = 1000000$, *epsilon* = 1, dan *gamma* = 0.0001 dengan MSE sebesar 240.03.

Hasil Perhitungan Rmse Dan Mape

Peneliti memasukkan *file csv* berupa data *testing* sebanyak 259 sampel pada periode April 2018 – April 2019 agar dapat dihitung tingkat kesalahan prediksinya dengan menggunakan *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE). Peneliti juga

menghitung RMSE dan MAPE pada data *training* agar dapat dibandingkan apakah terdapat perbedaan yang signifikan, hasil perhitungan dapat dilihat pada Tabel 4.

Berdasarkan hasil Tabel 4, nilai RMSE dan MAPE yang diperoleh dari data *training* dan *testing* tidak mengalami perbedaan yang signifikan, sehingga model prediksi sudah memiliki kemampuan prediksi dan generalisasi yang baik pada data *training* dan *testing*, dengan nilai RMSE *training* 14.334 dan MAPE *training* 0.211%, serta RMSE *testing* 20.281 dan MAPE *testing* 0.251%.

Visualisasi Hasil Prediksi

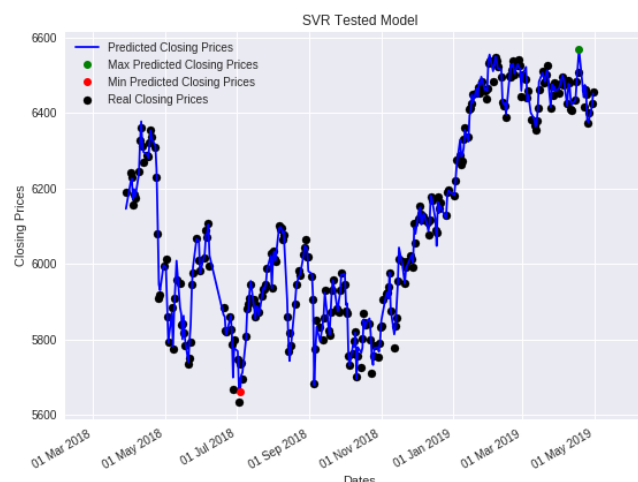
Visualisasi hasil prediksi pada *website* menghasilkan representasi yaitu

berupa grafik. *Website* menghasilkan dua grafik dengan properti sumbu *axis* yang berbeda yaitu *Date* dan *Close Real*. Grafik pertama dapat dilihat pada gambar 2.

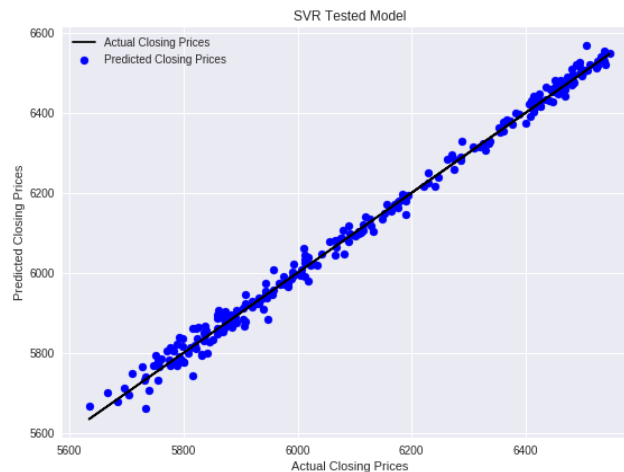
Gambar 2 menunjukkan grafik hasil prediksi (garis biru) yang berusaha untuk mengikuti pola pergerakan data aktualnya (titik hitam) berdasarkan urutan periode waktu pada data *testing*. Pola data pada hasil prediksi menghasilkan pola *trend* yang sempat menurun dari bulan Mei 2018 – Juli 2018, kemudian meningkat pada bulan November 2018 – Maret 2019. Titik hijau merupakan harga tertinggi (*max*) pada harga penutupan (*close*) diprediksi sebesar 6567.809 pada tanggal 18 April 2019, sedangkan titik merah merupakan harga terendah (*min*) diprediksi sebesar 5660.678 pada tanggal 4 Juli 2018.

Tabel 4. Hasil Perhitungan RMSE dan MAPE

| Data | RMSE | MAPE (%) |
|----------|--------|----------|
| Training | 14.334 | 0.211 |
| Testing | 20.281 | 0.251 |



Gambar 2. Grafik Hasil Prediksi *Time Series*



Gambar 3. Grafik Hasil Prediksi dan Data Aktual

Grafik kedua dapat dilihat pada Gambar 3. Gambar 3 menunjukkan seberapa besar perbedaan (*residual*) antara hasil prediksi (titik biru) dengan data aktual (garis hitam) pada data *testing*. Jika hasil prediksi terletak cukup jauh terhadap garis data aktualnya, maka dapat diperkirakan model menghasilkan prediksi dengan perbedaan (*residual*) yang besar. Berdasarkan Gambar 3, dapat diketahui bahwa hasil prediksi memiliki perbedaan (*residual*) yang kecil, karena setiap hasil prediksi terletak dekat terhadap garis data aktualnya, sehingga model menghasilkan kinerja yang baik dalam memprediksi data *testing* yang ditunjukkan dengan RMSE *testing* 20.281 dan MAPE *testing* 0.251%.

KESIMPULAN DAN SARAN

Kesimpulan dari penelitian ini adalah pembuatan model prediksi menggunakan algoritma *Support Vector Regression* (SVR) untuk memprediksi harga penutupan Indeks

Harga Saham Gabungan (IHSG) telah berhasil dilakukan, dengan penentuan *hyperparameter* optimal menggunakan metode *Grid Search* telah berhasil dilakukan, sehingga menghasilkan *hyperparameter* $C = 1000000$, $\epsilon = 1$, dan $\gamma = 0.0001$ dengan nilai *Mean Squared Error* (MSE) sebesar 240.03. Tingkat kesalahan terhadap hasil prediksi telah berhasil dihitung menggunakan *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE) yang menghasilkan RMSE *training* dan *testing* sebesar 14.334 dan 20.281, serta MAPE *training* dan *testing* sebesar 0.211% dan 0.251%. Tidak adanya perbedaan yang signifikan antara tingkat kesalahan data *training* dan *testing*, menandakan model prediksi sudah memiliki kemampuan prediksi dan generalisasi yang baik pada data *training* dan *testing*.

Visualisasi antara data asli dan hasil prediksi pada periode yang diujikan terhadap model prediksi telah berhasil dilakukan, sehingga menghasilkan grafik pertama yang

membentuk pola *trend* yang yang sempat menurun dari bulan Mei 2018 – Juli 2018, kemudian meningkat pada bulan November 2018 – Maret 2019, dengan harga tertinggi penutupan (*max close*) diprediksi sebesar 6567.809 pada tanggal 18 April 2019 dan harga terendah (*min close*) diprediksi sebesar 5660.678 pada tanggal 4 Juli 2018. Grafik kedua yang menunjukkan bahwa hasil prediksi memiliki perbedaan (*residual*) yang kecil, karena setiap hasil prediksi terletak dekat terhadap garis data aktualnya, sehingga model menghasilkan kinerja yang baik dalam memprediksi data *testing*.

Pengembangan lebih lanjut dapat dilakukan untuk menyempurnakan penelitian ini dengan ditambahkan variabel independen seperti laporan keuangan, analisis sentimen, atau kondisi perekonomian berbagai perusahaan yang bergerak searah dengan Indeks Harga Saham Gabungan (IHSG). Penggunaan metode optimasi lain yang lebih cepat dan efisien juga perlu diperhatikan dalam menentukan *hyperparameter* optimal pada *Support Vector Regression* (SVR). Selain itu, diharapkan kedepannya penggunaan SVR dapat dikembangkan dan digabungkan dengan algoritma *machine learning* lain untuk menghasilkan model prediksi dengan kinerja yang lebih presisi.

DAFTAR PUSTAKA

- [1] X. Ding, Y. Zhang, T. Liu, and J. Duan, “Deep Learning for Event-Driven Stock Prediction”, *International Joint Conference on Artificial Intelligence (IJCAI)*, pp 2327-2333, 2015.
- [2] A. Dutta, G. Bandopadhyay, and S. Sengupta, “Prediction of Stock Performance in Indian Stock Market Using Logistic Regression”, *International Journal of Business and Information*, vol 7 no. 1 pp 105–136, 2012.
- [3] Mariati., P. R. K. Sari, dan A. Salam, “Pengaruh Panama Papers terhadap Indeks Harga Saham Gabungan (IHSG)”, *Jurnal Manajemen dan Bisnis*, vol 2, No.1, 2019.
- [4] A.M. Razak, dan Suhadak, “Pengaruh Worldwide Governance Indicators dan Macroeconomic Terhadap IHSG”, *Jurnal Administrasi Bisnis*, Vol.70, No.1, 2019.
- [5] D. Wartati, dan N.A. Masrurroh, “Aplikasi Jaringan Saraf Tiruan Dan Particle Swarm Optimization Untuk Peramalan Indeks Harga Saham Bursa Efek Indonesia”, *Jurnal Teknosains*, vol 6, No 1, pp. 22-30, 2016.
- [6] T. Hidayatulloh, “Kajian Komparasi Penerapan Algoritma Support Vector Machine (SVM) Dan Multilayer Perceptron (MLP) Dalam Prediksi Indeks Saham Sektor Perbankan”, *Seminar Nasional Inovasi Dan Tren (SNIT)*, pp 262-272, 2014.
- [7] B. Harahap, R. F. Umbara, dan D. Triantoro, “Prediksi Indeks Harga Saham

Menggunakan Metode Gabungan Support Vector Regression dan Random Forest”, e-Proceeding of Engineering, vol 3, No 3, pp. 5353–5363, 2016.

- [8] Alfredo, Jondri, dan R. Rismala, “Prediksi Harga Saham menggunakan Support Vector Regression dan Firefly Algorithm”, e-Proceeding of Engineering, vol 2, No 2, 2015
- [9] H. Yasin, A. Prahutama, dan T.W. Utami, “Prediksi Harga Saham Menggunakan Support Vector Regression Dengan Algoritma Grid Search”, *Media Statistika*, vol 7, No 1, pp. 29–35, 2014.