

# STUDI KOMPARATIF MODEL *SUPPORT VECTOR REGRESSION* DAN *RANDOM FOREST* DALAM MEMPREDIKSI HARGA SAHAM BMRI

\*Bintang Nauli Mariana, I Wayan Mangku, dan I Gusti Putu Purnaba

Program Studi Matematika, Sekolah Sains Data, Matematika, dan Informatika,  
Institut Pertanian Bogor, Jl. Meranti, Kampus IPB Dramaga Bogor.  
[bintangnaulishnbintang@apps.ipb.ac.id](mailto:bintangnaulishnbintang@apps.ipb.ac.id) \*corresponding author,  
[wayanma@apps.ipb.ac.id](mailto:wayanma@apps.ipb.ac.id), [purnaba@apps.ipb.ac.id](mailto:purnaba@apps.ipb.ac.id)

## Abstrak

Berbagai faktor dapat memengaruhi dinamika pasar modal di Indonesia, khususnya dalam periode 2020-2025 sehingga menyebabkan harga saham menjadi fluktuatif dan sulit diprediksi dengan sederhana. Oleh karena itu, diperlukan model yang mampu memprediksi harga saham agar investor dapat mengambil keputusan dengan tepat. Penelitian ini bertujuan untuk membandingkan kinerja model *support vector regression* dan *random forest* dalam memprediksi harga penutupan saham BMRI dan memprediksi harga saham untuk satu hari ke depan. Penelitian dilakukan dengan mengambil data historis harga saham untuk digunakan dalam membangun model prediksi melalui penyesuaian parameter dan mengevaluasi model menggunakan ukuran kesalahan prediksi. Hasil penelitian menunjukkan bahwa model *support vector regression* menunjukkan kinerja lebih baik dalam memprediksi harga saham BMRI, dengan nilai RMSE sebesar 39,01, MAPE 0,56% dan  $R^2$  sebesar 0,99. Selain itu, prediksi harga saham BMRI untuk satu hari ke depan sebesar Rp4.410,79 per lembar.

**Kata kunci:** harga saham, *machine learning*, prediksi, *random forest*, *support vector regression*

## 1 Pendahuluan

Di zaman sekarang, investasi menjadi bagian yang penting untuk menjaga kestabilan finansial, baik bagi individu maupun institusi. Salah satu instrumen investasi yang paling diminati adalah saham, karena dapat memberikan keuntungan yang tinggi, namun risiko yang besar. Berinvestasi dapat dilakukan di pasar modal, salah satu pasar modal di Indonesia adalah Bursa Efek Indonesia (BEI). Berbagai faktor dapat memengaruhi dinamika pasar modal di Indonesia, khususnya dalam periode 2020-2025 sehingga menyebabkan harga saham menjadi fluktuatif dan sulit diprediksi dengan sederhana.

Pada Oktober 2020 hingga September 2025, terjadi berbagai fenomena penting dalam perekonomian Indonesia, yaitu pandemi COVID-19, fase pemulihan ekonomi, hingga muncul ketidakpastian global akibat tekanan inflasi dan kebijakan pemerintah.

Fenomena tersebut tentu berdampak pada sektor perbankan, di mana saham yang terdampak cukup signifikan adalah saham PT Bank Mandiri Tbk (BMRI) yang merupakan salah satu bank terbesar di Indonesia. Pada periode ini, harga saham BMRI menunjukkan pola pergerakan yang dinamis. Hal ini menunjukkan bahwa banyaknya faktor yang saling berinteraksi secara kompleks dapat memengaruhi harga saham, sehingga perlu dilakukan analisis prediksi.

Prediksi harga saham penting untuk diteliti agar investor dapat mengambil keputusan yang tepat, baik dalam memperkirakan risiko maupun peluang investasi. Berbagai model telah dikembangkan untuk memprediksi harga saham, mulai dari statistika klasik hingga pendekatan berbasis *machine learning*. Beberapa model yang banyak digunakan adalah *Support Vector Regression* (SVR) dan *Random Forest* (RF). SVR adalah bagian dari metode SVM, yakni sebuah model matematika berbasis *machine learning* yang mampu menangani data non-linear dan mampu mengatasi *overfitting* [1]. SVR menggunakan konsep regresi sebagai optimasi kuadratik dengan fungsi *kernel* sehingga data dapat direpresentasikan ke dimensi yang lebih tinggi. Sementara itu, *random forest* merupakan model regresi yang membangun banyak pohon keputusan berdasarkan statistika dan teori peluang untuk meningkatkan akurasi prediksi [2].

Kedua model ini banyak digunakan dalam penelitian terdahulu karena keunggulannya dalam menghadapi data yang kompleks dan fluktuatif. Misalnya dalam penelitian yang dilakukan oleh [3] menunjukkan bahwa SVR mampu menghasilkan prediksi harga saham dengan tingkat akurasi yang baik terutama pada data yang memiliki pola nonlinear. Penelitian tersebut menyarankan untuk mengeksplorasi teknik optimasi parameter dan pembaruan model secara dinamis untuk meningkatkan akurasi prediksi. Penelitian selanjutnya dilakukan oleh [4] di mana diperoleh hasil bahwa kinerja model prediksi dapat berbeda pada setiap kondisi pasar dan karakteristik data, sehingga pengujian berbagai model pada objek dan periode pengamatan yang berbeda masih diperlukan. Selanjutnya, model *random forest* juga sudah pernah diteliti yakni [5] di mana model *random forest* menghasilkan performa yang baik dalam memprediksi harga saham karena mampu menangani pola hubungan yang kompleks antarvariabel. Penelitian tersebut kemudian menyarankan untuk memanfaatkan indikator tambahan untuk meningkatkan akurasi prediksi. Selain itu, penelitian mengenai *random forest* juga dilakukan oleh [6] dengan hasil model *random forest* sangat dipengaruhi oleh karakteristik data. Penelitian tersebut menyarankan untuk melakukan perbandingan antaralgoritma model pada data dan periode yang sama untuk mengetahui model yang paling sesuai. Berdasarkan penelitian terdahulu, kedua model menunjukkan performa cukup baik dengan karakteristik dan keunggulan yang berbeda dalam memprediksi harga saham, sehingga diperlukan kajian lebih lanjut untuk mengevaluasi kinerjanya pada objek dan periode pengamatan yang sama.

Oleh karena itu, penelitian ini bertujuan membandingkan kinerja model *Support Vector Regression* (SVR) dan *Random Forest* (RF) untuk memprediksi harga saham BMRI serta melakukan prediksi harga saham satu hari ke depan. Pemilihan kedua model tersebut didasarkan pada perbedaan pendekatan yang digunakan, di mana SVR merupakan model berbasis *kernel* yang efektif dalam memodelkan hubungan data nonlinear, sedangkan *random forest* merupakan model *ensemble learning* yang mampu menangani data kompleks juga mengurangi risiko *overfitting*. Meskipun keduanya telah menunjukkan performa yang baik pada penelitian terdahulu, penelitian yang membandingkan secara langsung model *support vector regression* dan *random forest* pada saham BMRI selama periode 2020-2025 masih terbatas. Penelitian ini berfokus pada

harga penutupan saham BMRI selama lima tahun terakhir yang belum banyak diteliti sebelumnya, serta melakukan perbandingan performa secara kuantitatif berbasis kesalahan prediksi seperti RMSE, MAPE dan  $R^2$ . Dengan demikian, diharapkan penelitian ini dapat memberikan wawasan mengenai model mana yang memberikan tingkat kesalahan lebih kecil dan akurasi prediksi lebih tinggi, sehingga dapat menjadi acuan yang lebih andal bagi investor maupun akademisi dalam mengambil keputusan investasi di sektor perbankan.

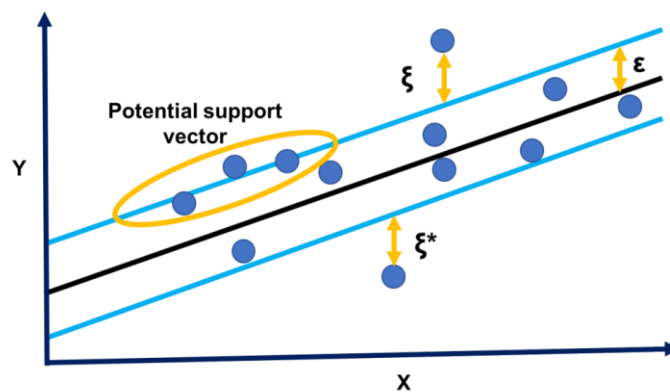
## 2 Metode Penelitian

### 2.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah harga penutupan saham harian PT Bank Mandiri Tbk dengan kode saham BMRI. Data diambil dalam rentang 1 Oktober 2020 hingga 30 September 2025 dengan jumlah 1200 Data yang diperoleh dari laman *finance.yahoo.com*. Analisis data pada penelitian ini menggunakan Python.

### 2.2 Support Vector Regression

SVR digunakan untuk kasus regresi dengan memprediksi nilai kontinu yang ditemukan pada tahun 1999 oleh Vladimir N. Vapnik [7]. Visualisasi SVR dapat dilihat pada Gambar 1 di bawah ini.



Gambar 1 Visualisasi *Support Vector Regression*  
Sumber: <https://snowflect.com/regression/>

Dalam model *support vector regression*, fungsi ditulis dengan bentuk sebagai berikut:

$$f(x) = \omega^T x + b \quad (1)$$

dengan  $\omega$  vektor bobot,  $x$  vektor input, dan  $b$  bias. Fungsi pada Persamaan 1 memiliki arti bahwa SVR bertujuan dalam mencari fungsi regresi yang sederhana. Hal ini dapat dilakukan dengan meminimumkan nilai norm dari vektor bobot ( $\omega$ ) yaitu  $\omega^T \omega$ . Jika nilai dari norm semakin kecil, maka akan semakin sederhana model yang dihasilkan, sehingga permasalahan optimasi tersebut dapat dituliskan dalam bentuk Persamaan 2 sebagai berikut:

$$\min \frac{1}{2} \|\omega\|^2. \quad (2)$$

Selanjutnya, terdapat konsep *e-sensitive loss function*. Konsep ini mengukur sejauh mana tingkat kesalahan dapat ditoleransi. Jika ada kesalahan (*error*) yang nilainya masih di dalam batas toleransi  $\varepsilon$ , maka tidak akan dikenakan penalti. Konsep tersebut didefinisikan oleh fungsi pada Persamaan 3 sebagai berikut:

$$L_\varepsilon(y, f(x)) = \begin{cases} 0 & , \text{jika } |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon, & \text{jika } |y - f(x)| > \varepsilon \end{cases} \quad (3)$$

dengan  $y$  merupakan nilai observasi/aktual,  $f(x)$  adalah nilai prediksi dari model SVR, dan  $\varepsilon$  merupakan batas toleransi kesalahan. Dengan demikian, hanya kesalahan yang melebihi  $\varepsilon$  yang akan diperhitungkan dalam pembelajaran model, sehingga SVR akan lebih tahan terhadap *noise* dalam data ataupun fluktuasi dalam data. Akan tetapi, dalam pengaplikasiannya, tidak semua data dapat diprediksi dengan tingkat kesalahan kurang dari  $\varepsilon$ . Oleh karena itu, perlu dilakukan penambahan variabel *slack* berupa  $\xi$  dan  $\xi^*$  untuk memfasilitasi data yang melanggar batas  $\varepsilon$ . Dengan demikian, persamaan optimasi SVR dapat ditulis sebagai:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi + \xi^*) \quad (4)$$

dengan kendala:

$$\begin{cases} y_i - f(x) \leq \varepsilon + \xi_i, \\ f(x) - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i + \xi_i^* \geq 0. \end{cases}$$

Dalam Persamaan 4, parameter  $C > 0$  berperan sebagai pengendali *trade-off* antara kerataan fungsi dan tingkat toleransi terhadap batas kesalahan  $\varepsilon$ . Penalti besar akan diberlakukan terhadap *error* jika nilai  $C$  besar, sehingga model akan lebih ketat mengikuti data. Sebaliknya, jika nilai  $C$  kecil, model akan lebih toleran terhadap *error*, namun mungkin saja kurang akurat.

Solusi dari permasalahan optimasi dalam *support vector regression* dapat dicari menggunakan metode pengali Lagrange seperti pada Persamaan 5 di bawah ini.

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i - y_i + f(x_i)) \\ & - \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* + y_i - f(x_i)) \end{aligned} \quad (5)$$

dengan  $\alpha_i$ ,  $\alpha_i^*$ ,  $\xi_i$ , dan  $\xi_i^*$  merupakan pengali Lagrange yang digunakan untuk menyelesaikan optimasi dengan kendala. Kendala ini akan diubah menjadi masalah optimasi tanpa kendala. Hal ini dapat diselesaikan dengan menggunakan Kondisi Karush-Kuhn-Tucker [7]. Kondisi ini memastikan bahwa nilai  $\alpha_i - \alpha_i^*$  yang diperoleh adalah solusi optimum dari fungsi Lagrange. Dalam penelitian ini, optimasi SVR tidak dilakukan secara analitik, melainkan dengan *solver* bawaan pustaka *scikit-learn* yang mampu menyelesaikan masalah dual dengan memenuhi kondisi KKT. Dengan demikian, persamaan berikut disajikan sebagai dasar teoritis untuk menjelaskan parameter model.

Oleh karena itu, dengan menurunkan fungsi Lagrange terhadap parameter  $w$  dan  $b$  diperoleh Persamaan 6 dan 7 di bawah ini.

$$w = \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i, \quad (6)$$

$$\sum_{i=1}^n \alpha_i - \sum_{i=1}^n \alpha_i^* = 0. \quad (7)$$

Langkah selanjutnya adalah nilai bias  $b$  dapat diperoleh dari salah satu *support vector* yang memenuhi syarat  $0 < \alpha_i < C$  atau  $0 < \alpha_i^* < C$ , seperti Persamaan 8a dan 8b berikut:

$$b = y_i - w^T x_i - \varepsilon, \quad 0 < \alpha_i < C \quad (8a)$$

$$b = y_i - w^T x_i - \varepsilon, \quad 0 < \alpha_i^* < C \quad (8b)$$

dengan  $x_i$  merupakan vektor peubah  $x$  periode ke- $i$ . Kemudian, model optimasi SVR dapat ditulis kembali dalam bentuk dual seperti Persamaan 9 di bawah ini.

$$\max \left[ -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) k(x_i, x_j) - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) \right] \quad (9)$$

dengan kendala sebagai berikut:

$$\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, \quad (9a)$$

$$\alpha_i, \alpha_i^* \in [0, C], i = 1, 2, 3, \dots, n. \quad (9b)$$

Menurut Teorema Representer [8], dalam *support vector regression*, solusi optimal dapat dinyatakan sebagai kombinasi linear dari fungsi *kernel* yang bergantung pada data *training*. Dengan demikian, diperoleh bentuk akhir dari persamaan (1) yang merupakan hasil dari teorema Representer, di mana setiap solusi dari  $f(x)$  dapat ditulis sebagai persamaan (10) berikut:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(x_i, x_j) + b \quad (10)$$

dengan keterangan:  $\alpha_i - \alpha_i^*$  = pengali Lagrange (data *training* tersimpan pada parameter ini);  $k(x_i, x_j)$  = fungsi *kernel*;  $b$  = bias.

Dalam kasus regresi, harus ditentukan parameter epsilon dan *cost* ( $\varepsilon$  dan  $C$ ). Pemilihan parameter ini akan memengaruhi terbentuknya model akhir SVR. Selain itu, SVR juga menggunakan fungsi *kernel* guna memetakan langsung input ke ruang fitur berdimensi tinggi lalu diselesaikan secara linear. Beberapa contoh *kernel* di antaranya linear, polinomial, dan *Radial Based Function* (RBF), dengan masing-masing *kernel* memiliki fungsi yang tertuang dalam persamaan (11), (12), dan (13) berikut:

Linear:

$$K(x_i, x_j) = x_i^T x_j. \quad (11)$$

Polinomial:

$$K(x_i, x_j) = (\gamma x_i^T x_j + 1)^d, d > 1. \quad (12)$$

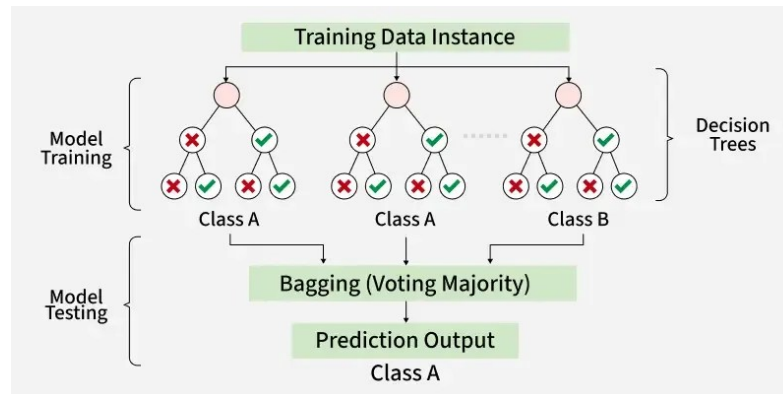
Radial Based Function (RBF):

$$K(x_i, x_j) = \exp(-\gamma |x_i - x_j|^2). \quad (13)$$

Berdasarkan Teorema Mercer [9], fungsi *kernel*  $K(x_i, x_j)$  dapat digunakan untuk memetakan data ke ruang fitur berdimensi tinggi apabila memenuhi sifat simetri dan positif semidefinit. Teorema ini menyatakan bahwa setiap *kernel* yang memiliki kedua sifat tersebut dapat diinterpretasikan sebagai hasil kali dalam suatu *Reproducing Kernel Hilbert Space* (RKHS). Dengan demikian, permasalahan yang bersifat non-linear di ruang asli dapat diubah menjadi bentuk linear di ruang fitur. Konsekuensinya, penerapan *kernel* seperti linear, polinomial, dan *radial basis function* (RBF) pada metode SVR tetap menjamin adanya solusi tunggal dalam proses optimasinya.

### 2.3 Random Forest

*Random forest* pertama kali diperkenalkan oleh Breiman pada tahun 2001 [10]. Dalam pengaplikasiannya, *random forest* menggabungkan banyak pohon keputusan (*decision tree*) untuk meningkatkan akurasi prediksi. *Random forest* membangun pohon keputusan secara acak, lalu hasil akhir ditentukan dengan regresi, sehingga mampu mengurangi ragam dan mengatasi kelemahan yang biasanya muncul pada pohon tunggal [11]. Visualisasi *random forest* dapat dilihat pada Gambar 2 di bawah ini.



Gambar 2 Algoritma *Random Forest*

Sumber: <https://www.geeksforgeeks.org/machine-learning/random-forest-algorithm-in-machine-learning/>

Metode *sampling bootstrap* digunakan untuk membangun setiap pohon, yaitu pengambilan sampel secara acak dengan pengembalian dari dataset utama. Metode ini menghasilkan variasi antara pohon, sehingga mengurangi korelasi dan meningkatkan kemampuan generalisasi model. Hasil prediksi dari seluruh pohon digabungkan dengan cara dirata-ratakan. Rumus dari pembentukan *decision tree* sebagai berikut:

$$D_t = \{ (x_i, y_i) \mid i \in S_t \} \quad (14)$$

dengan  $D_t$  adalah dataset *bootstrap* ke- $t$  dan  $(x_i, y_i)$  merupakan data fitur, target, dan  $S_t$  adalah sampel acak hasil *bootstrap* dari dataset awal  $D_t$ . Pada Persamaan 14, Berdasarkan Teorema Konsistensi *random forest regression* [12], pembentukan banyak pohon secara acak dan independen menjamin bahwa ketika jumlah pohon  $T \rightarrow \infty$  serta  $n \rightarrow \infty$ , maka hasil prediksi dari model akan konvergen ke nilai ekspektasi dari variabel target, yaitu:

$$f_{RF}(x) \rightarrow E[Y \mid X = x]. \quad (15)$$

Persamaan (15) bermakna bahwa semakin banyak jumlah pohon dan data, semakin stabil hasil prediksi yang dihasilkan model.

Selanjutnya, dalam setiap pembentukan simpul pada pohon keputusan, *random forest* memilih subset fitur secara acak (*feature bagging*) dalam setiap pembentukan simpul pada pohon keputusan untuk menghasilkan pembagian yang terbaik dengan *Mean Squared Error (MSE)*. Strategi ini mengakibatkan peningkatan akurasi model secara keseluruhan dan mengurangi korelasi antar pohon. Secara matematis, untuk setiap node dipilih subset fitur  $m < p$ , akan ditentukan fitur yang menghasilkan *MSE* paling kecil, dengan rumus sebagai berikut:

$$MSE(j, s) = N_L \frac{1}{N_L} \sum_{x_i \in L} (y_i - y_L)^2 + N_R \frac{1}{N_R} \sum_{x_i \in R} (y_i - y_R)^2, \quad (16)$$

dengan:

$L$  dan  $R$  : himpunan data hasil pemisahan di sisi kiri dan kanan,

$N_L, N_R$  : jumlah sampel masing-masing sisi,

$y_L, y_R$  : rata-rata target pada sisi kiri dan kanan.

Persamaan (16) memiliki arti bahwa fitur serta titik pemisahan yang menghasilkan nilai  $MSE(j, s)$  paling kecil kemudian dipilih sebagai *split* terbaik. Berdasarkan teorema Batas Generalisasi *Random Forest* [10], akurasi model bergantung pada 2 hal, yaitu korelasi antar pohon ( $\rho$ ), dan kekuatan rata-rata tiap pohon ( $s$ ). Hal tersebut dirumuskan dalam persamaan (17) sebagai berikut:

$$PE \leq \frac{\rho(1 - s^2)}{s^2}. \quad (17)$$

Jika kita menyeleksi subset fitur secara acak di setiap node-nya, maka korelasi antar pohon ( $\rho$ ) dapat dikurangi. Hal ini mengakibatkan *error* akan menurun dan model akan lebih stabil. Setelah seluruh pohon keputusan terbentuk, hasil prediksi dari masing-masing pohon dirata-rata untuk mendapatkan hasil prediksi akhir. Hasil estimasi harga saham yang lebih stabil dengan mengurangi ragam antar model dapat diperoleh dengan pendekatan *averaging*. Rumus prediksi akhir dari model *random forest* dinyatakan dalam persamaan (18) sebagai berikut:

$$y(x) = \frac{1}{T} \sum_{t=1}^T y_t(x), \quad (18)$$

dengan  $T$  adalah jumlah pohon dalam hutan, dan  $y_t(x)$  adalah hasil prediksi pohon ke- $t$  untuk input ke- $x$ .

Terdapat teorema yang memperjelas rumus prediksi akhir dari model *random forest*, yaitu teorema Reduksi Varians (*Bagging Theorem*) oleh [10]. Jika korelasi rata-rata  $\rho$  dan setiap pohon memiliki varians  $\sigma^2$ , maka varians keseluruhan hasil rata-rata yaitu:

$$\text{Var}(f_{RF}) = \rho\sigma^2 + \frac{1-\rho}{T}\sigma^2. \quad (19)$$

Persamaan (19) memiliki makna bahwa jika jumlah pohon  $T$  semakin banyak dan korelasi antar pohon  $\rho$  semakin rendah, maka total varians pada model akan menurun secara signifikan sehingga prediksi yang dihasilkan akan lebih stabil dan tidak *overfitting*.

Terdapat beberapa parameter penting yang memengaruhi model *random forest regression*, yaitu *n\_estimators*, *max\_features*, *max\_depth*, *min\_samples\_split*, dan *min\_samples\_leaf* di mana nilai terbaik akan diperoleh melalui proses *tuning parameter*.

## 2.4 Tahapan Penelitian

Penelitian ini menganalisis pola data penutupan harga saham harian pada PT Bank Mandiri Tbk (BMRI) dengan rentang waktu lima tahun agar dapat membandingkan dua model sehingga diperoleh model terbaik dalam memprediksi harga saham. Langkah-langkah penelitian adalah sebagai berikut:

1. Mengumpulkan data harga penutupan saham harian BMRI dengan rentang waktu 1 Oktober 2020 hingga 30 September 2025.
2. Input data.
3. Analisis deskriptif.
4. *Preprocessing* data:
  - a. Penentuan variabel penelitian.
  - b. Pembentukan variabel turunan (MA, EMA, Log Return).
  - c. Transformasi variabel target (*exponential smoothing*).
5. Membagi data menjadi data *training* sebanyak 70% , *validation* 15%, data *testing* sebanyak 15% secara berurutan.
6. Standarisasi data menggunakan *StandardScaler* pada pemodelan SVR, di mana parameter standarisasi ditentukan berdasarkan data *training* dan diterapkan pada data *validation* serta *testing*.
7. Melakukan pemodelan *support vector regression*
  - a. Uji *kernel* terbaik (linear, polinomial, RBF).
  - b. *Tuning* parameter.
  - c. Pembentukan model SVR terbaik.
8. Melakukan pemodelan *random forest*:
  - a. *Tuning* parameter menggunakan *TimeSeriesSplit*.
  - b. Pembentukan model *random forest* terbaik.
9. Evaluasi kinerja model SVR dan *random forest* dengan membandingkan nilai RMSE, MAPE,  $R^2$  kedua model.
10. *Postprocessing* data hasil prediksi dengan inversi ke skala asli.
11. Prediksi harga saham satu hari ke depan.

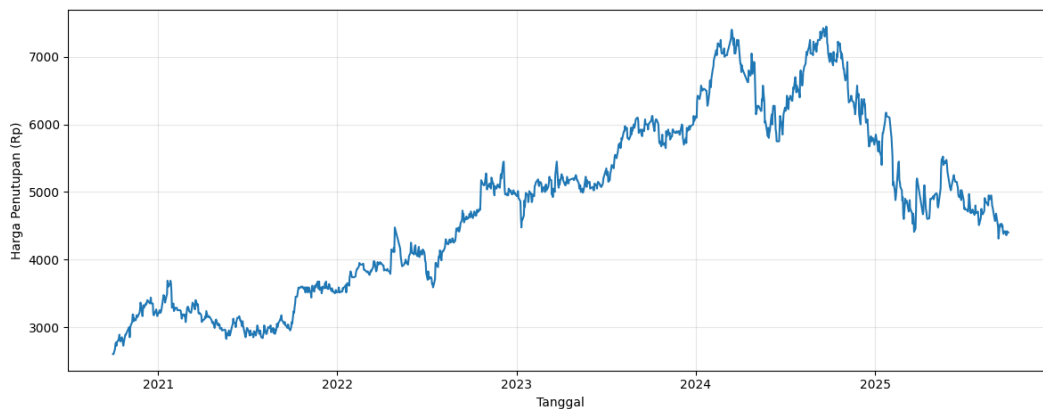
### 3 Hasil dan Pembahasan

#### 3.1 Analisis Deskriptif

Hasil dari analisis deskriptif pada harga penutupan saham BMRI dapat dilihat pada Tabel 1, sedangkan untuk hasil plot ditampilkan pada Gambar 3.

Tabel 1 Hasil analisis deskriptif

Saham	$N$	Mean	Minimum	Maximum	Std. Deviation
BMRI	1200	4.802,37	2.600,00	7.450,00	1.285,37



Gambar 3 Grafik harga saham penutupan harian BMRI

Tabel 1 dan Gambar 3 menunjukkan bahwa harga saham penutupan terendah (minimum) dalam periode 1 Oktober 2020 hingga 30 September 2025 adalah Rp2.600,00 per lembar, tepatnya pada tanggal 1 Oktober 2020, sedangkan harga penutupan saham tertinggi (maksimum) mencapai Rp7.450,00 per lembar yang terjadi pada tanggal 23 September 2024. Pada penelitian ini, diperoleh harga rata-rata penutupan saham (*mean*) BMRI sebesar Rp4.802,37 per lembar dengan standar deviasi sebesar 1.285,37 per lembar. Gambar 3 menunjukkan bahwa pola data deret waktu yang terbentuk menunjukkan harga penutupan saham BMRI memiliki suatu kecenderungan naik turun dalam waktu tertentu. Hal ini dapat disebabkan oleh faktor internal dan eksternal perusahaan.

#### 3.2 Preprocessing Data

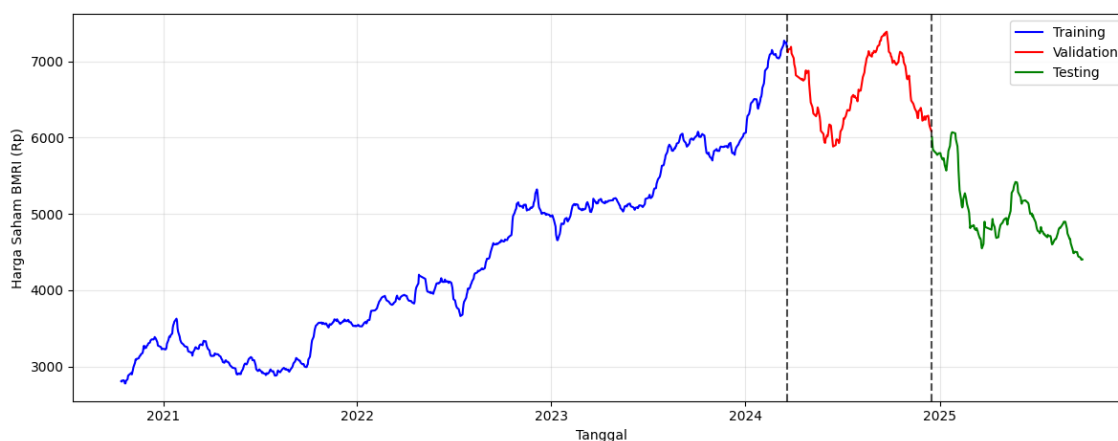
Pada penelitian ini, variabel input yang digunakan yaitu harga penutupan BMRI. Pemilihan variabel ini didasarkan pada relevansi terhadap pergerakan harga penutupan saham BMRI, lalu variabel output ( $Y$ ) didefinisikan sebagai harga saham BMRI satu periode ke depan. Selanjutnya, variabel turunan dibuat untuk meningkatkan data yang digunakan oleh model guna mempelajari pola pergerakan harga saham BMRI. Variabel turunan yang digunakan termasuk *moving average* periode 5 dan 10 hari, *exponential moving average* periode 5 dan 10 hari, dan *log return*. Berdasarkan 1.200 observasi harga penutupan saham BMRI, pembentukan variabel *moving average* periode 5 menghasilkan 1.196 observasi, sedangkan *moving average* periode 10 menghasilkan 1.191 observasi. Perbedaan jumlah observasi tersebut disebabkan oleh kebutuhan data historis pada beberapa periode awal dalam proses perhitungannya. Sementara itu, variabel *exponential moving average* periode 5 dan 10 hari masing-masing menghasilkan 1.200 observasi, sedangkan variabel *log return* menghasilkan 1.199 observasi. Setelah seluruh variabel

input dan output ditentukan, tahap selanjutnya adalah transformasi variabel *output* dengan *exponential smoothing*, yang dilakukan dengan memberikan bobot yang lebih besar pada data harga penutupan terbaru dibandingkan data sebelumnya, sehingga menghasilkan variabel output yang lebih halus namun tetap merepresentasikan tren pergerakan harga saham.

Proses pembentukan variabel turunan menyebabkan jumlah observasi yang digunakan dalam pemodelan berkurang dari 1.200 menjadi 1.190 observasi. Hal tersebut terjadi dikarenakan perhitungan *moving average* periode 10 hari memerlukan data historis beberapa periode awal sehingga nilai variabel belum dapat dihitung secara lengkap, sedangkan pembentukan variabel output untuk prediksi satu hari ke depan menyebabkan observasi pada periode terakhir tidak dapat digunakan dalam proses pemodelan.

### 3.3 Pembagian Data

Data kemudian dibagi menjadi data *training*, *validation*, dan data *testing*. Dalam *support vector regression* dan *random forest*, data *training* digunakan untuk meningkatkan kinerja model terhadap data *testing*, serta untuk menentukan parameter terbaik (*kernel*) untuk pembentukan model. Kemudian, data *testing* yang merupakan data baru digunakan untuk menguji hasil dari masing-masing model yang terbentuk dari pelatihan di data *training*. Detail pembagian data *training* dan data *testing* dapat dilihat pada Gambar 4 dan Tabel 2 berikut.



Gambar 4 Pembagian data *training*, *validation*, dan *testing*

Tabel 2 Pembagian data *training*, *validation*, dan *testing*

Nama Saham	Data			Total
	<i>Training</i>	<i>Validation</i>	<i>Testing</i>	
BMRI	833	178	179	1190
Persentase	70%	15%	15%	100%

Berdasarkan Gambar 4 dan Tabel 2, dapat dilihat bahwa pembagian data yang digunakan untuk penelitian ini sejumlah 70% untuk data *training* (periode 14 Oktober 2020 – 19 Maret 2024), 15% untuk data *validation* (periode 20 Maret 2024 – 16 Desember 2024) dan data *testing* sebanyak 15% (periode 17 Desember 2024 – 26 September 2025). Data *training* memiliki persentase yang lebih besar agar model dapat berlatih dengan baik.

Pembagian data ini menggunakan *python*, khususnya fungsi *time series split* dengan teknik pengindeksan (*index slicing*).

### 3.4 Standarisasi Data

Data kemudian akan distandarisasi menggunakan *python* dengan library *StandardScaler* yang disediakan oleh *scikit-learn* dapat dilihat pada Tabel 3 berikut.

Tabel 3 Hasil standarisasi variabel input dan output harga saham harian BMRI

No	Close	MA 5	MA 10	EMA 5	EMA 10	Log Return	Y Scaled
1	1,346	1,490	1,596	1,472	1,559	-1,281	1,410
2	1,238	1,408	1,548	1,400	1,508	-1,307	1,324
3	1,109	1,330	1,507	1,309	1,442	-1,590	1,263
4	1,109	1,269	1,458	1,248	1,388	-0,066	1,260
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
176	0,019	0,044	0,079	0,043	0,085	-0,198	0,033
177	0,001	0,023	0,079	0,032	0,074	-0,332	0,017
178	-0,024	0,013	0,068	0,016	0,060	-0,467	0,022
179	0,027	0,019	0,059	0,023	0,058	0,733	0,020

Tabel 3 menampilkan hasil standarisasi pada data *testing* yang terdiri atas 179 observasi. Penyajian dibatasi pada lima observasi pertama dan lima observasi terakhir untuk merepresentasikan hasil transformasi data yang digunakan dalam evaluasi model. Proses standarisasi tetap dilakukan pada seluruh data menggunakan parameter rata-rata dan simpangan baku yang diperoleh dari data *training*, lalu diterapkan pada data *validation* dan *testing* untuk menghindari data *leakage*. Hasil standarisasi pada Tabel 3 menunjukkan bahwa setiap variabel telah berada pada skala yang sama, dengan nilai positif dan negatif. Nilai negatif menunjukkan bahwa data lebih rendah dari nilai rata-rata historis, sedangkan nilai positif menunjukkan bahwa data lebih tinggi dari nilai rata-rata. Ini adalah karakteristik normal metode *StandardScaler* dan tidak memengaruhi makna ekonomi dari data harga saham.

### 3.5 Model Support Vector Regression

Pada tahap awal pemodelan SVR, dilakukan pengujian terhadap 3 (tiga) jenis *kernel* yang berbeda, yaitu linear, polinomial dan *Radial Basis Function* (RBF). Masing-masing *kernel* digunakan untuk membentuk model pada data *training*, kemudian kinerjanya dievaluasi menggunakan data *validation* berdasarkan RMSE, MAPE, dan  $R^2$ . Hasil evaluasi digunakan sebagai dasar dalam menentukan *kernel* yang paling sesuai sebelum dilakukan proses *tuning* parameter. Hasil dari uji *kernel* pada SVR disajikan dalam Tabel 4.

Berdasarkan uji yang telah dilakukan, diperoleh kernel linear sebagai kernel terbaik dengan nilai pada masing-masing metrik evaluasi terendah dibanding kedua kernel lainnya. Selanjutnya, kernel linear akan digunakan untuk mencari parameter terbaik. Berdasarkan *tuning* parameter, diperoleh nilai parameter  $C$  sebesar 10, dan  $\epsilon$  sebesar  $10^{-2}$  yang selanjutnya digunakan untuk membuat prediksi harga penutupan saham BMRI. Parameter ini menghasilkan kinerja terbaik pada data *training* dan *validation*. Hasil evaluasi model ditunjukkan oleh Tabel 5 dan Gambar 5.

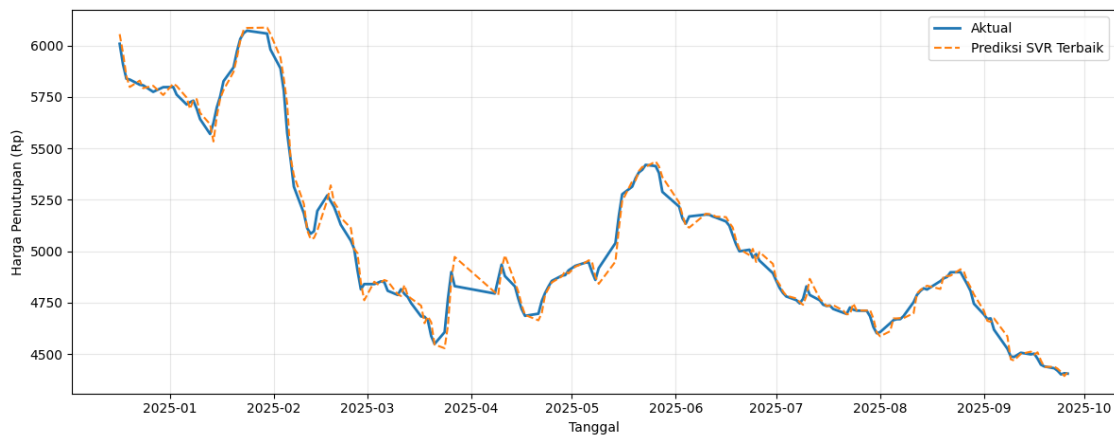
Tabel 4 Hasil uji kernel pada model *Support Vector Regression* (SVR)

Jenis <i>Kernel</i>	Metrik Evaluasi		
	RMSE	MAPE	R <sup>2</sup>
Linear	47,20	0,59	0,99
Polinomial	3.611,20	42,77	-71,82
RBF	165,28	1,56	0,85

Tabel 5 Perbandingan nilai aktual dan prediksi SVR (data *testing*)

No	Tanggal	Aktual	Prediksi
1	2024-12-17	6.008,19	6.052,52
2	2024-12-18	5.908,23	5.961,30
3	2024-12-19	5.838,26	5.848,87
4	2024-12-20	5.834,28	5.795,07
⋮	⋮	⋮	⋮
176	2025-09-23	4.418,54	4.426,60
177	2025-09-24	4.400,98	4.411,53
178	2025-09-25	4.406,68	4.391,33
179	2025-09-26	4.404,68	4.410,79

Tabel 5 menunjukkan perbandingan antara nilai hasil prediksi harga penutupan saham BMRI yang dihasilkan menggunakan model *Support Vector Regression* (SVR) terbaik pada data *testing*. Nilai aktual sebesar Rp6.008,19 dan hasil prediksi SVR sebesar Rp6.052,52 berbeda sedikit pada awal periode *testing*, tepatnya pada 17 Desember 2024. Pada beberapa tanggal berikutnya, pola kedekatan antara nilai aktual dan prediksi semakin terlihat. Pada tanggal 26 September 2025, pada akhir periode *testing*, harga penutupan aktual tercatat sebesar Rp4.404,68, sedangkan hasil prediksi SVR sebesar Rp4.410,79. Secara keseluruhan, temuan ini menunjukkan bahwa model SVR mampu mengikuti pola pergerakan harga saham dengan cukup baik dan menghasilkan prediksi yang hampir sama dengan nilai aktual berdasarkan data *testing*.



Gambar 5 Grafik perbandingan harga aktual dan prediksi SVR

Berdasarkan Gambar 5, dapat dilihat bahwa pergerakan harga saham BMRI menunjukkan fluktuasi yang cukup signifikan selama beberapa periode. Selain itu, terjadi

penurunan harga cukup tajam pada akhir Februari hingga awal April 2025, saat harga aktual turun dari kisaran Rp6.100,00 ke Rp4.700,00. Prediksi model SVR mengikuti pola penurunan tersebut, meskipun sedikit lebih halus. Selanjutnya, kenaikan harga cukup kuat dari Mei hingga Juni 2025, mencapai sekitar Rp5.300,00. Namun, terjadi penurunan harga secara bertahap di bulan Juli hingga September 2025 dan mendekati Rp4.400,00 pada akhir periode. Secara keseluruhan, SVR mampu menangkap arah perubahan harga dengan baik, hasil prediksi cenderung lebih halus dibandingkan harga aktual dalam beberapa periode dengan pergerakan harga yang fluktuatif.

### 3.6 Model *Random Forest*

Pada model *random forest*, kompleksitas model diseimbangkan menggunakan *tuning* parameter. Model dibangun menggunakan data *training* dan selanjutnya dilakukan *tuning* parameter menggunakan *GridSearchCV*. Kinerja setiap kombinasi parameter dievaluasi pada data *validation* untuk memperoleh parameter yang memberikan performa model terbaik. Parameter terpilih kemudian digunakan untuk membentuk model akhir yang akan dievaluasi pada data *testing*. Hasil *tuning* parameter disajikan pada Tabel 6 berikut.

Tabel 6 *Tuning* parameter *random forest*

Parameter	Nilai Terbaik
<i>n_estimators</i>	300
<i>max_features</i>	<i>sqrt</i>
<i>max_depth</i>	10
<i>min_samples_split</i>	2
<i>min_samples_leaf</i>	1

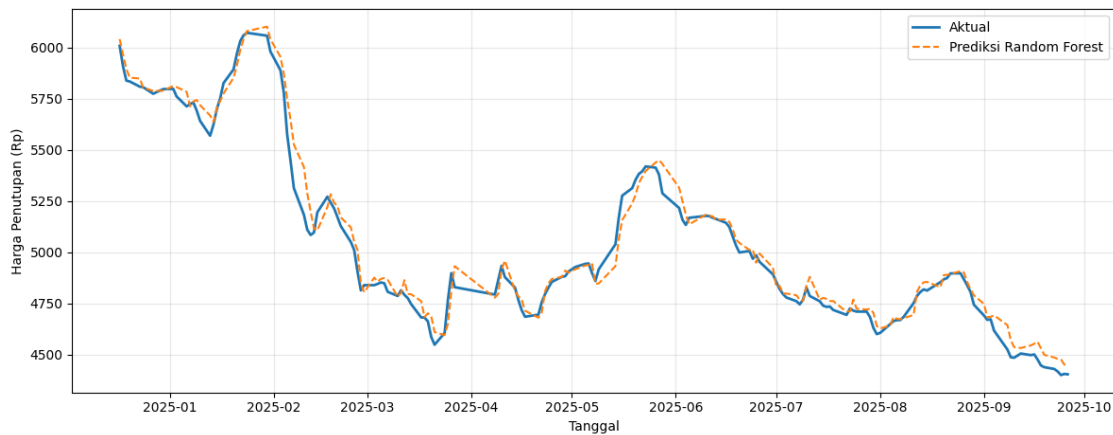
Berdasarkan kombinasi parameter pada Tabel 6, diperoleh nilai *cross-validation* MAPE pada data *validation* sebesar 5,1%, yang menunjukkan model *random forest* memiliki tingkat kesalahan yang relatif rendah pada data *validation*. Nilai MAPE yang kecil mengindikasikan bahwa kombinasi parameter yang dihasilkan sebelumnya mampu membuat model yang stabil dan tidak *overfitting*.

Parameter terbaik yang telah diperoleh selanjutnya digunakan untuk membentuk model *random forest*. Model kemudian dilatih kembali menggunakan seluruh data *training* agar mempelajari pola data secara optimal. Setelah proses *training* selesai, model diterapkan pada data *testing* yang tidak terlibat dalam proses *training* maupun pemilihan parameter untuk menghasilkan prediksi harga saham BMRI. Hasil prediksi pada data *testing* disajikan pada Tabel 7 dan Gambar 6 berupa perbandingan antara nilai aktual dan prediksi.

Berdasarkan Tabel 7, dapat dilihat pada awal periode *testing* tepatnya tanggal 17 Desember 2024, tercatat harga aktual sebesar Rp6.008,19, sementara hasil prediksi model *random forest* sebesar Rp6.040,96. Pada beberapa periode selanjutnya, hasil prediksi harga penutupan saham BMRI cenderung berada di atas harga aktual, namun tetap mengikuti tren pergerakan harga saham. Pada akhir periode *testing*, tepatnya 26 September 2025, harga penutupan aktual tercatat sebesar Rp4.404,68 sementara hasil prediksi sebesar Rp4.454,27. Berdasarkan hasil prediksi tersebut, model *random forest* mampu mengikuti pergerakan harga saham BMRI dengan cukup baik, namun untuk beberapa periode tertentu, hasil prediksi harga sedikit lebih tinggi dari harga aktualnya.

Tabel 7 Perbandingan nilai aktual dan prediksi *random forest* (data *testing*)

No	Tanggal	Aktual	Prediksi
1	2024-12-17	6.008,19	6.040,96
2	2024-12-18	5.908,23	5.965,86
3	2024-12-19	5.838,26	5.895,54
4	2024-12-20	5.834,28	5.852,98
⋮	⋮	⋮	⋮
176	2025-09-23	4.418,54	4.480,26
177	2025-09-24	4.400,98	4.476,80
178	2025-09-25	4.406,68	4.452,02
179	2025-09-26	4.404,68	4.454,27

Gambar 6 Grafik perbandingan harga aktual dan prediksi *random forest*

Berdasarkan Gambar 6, grafik menunjukkan bahwa pergerakan harga saham BMRI berdasarkan data *testing* menunjukkan pergeseran yang cukup besar selama beberapa periode. Harga saham turun secara bertahap pada awal periode pengujian, dan hasil prediksi *random forest* mengikuti pola penurunan tersebut, meskipun sedikit di atas nilai aktual. Pada pertengahan periode pengujian, terjadi kenaikan harga saham yang cukup jelas, dan prediksi model mampu mengikuti pola kenaikan tersebut dengan baik. Pada titik tertentu, respons model terhadap lonjakan harga tampak lebih halus daripada data aktual. Secara keseluruhan, model *random forest* mampu menangkap pola pergerakan harga saham dengan cukup baik dan menghasilkan prediksi yang relatif dekat dengan nilai aktual berdasarkan data *testing*.

### 3.7 Postprocessing Data

Tahap *postprocessing* dilakukan dengan mengembalikan hasil prediksi yang telah distandarisasi ke dalam skala harga penutupan saham BMRI yang sebenarnya melalui proses invers standarisasi. Proses ini bertujuan agar hasil prediksi dapat diinterpretasikan secara langsung dan dibandingkan dengan data aktual. Hasil prediksi yang telah dikembalikan ke skala asli selanjutnya digunakan pada tahap evaluasi dan perbandingan kinerja model.

### 3.8 Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan menggunakan data testing yang tidak terlibat pada tahap pelatihan maupun pemilihan parameter. Kinerja masing-masing model kemudian dibandingkan dengan melihat tiga metrik evaluasi, yaitu RMSE, MAPE, dan  $R^2$ . Hasil perbandingan metrik evaluasi pada data testing disajikan pada Tabel 9 berikut.

Tabel 9 Perbandingan kinerja model pada data *testing*

Metrik Evaluasi	SVR	<i>Random Forest</i>
RMSE	39,01	59,95
MAPE	0,56%	0,89%
$R^2$	0,99	0,98

Model SVR menghasilkan nilai RMSE sebesar 39,01, lebih kecil dibandingkan *random forest* sebesar 59,95, hal ini menunjukkan bahwa tingkat kesalahan prediksi pada model SVR lebih rendah. Selain itu, pada metrik evaluasi MAPE, model SVR memiliki persentase yang lebih kecil dibandingkan *random forest*, yaitu sebesar 0,56%. Metrik evaluasi yang terakhir adalah koefisien determinasi ( $R^2$ ), di mana pada metrik SVR, koefisien determinasi sebesar 0,99 dan pada *random forest* sebesar 0,98.

Berdasarkan ketiga metrik evaluasi tersebut, model *Support Vector Regression* (SVR) terbukti unggul dengan tingkat kesalahan yang lebih kecil, dan kemampuan prediksi lebih tinggi. Hal inilah yang menjadikan model SVR sebagai model prediksi yang lebih baik dibandingkan *random forest*.

### 3.9 Hasil Prediksi Satu Hari Ke Depan

Setelah memperoleh model terbaik untuk memprediksi harga saham BMRI, yaitu SVR, maka tahap selanjutnya adalah memprediksi harga saham satu hari ke depan. Model SVR terbaik dibangun dengan parameter  $C$  sebesar 10 dan  $\epsilon$  sebesar  $10^{-2}$ . Harga saham diprediksi menggunakan variabel input yang sama seperti saat pemodelan SVR, yaitu harga penutupan saham BMRI, *moving average* periode 5 dan 10 hari (MA\_5, MA\_10), *exponential moving average* 5 dan 10 hari (EMA\_5, EMA\_10), dan *log return*. Variabel tersebut digunakan untuk merepresentasikan harga saham dan tren pergerakannya, sehingga diasumsikan harga saham satu hari ke depan dipengaruhi oleh perubahan harga pada hari sebelumnya dan pola historis yang ada.

Hasil prediksi harga penutupan saham BMRI satu hari ke depan, tepatnya tanggal 29 September 2025 diperkirakan sebesar Rp4.410,79 per lembar, dengan harga penutupan terakhir pada tanggal 26 September 2025 sebesar Rp4.404,68 per lembar. Hasil prediksi ini mengindikasikan adanya kenaikan harga pada periode berikutnya, sehingga memberikan gambaran awal bagi investor terkait arah pergerakan harga saham BMRI.

## 4 Simpulan dan Saran

Berdasarkan ketiga metrik evaluasi yang digunakan, model *Support Vector Regression* (SVR) menunjukkan nilai *error* yang lebih kecil dibandingkan *random forest* dalam memprediksi harga saham. SVR menghasilkan nilai RMSE sebesar 39,01, lalu MAPE 0,56% dan  $R^2$  sebesar 0,99. Namun demikian, belum dilakukan uji statistik sehingga belum dapat disimpulkan bahwa model SVR secara signifikan lebih unggul

dibandingkan *random forest* dalam memprediksi harga saham. Prediksi harga saham satu hari ke depan, yang diperoleh dengan menggunakan data pada tanggal 26 September 2025 sebagai variabel input, menunjukkan bahwa harga penutupan saham BMRI pada tanggal 29 September 2025 diperkirakan sebesar Rp4.410,79 per lembar. Harga ini sedikit meningkat Rp6,11 per lembar dibandingkan harga penutupan aktual periode satu hari sebelumnya yaitu Rp4.404,68 per lembar.

Beberapa saran untuk penelitian selanjutnya di antaranya menambahkan variabel input lainnya sebagai prediktor, seperti indikator fundamental perusahaan, harga pembukaan, maupun sentimen pasar. Proses *tuning parameter* perlu dilakukan secara lebih luas tanpa dibatasi di awal, sehingga diperoleh parameter optimal yang lebih representatif terhadap karakteristik data. Selain itu, penelitian selanjutnya dapat melakukan uji statistik seperti Diebold-Mariano untuk mengetahui perbedaan akurasi prediksi antar model dengan lebih signifikan, serta menggunakan model *machine learning* lainnya, seperti *Long Short Term Memory* (LSTM), *XGBoost*, atau menggabungkan keduanya menjadi model hibrida sehingga dapat dipertimbangkan sebagai pembandingan untuk model prediksi dengan akurasi yang lebih tinggi.

## 5 Penutup

Peneliti mengucapkan terima kasih kepada tim *editor* dan *reviewer* untuk arahnya selama penelitian ini yang dapat meningkatkan kualitas dari *paper* ini sehingga dapat berguna untuk berbagai kebutuhan akademik.

## Daftar Pustaka

- [1] Kurniawati A, Arima A. 2021. Analisis prediksi harga saham PT. Astra International Tbk menggunakan metode autoregressive integrated moving average (ARIMA) dan support vector regression (SVR). *Jurnal Ilmu Komputasi*. 20(4):417–423. doi: <https://doi.org/10.32409/jikstik.20.3.2732>.
- [2] Fitri E, Riana D. 2022. Analisa perbandingan model prediction dalam prediksi harga saham menggunakan metode linear regression, random forest regression dan multilayer perceptron. *Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*. 6(1):69–78. doi: <https://doi.org/10.46880/jmika.Vol6No1.pp69-78>.
- [3] Henrique BM, Sobreiro VA, Kimura H. 2018. Stock price prediction using support vector regression on daily and up to the minute prices. *Journal of Finance and Data Science*. 4(3):183–201. doi: <https://doi.org/10.1016/j.jfds.2018.04.003>.
- [4] Khan AH, Cao B, Alharbi A, et al. 2023. A performance comparison of machine learning models for stock market prediction with novel investment strategy. *PLoS ONE*. 18(9):e0286362. doi: <https://doi.org/10.1371/journal.pone.0286362>.
- [5] Omar AB, et al. 2022. Stock market forecasting using the random forest and support vector machine models. *Frontiers in Environmental Science*. 10:917047. doi: <https://doi.org/10.3389/fenvs.2022.917047>.
- [6] Jin X. 2022. The comparison of stock price prediction based on linear models and machine learning approaches. In: *Proceedings of the International Conference on Business and Big Data Economy and Management*. Atlantis Press. doi: [https://doi.org/10.2991/978-94-6463-052-7\\_6](https://doi.org/10.2991/978-94-6463-052-7_6).
- [7] Smola AJ, Schölkopf B. 2003. A tutorial on support vector regression. *Statistics and Computing*. 14(3):199–222.
- [8] Schölkopf B, Herbrich R, Smola AJ. 2001. A generalized representer theorem. Dalam: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 2111. Berlin/Heidelberg: Springer. hlm 416–426.
- [9] Ghogh B, Ghodsi A, Karray F, Crowley M. 2021. Reproducing kernel Hilbert space, Mercer's theorem, eigenfunctions, Nyström method, and use of kernels in machine learning: tutorial and survey. *Foundations and Trends in Machine Learning*. 13(1):1–142.

- [10] Breiman L. 2001. Random forests. *Machine Learning*. 45(1):5–32. doi: <https://doi.org/10.1023/A:1010933404324>.
- [11] Speiser JL, Miller ME, Tooze JA, Ip E. 2019. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*. 134:93–101. doi: <https://doi.org/10.1016/j.eswa.2019.05.028>.
- [12] Scornet E, Biau G, Vert JP. 2015. Consistency of random forests. *The Annals of Statistics*. 43(4):1716–1741. doi: <https://doi.org/10.1214/15-AOS1321>.