

Analisis Sentimen Pasar melalui Berita Finansial untuk Prediksi Harga Saham PT Bank Rakyat Indonesia Tbk

Ferdiansyah Permana Putra¹, Mukti Ratna Dewi^{1*}, Fausania Hibatullah¹

¹Departement Statistika Bisnis, Insitut Teknologi Sepuluh Nopember

*Email: mukti_ratna@its.ac.id

Info Artikel	Abstrak
<p>Kata Kunci : analisis sentimen, prediksi harga saham, <i>support vector regression</i>, <i>fruit fly optimization algorit</i>m, IndoBERT</p> <p>Keywords : <i>sentiment analysis</i>, <i>stock prediction</i>, <i>support vector regression</i>, <i>fruit fly optimization algorithm</i>, <i>IndoBERT</i></p> <p>Tanggal Artikel Dikirim : 31 Oktober 2024 Direvisi : 15 November 2024 Diterima : 17 November 2024</p>	<p>Abstrak</p> <p>Sentimen pasar merupakan salah satu faktor yang mempengaruhi fluktuasi harga saham yang dapat bersumber dari masyarakat umum maupun berita-berita yang terkait dengan saham. Dalam penelitian ini, pengaruh sentimen pasar melalui berita keuangan dianalisis terhadap harga saham PT Bank Rakyat Indonesia Tbk (BBRI). Penelitian ini menggunakan pendekatan <i>machine learning</i> dengan metode <i>Support Vector Regression</i> (SVR) untuk memprediksi harga penutupan saham BBRI berdasarkan sentimen berita. Model SVR dioptimalkan dengan algoritma <i>Fruit Fly Optimization Algorithm</i> (FOA). Sentimen pasar terlebih dahulu dievaluasi menggunakan metode IndoBERT yang menunjukkan tingkat akurasi sentimen keseluruhan di atas 90%. Setelah itu, empat skenario pemodelan diusulkan untuk menemukan model prediksi terbaik: (1) model tanpa sentimen, (2) model dengan sentimen pada periode t, (3) model dengan sentimen pada periode $t - 1$, dan (4) model dengan sentimen pada periode t dan periode $t - 1$. Hasil akhir menunjukkan bahwa model pada skenario (1) memiliki kesalahan prediksi terendah dibandingkan dengan model lainnya.</p> <p>Abstract</p> <p><i>Market sentiment is one of the factors that influences the fluctuation of stock prices, which can originate from the general public or news related to stocks. In this study, we explore the effect of market sentiment through financial news on the stock price of PT Bank Rakyat Indonesia Tbk (BBRI). This research adopts a machine learning approach using the Support Vector Regression (SVR) method to predict the closing price of BBRI stock based on news sentiment, and the function is later optimized with the Fruit Fly Optimization Algorithm (FOA) algorithm. The market sentiment is first evaluated using the IndoBERT method, which shows an overall sentiment accuracy level above 90%. Afterward, four modeling scenarios are proposed to find the best prediction model: (1) a model without sentiment, (2) a model with sentiment at period t, (3) a model with sentiment at period $t - 1$, and (4) a model with sentiment at both period t and period $t - 1$. The final results indicate that the model in scenario (1) has the lowest prediction error compared to other models.</i></p>

1. PENDAHULUAN

Investasi saham di pasar modal memberikan keuntungan bagi investor dengan peluang memperoleh laba dari penanaman modal pada perusahaan. Selain itu, investasi juga turut mendorong pertumbuhan ekonom negarai. Sektor keuangan, khususnya perbankan, menjadi pilihan favorit investor muda di pasar modal Indonesia. Berdasarkan data survei PT. Kustodian Sentral Efek Indonesia (KSEI) pada Maret 2022, tercatat sebanyak 690.250 orang berinvestasi di sektor keuangan. Salah satu saham yang mencatat prestasi tinggi adalah PT Bank Rakyat Indonesia Tbk.

PT Bank Rakyat Indonesia Tbk, yang memiliki kode emiten BBRI, adalah salah satu bank terkemuka di Indonesia yang menarik perhatian investor besar seperti Schroders Indonesia dan BNP Paribas Asset Management. Pada tahun 2023, saham BBRI mengalami pencapaian positif dengan mencapai *all-time high* dan pertumbuhan laba kuartal pertama yang signifikan [1]. Meskipun demikian, beberapa kejadian tertentu, seperti penurunan Indeks Harga Saham Gabungan (IHSG) dan krisis di Silicon Valley Bank (SVB), menyebabkan sentimen negatif pada pasar saham yang secara tidak langsung turut menyumbang penurunan harga saham BBRI pada periode tertentu [2]. Sebaliknya, pada periode netral di mana tidak ada berita finansial yang signifikan, harga saham BBRI tetap stabil. Oleh karena itu, penelitian ini mengusulkan penggunaan analisis sentimen pasar untuk memprediksi harga saham PT Bank Rakyat Indonesia Tbk.

SVR merupakan pengembangan lanjutan dari model *Support Vector Machine* (SVM) yang digunakan untuk pemodelan regresi dan *time series*. Kelebihan dari metode ini adalah mampu menyelesaikan permasalahan non-linear dengan memanfaatkan fungsi kernel non-linier, seperti *Radial Basis Function* (RBF), *polynomial*, *spline*, dll [3]. Dari beberapa fungsi kernel non-linier tersebut, fungsi kernel RBF adalah yang paling umum digunakan dan memiliki performansi yang lebih baik dibandingkan dengan fungsi kernel lainnya [4]. Sehubungan dengan kemampuan dalam memprediksi dan meramalkan kejadian mendatang, beberapa penelitian terdahulu menunjukkan bahwa SVR memiliki performa yang lebih baik dibandingkan metode peramalan tradisional seperti *Moving Average* (MA) [5] dan *Autoregressive Integrated Moving Average* (ARIMA) [6] maupun metode pembelajaran mesin lainnya, seperti Jaringan Syaraf Tiruan (JST) [7]. Oleh karena itu, penelitian ini menggunakan SVR dengan fungsi kernel RBF untuk membangun model peramalan harga saham PT Bank Rakyat Indonesia Tbk.

Fungsi kernel RBF memiliki tiga parameter utama, yaitu γ , C , dan ϵ . Parameter γ digunakan untuk memetakan data ke ruang fitur yang lebih tinggi sementara parameter C mengontrol toleransi kesalahan model untuk mencegah *overfitting*. Di sisi lain, parameter ϵ mengontrol batas kesalahan pada fungsi kernel [8]. Meskipun fungsi kernel RBF memiliki kelebihan dalam menangani estimasi non-linear, metode SVR dengan kernel RBF menghadapi tantangan kombinatorial optimasi karena parameter yang kompleks. Solusi untuk mengatasi hal ini adalah menggunakan teknik optimasi cerdas, salah satunya adalah *Fruit Fly Optimization Algorithm* (FOA). Algoritma FOA memiliki keunggulan dalam implementasi yang mudah, perhitungan cepat, konvergensi yang baik, dan kemampuan untuk mencapai solusi optimal dengan cepat. FOA juga memberikan kemudahan dalam pengaturan parameter yang memperkuat kualitas optimasi [8].

Penelitian ini diawali analisis sentimen pasar menggunakan IndoBERT dengan memanfaatkan berita harian finansial. Setelah itu, pemodelan terhadap harga saham BBRI dilakukan dengan empat skenario, yaitu (1) model tanpa sentimen, (2) model dengan sentimen pada periode t , (3) model dengan sentimen pada periode $t - 1$, dan (4) model dengan sentimen pada periode t dan periode $t - 1$. Dari keempat model tersebut kemudian dipilih model dengan kemampuan prediksi terbaik. Harapannya, model terbaik tersebut dapat digunakan oleh investor BBRI dalam pengambilan keputusan investasi di PT Bank Rakyat Indonesia Tbk.

Secara rinci, pertanyaan penelitian yang ingin dijawab dalam penelitian ini adalah sebagai berikut:

1. Bagaimana hasil analisis sentimen data opini berita saham BBRI?
2. Bagaimana pemodelan harga saham BBRI menggunakan *Support Vector Regression* (SVR)?
3. Bagaimana hasil ramalan harga penutupan saham BBRI menggunakan *Support Vector Regression* (SVR) berdasarkan model terbaik?

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen atau *opinion mining* merupakan aplikasi dari *text mining* yang digunakan untuk mengidentifikasi opini dalam sekelompok data teks terkait suatu peristiwa atau topik tertentu [9]. Dalam konteks analisis sentimen, terdapat keterkaitan dengan *text mining* yang merupakan metode untuk mendapatkan data dari teks dokumen dengan tujuan mencari kata-kata yang mewakili isi dokumen tersebut [10]. Analisis sentimen diawali dengan *text preprocessing* [10] dan pengambilan data sentimen melalui proses *scrapping* [11]. Data yang diambil biasanya bersifat tidak terstruktur.

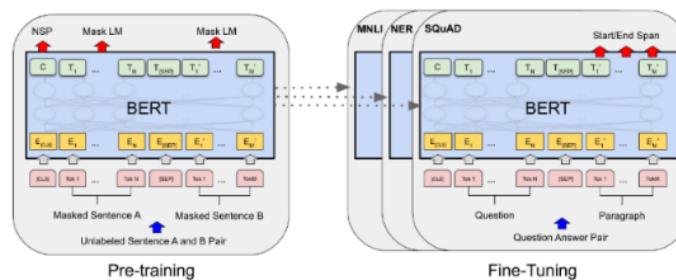
2.2. Bidirectional Encoder Representations from Transformer (BERT)

Salah satu metode analisis sentimen terbaru adalah *Bidirectional Encoder Representations from Transformers* (BERT). BERT yang diperkenalkan oleh Google pada 11 Oktober 2018, merupakan model representasi bahasa terlatih yang menggunakan arsitektur *transformer* secara *bidirectional* untuk memahami konteks kata dalam data teks [12]. Dalam konteks penelitian ini, IndoBERT adalah versi Bahasa Indonesia dari BERT yang menggunakan dataset Indo4B yang mencakup 4 miliar kata dari berbagai sumber. Teknik *SentencePiece* dengan *BPE tokenizer* digunakan dalam IndoBERT untuk membentuk kamus subtoken dari teks-teks dalam dataset [13].

Arsitektur *transformers*, yang menjadi inti dalam BERT, merupakan pendekatan kunci dalam pemrosesan alami, menggunakan *self-attention* untuk memperbaiki pemahaman dan menghasilkan teks pada *encoder* dan *decoder*. *Transformers*, diperkenalkan oleh [14] dan telah menjadi landasan bagi banyak perkembangan dalam pemrosesan bahasa alami (NLP). Dalam arsitektur ini, terdapat *encoder* dan *decoder*. *Encoder* dan *decoder* memproses input teks dan menghasilkan representasi teks yang kaya dan abstrak sehingga lebih kontekstual. *Embedding's* yang digunakan adalah *WordPiece embeddings* yang menjadi langkah awal dalam *input* teks pada BERT.

Setelah melakukan tahapan *embeddings* dan arsitektur *transformers*, tahapan selanjutnya adalah *pre-trained* BERT yang merupakan tahapan ketika model mempelajari bahasa dan konteksnya. *Pre-trained* memiliki serangkaian komponen, yaitu *classification layer* dengan bantuan *fully-connected layer*, sekaligus *Gaussian Error Linear Unit* (GELU) dengan ReLU [14].

Fine tuning BERT adalah proses melatih model dengan tugas yang lebih khusus menggunakan model yang telah diawali pelatihan dengan dataset besar yang disebut *pre-trained*. Proses ini tetap memanfaatkan mekanisme *self-attention* dalam arsitektur *transformers* yang memberikan kemampuan kepada BERT untuk dengan cepat memproses data, baik dalam konteks kalimat tunggal maupun pasangan kalimat. Adapun proses *pre-trained* dan *fine tuning* dapat dijelaskan pada Gambar 1.



Gambar 1. *Pre-trained* dan *fine tuning* BERT (Sumber: [14])

Fine tuning menggunakan *hyperparameter* model untuk meningkatkan performa dalam melatih model. Nilai *hyperparameter* yang optimal untuk setiap tugas pada NLP selalu berbeda, namun model bekerja sangat baik pada *batch size* nya sebanyak 16 dan 32, *learning rate* (Adam) 5×10^{-5} , 5×10^{-5} , dan $w \times 10^{-5}$, sekaligus dengan jumlah iterasi sebanyak 2, 3, dan 4. Selain itu, *fine tuning* dalam BERT menggunakan model *Compute Unified Device Architecture* (CUDA) sebagai pemrosesan *hyperparameter* [14]. Hasil dari analisis IndoBERT ini berupa *softmax probabilities* dengan mekanisme *self-attention* pada arsitektur *transformers* yang dijelaskan pada persamaan (1) dan (2).

$$Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (1)$$

$$\text{softmax}(x)_i = \frac{e^{x_i}}{\sum_{i=1}^n e^{x_i}} \quad (2)$$

di mana \mathbf{Q} adalah matriks *query*, \mathbf{K} sebagai matriks *key*, \mathbf{V} sebagai matriks *value*, nilai d_k sebagai dimensi vektor *query* dan *key*, dan *softmax* sebagai fungsi aktivasi untuk menghasilkan *probabilities*. Selain itu, x_i merupakan setiap *input* dalam vektor x , dengan i adalah indeks dari elemen vektor x , dan n merupakan jumlah elemen dalam vektor x .

Output dari *self-attention* kemudian dimasukkan melalui *Feed Forward Network* (FFN) dengan fungsi aktivasi ReLU yang nantinya menghasilkan skor sentimen untuk pengklasifikasian sentimen positif, netral, dan negatif [14]. Adapun perhitungannya dijelaskan pada persamaan (3) dan (4):

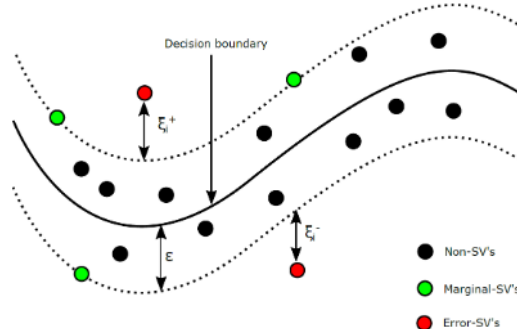
$$FFN(x) = \max(0, x\mathbf{W}_1 + b_1)\mathbf{W}_2 + b_2 \quad (3)$$

$$ReLU(x) = \max(0, x) \begin{cases} x, & \text{jika } x \geq 0 \\ 0, & \text{jika } x < 0 \end{cases} \quad (4)$$

di mana W_1, W_2 merupakan matriks bobot, b_1, b_2 adalah bias, dan x merupakan *input* pada lapisan FFN dan ReLU atau *output* dari *self-attention*.

2.3 Support Vector Regression (SVR)

Support Vector Regression (SVR) adalah suatu model dan pengembangan lanjutan dari Support Vector Machine (SVM) dalam pemodelan, khususnya pada regresi *time series* pada penelitian ini. SVR merupakan metode yang juga dapat menyelesaikan permasalahan estimasi *non-linear* [3]. Konsep SVR didasarkan dari *structural risk minimization* yaitu untuk mengestimasi batas atas dari *generalization error*, sehingga SVR mampu mengatasi *overfitting*. Tujuan dari SVR adalah untuk mendapatkan suatu fungsi dengan tingkat kesalahan paling kecil sehingga menghasilkan suatu prediksi yang bagus [3]. Adapun konsep SVR yang divisualisasikan pada Gambar 2.



Gambar 2. Konsep SVR (Sumber: [8])

Gambar 2 menunjukkan bagaimana *error* pada SVR dihitung. Nilai *error* bernilai sama dengan 0 ketika sampai dengan garis batas *error* ϵ , sedangkan di luar batas tersebut nilai *error* akan dihitung sebagai ξ atau ξ^* . Adapun fungsi regresi dari metode SVR dapat dituliskan pada persamaan (9).

$$f(x) = W^T \varphi(x) + b \quad (5)$$

di mana W adalah matriks pembobot berdimensi ℓ , $\varphi(x)$ merupakan fungsi pemetaan *non-linear* pada ruang dengan ℓ dimensi, dan b adalah konstanta bias, dan $f(x)$ adalah fungsi objektif SVR.

Permasalahan regresi di atas bertujuan untuk meminimalisir risiko dari f . Matriks W dan koefisien b berfungsi untuk meminimalkan fungsi risiko dengan regularisasi yang memiliki formulasi seperti pada persamaan (6).

$$R(f(x)) = \frac{1}{2} \|W\|^2 + C \times \frac{1}{n} \sum_{i=1}^n L_{\epsilon}(y_i, f(x_i)) \quad (6)$$

di mana

$$\sum_{i=1}^n L_{\epsilon}(y_i, f(x_i)) = \begin{cases} 0, & |y_i - f(x_i)| \leq \epsilon \\ |y_i - f(x_i)| - \epsilon, & \text{lainnya} \end{cases} \quad (7)$$

dengan L_{ϵ} adalah ϵ - *insensitive loss function*, y_i adalah nilai sebenarnya pada data ke- i , $i=1, 2, \dots, n$, dan $f(x_i)$ didefinisikan sebagai nilai prediksi model pada data ke- i , $i=1, 2, \dots, n$. Selain itu, C dan ϵ diartikan sebagai *hyperparameter* yang sudah ditentukan, dan n adalah banyaknya data.

Semua titik yang berada dalam rentang $f \pm \epsilon$ disebut *feasible*, sedangkan *infeasible* berada diluar rentang tersebut. Titik-titik yang *infeasible* bisa ditambahkan ke variabel *slack* ξ_i dan ξ_i^* untuk mengatasi permasalahan pembatas yang tidak layak (*infeasible constraint*) sehingga dilakukan optimasi pada persamaan (6) yang dapat ditransformasikan ke dalam bentuk persamaan (8).

$$\min \frac{1}{2} \|W\|^2 + C \times \frac{1}{n} \sum_{i=1}^n (\xi_i + \xi_i^*), \quad (8)$$

dengan batasan

$$W\varphi(x_i) + b - y_i \leq \epsilon + \xi_i^* \text{ dan} \quad (9)$$

$$y_i - W\varphi(x_i) - b \leq \epsilon + \xi_i \text{ dengan } \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, n \quad (10)$$

Optimalisasi pada batasan (9) dan (10) tersebut bisa diselesaikan dengan *primal lagrangian* dalam bentuk persamaan (11) [15].

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

$$-\sum_{i=1}^n \beta_i^* [y_i - \mathbf{W}\varphi(x_i) - b + \varepsilon + \xi_i] - \sum_{i=1}^n (a_i \xi_i + a_i^* \xi_i^*) \quad (11)$$

Persamaan (11) diminimalkan pada variabel *primal* w, b, ξ_i, ξ_i^* dan dimaksimalkan dalam bentuk *lagrangian multiplier non-negative* $a_i, a_i^*, \beta_i, \beta_i^*$ seperti ditampilkan pada persamaan (12)-(15) [15].

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{W} - \sum_{i=1}^n (\beta_i - \beta_i^*) \varphi(x_i) = 0, \quad (12)$$

$$\frac{\partial L}{\partial b} = \mathbf{W} - \sum_{i=1}^n (\beta_i - \beta_i^*) = 0, \quad (13)$$

$$\frac{\partial L}{\partial \xi} = C - \beta_i - a_i = 0, \quad (14)$$

$$\frac{\partial L}{\partial \xi^*} = C - \beta_i^* - a_i^* = 0, \quad (15)$$

sehingga kondisi *Karush-Kuhn-Tucker* terapkan untuk model regresi. Dengan menggunakan persamaan (8) didapatkan *dual lagrangian* pada persamaan (16) dengan mensubstitusikan persamaan (12)-(15). Persamaan *dual lagrangian* sendiri didapatkan ketika fungsi kernel $K(x_i, x_j) = \varphi(x_i)\varphi(x_j)$.

Salah satu fungsi kernel yang paling umum digunakan adalah *Gaussian Radial Basis Function* (RBF) yaitu kernel yang biasa digunakan pada analisis apabila data tidak dapat terpisah secara *linear*. Kernel RBF dapat dijelaskan pada persamaan (16) [16].

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \quad (16)$$

Selain itu juga terdapat kernel *linear* yang merupakan fungsi kernel paling sederhana yang memiliki formulasi pada persamaan (17). Kernel ini digunakan jika data yang akan dianalisis sudah terpisah secara *linear*.

$$K(x_i, x_j) = x_i \cdot x_j \quad (17)$$

Hasil persamaan *dual lagrangian* yang dijelaskan pada persamaan (18).

$$\partial(\beta_i, \beta_i^*) = \sum_{i=1}^n y_i (\beta_i - \beta_i^*) - \varepsilon \sum_{i=1}^n (\beta_i - \beta_i^*) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\beta_i - \beta_i^*) (\beta_j - \beta_j^*) K(x_i, x_j) \quad (18)$$

dengan batasan

$$\sum_{i=1}^n (\beta_i - \beta_i^*) = 0; 0 \leq \beta_i \leq C, i = 1, 2, \dots, n \quad \text{dan} \quad 0 \leq \beta_i^* \leq C, i = 1, 2, \dots, n \quad (19)$$

Lagrange Multipliers pada persamaan (18) memenuhi persamaan $\beta_i \beta_i^* = 0$. *Lagrange Multipliers* β_i dan β_i^* dihitung dan vektor bobot optimal yang diharapkan dari *hyperplane* regresi dijelaskan pada persamaan (20).

$$\mathbf{w} = \sum_{i=1}^n (\beta_i - \beta_i^*) \varphi(x) \quad (20)$$

sehingga didapatkan fungsi regresinya pada persamaan (21).

$$f(x, \beta_i, \beta_i^*) = \sum_{i=1}^n (\beta_i - \beta_i^*) K(x_i, x_j) + b \quad (21)$$

2.4 Fruit Fly Optimization Algorithm (FOA)

SVR memiliki beberapa parameter dengan kernel yang digunakan, seperti *gamma*, *C*, dan *epsilon*. Kombinasi parameter tersebut sulit dilakukan, sehingga perlu fungsi optimasi untuk menentukan parameter yang terbaik pada SVR, salah satunya FOA. Metode FOA merupakan suatu pendekatan untuk mencapai optimasi secara umum dengan mengadaptasi perilaku pencarian makanan yang dilakukan oleh lalat buah. Lalat buah memiliki organ *osphresis* yang memungkinkannya untuk mendeteksi berbagai jenis aroma makanan di udara, bahkan dari jarak yang jauh hingga mencapai 40 km [17]. Adapun langkah-langkah pada proses optimasi untuk meminimalkan adalah sebagai berikut.

1. Menentukan posisi awal dari *fruit fly population*, iterasi gen, dan koordinat posisi X_0 dan Y_0 dari populasi lalat buah acak.

2. Perpindahan lalat dengan berpatokan jarak terdekat dapat dihitung menggunakan rumus yang ditunjukkan pada persamaan (22) dan (23).

$$x_i = X_0 + \text{Random Value}, \quad (22)$$

$$y_i = Y_0 + \text{Random Value}, \quad (23)$$

di mana $i \in \{1, \dots, n - 1, n\}$

3. Menentukan jarak $Dist_i$ antara individu terbang dan asal lalat diperkirakan terlebih dahulu dan nilai penentuan konsentrasi bau S_i dirumuskan pada persamaan (24) dan (25).

$$Dist_i = \sqrt{X_i^2 + Y_i^2} \quad (24)$$

$$S_i = \frac{1}{Dist_i} \quad (25)$$

4. Nilai konsentrasi bau S_i disubstitusikan dalam fungsi *Fitness* untuk menentukan konsentrasi bau $Smell_i$ dari posisi individu *fruit fly* yang ditunjukkan pada persamaan (26).

$$Smell_i = \text{Fitness}(S_i) \quad (26)$$

5. Individu dengan nilai konsentrasi bau terbaik dipilih di antara kumpulan lalat buah, termasuk konsentrasi dan koordinatnya. Nilai $Smell_i$ merupakan nilai yang menentukan seberapa baik konsentrasi bau pada setiap individu *fruit fly*. Sehingga dapat ditunjukkan pada persamaan (27).

$$[best\ smell, best\ index] = \max(Smell_i) \quad (27)$$

6. Ketika nilai $bestSmell$ lebih kecil dari nilai $Smell_{best}$ saat ini, maka ganti nilai $Smell_{best}$ dengan $bestSmell$. Set lokasi baru dari kumpulan *fruit fly* ke koordinat dari individu *fruit fly* dengan nilai $bestSmell$ yang ditunjukkan pada persamaan (27), (28), dan (29).

$$Smell_{best} = bestSmell \quad (28)$$

$$X_0 = X(bestIndex) \quad (29)$$

$$Y_0 = Y(bestIndex) \quad (30)$$

7. Mengulangi langkah 2-5 untuk menentukan apakah konsentrasi bau maksimum lebih tinggi dari generasi sebelumnya, lalu melanjutkan langkah ke 6. Proses iterasi berhenti apabila iterasi telah mencapai nilai maksimal.

2.5 Akurasi Model Peramalan

Akurasi model dari SVR yang telah dilakukan merupakan sangat penting untuk menilai model manakah yang mampu meramalkan paling baik prediksi nilai saham BBRI di Indonesia. Evaluasi performansi model dapat diukur menggunakan MAPE dan RMSE.

2.5.1 Mean Absolute Percentage Error (MAPE)

MAPE merupakan rata-rata persentase absolut akibat kesalahan peramalan. MAPE menunjukkan seberapa besar kesalahan peramalan dengan nilai aktual. Jika nilai MAPE dari suatu peramalan kecil, maka peramalan tersebut akan semakin akurat. Berikut persamaan MAPE (30) [18].

$$MAPE = \frac{\sum_t^n |y_t - \hat{y}_t|}{n} \times 100\% \quad (31)$$

di mana y_t adalah nilai sebenarnya pada waktu ke- t dengan $t=1, 2, \dots, n$. Sementara itu, \hat{y}_t diartikan sebagai nilai dugaan pada waktu ke- t , $t = 1, 2, \dots, n$. Dalam hal ini, n adalah jumlah observasi dari data *time series*.

2.5.2 Root Mean Squared Error (RMSE)

Prediksi yang telah dibuat oleh model sebelumnya akan dipastikan terlebih dahulu agar dapat digunakan untuk analisis lebih lanjut atau pengambilan keputusan. Evaluasi model yang dilakukan untuk membandingkan nilai aktual dengan nilai prediksi. Jika nilai RMSE semakin kecil, maka semakin akurat kinerja model tersebut [19]. Berikut persamaan RMSE (31).

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2} \quad (32)$$

dengan y_t adalah nilai sebenarnya pada waktu ke- t di mana $t=1, 2, \dots, n$. Sementara itu, \hat{y}_t diartikan sebagai nilai dugaan pada waktu ke- t , $t=1, 2, \dots, n$. Dalam hal ini, n merupakan jumlah observasi dari data *time series*.

2.6 Ketepatan Klasifikasi Hasil Analisis Sentimen

Performa klasifikasi dapat diukur dengan ketepatan klasifikasi. Pengukurannya sendiri diketahui dari jumlah sentimen berita saham BBRI yang bersentimen positif, netral, dan negatif. *Confussion matrix* yang memuat nilai tersebut disajikan pada Tabel 1.

Tabel 1. Confussion Matrix

Kelas Aktual	Kelas Prediksi		
	Positif	Netral	Negatif
Positif	<i>True Positive</i>	<i>False Neutral</i>	<i>False Negative</i>
Netral	<i>False Positive</i>	<i>True Neutral</i>	<i>False Negative</i>
Negatif	<i>False Positive</i>	<i>False Neutral</i>	<i>True Negative</i>

True Positive (TP) adalah kondisi ketika kelas positif dapat diprediksi dengan benar. Begitu pula dengan *True Neutral* (TNt) maupun *True Negative* (TN). Di sisi lain, *False Positive* (FP) adalah jumlah dari observasi yang sebenarnya negatif atau netral, namun diklasifikasikan sebagai kelas positif. *False Neutral* (FNt) adalah jumlah kelas negatif atau netral yang diprediksi netral. Terakhir, *False Negative* (FN) adalah kondisi ketika kelas positif atau netral diprediksi sebagai kelas negatif.

$$\text{Akurasi keseluruhan} = \frac{TP + TNt + TN}{TP + FP + TNt + FNt + TN + FN} \quad (33)$$

$$\text{Akurasi kelas positif} = \frac{TP}{TP + FP} \quad (34)$$

$$\text{Akurasi kelas netral} = \frac{TNt}{TNt + FNt} \quad (35)$$

$$\text{Akurasi kelas negatif} = \frac{TN}{TN + FN} \quad (36)$$

Salah satu metrik evaluasi yang digunakan dalam penelitian ini adalah akurasi. Akurasi digunakan untuk menghitung ketepatan klasifikasi sebuah dokumen yang mempunyai data yang seimbang pada setiap kategorinya [20]. Adapun persamaan untuk menghitung akurasi yang dapat dijelaskan pada persamaan (32), (33), (34), dan (35).

3. METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang akan digunakan dalam penelitian ini merupakan data sekunder yang didapatkan melalui situs CNBC Indonesia pada <https://www.cnbcindonesia.com/> mengenai data berita saham BBRI dan situs <https://finance.yahoo.com/> mengenai data harga penutupan saham BBRI periode 02 Januari 2020 hingga 18 Juli 2023 dengan total data bersih yaitu 861 data.

3.2 Variabel Penelitian

Skenario pemodelan dan variabel yang digunakan dalam penelitian ini masing-masing disajikan pada Tabel 2 dan Tabel 3.

Tabel 2. Model Penelitian

Tipe	Model	Keterangan
Dengan Lag	Model A (Tanpa Sentimen)	Harga saham BBRI diprediksi berdasarkan harga penutupan saham BBRI pada hari sebelumnya (lag).
	Model B (Model Dengan Sentimen Berita Periode ke t)	Harga saham BBRI diprediksi berdasarkan harga penutupan saham BBRI pada hari sebelumnya (lag) dengan sentimen periode ke t .
	Model C (Model Dengan Sentimen Berita Periode ke $t - 1$)	Harga saham BBRI diprediksi berdasarkan harga penutupan saham BBRI pada hari sebelumnya (lag) dengan sentimen periode ke $t - 1$.
	Model D (Model Dengan Sentimen Berita Periode ke t dan Periode ke $t - 1$)	Harga saham BBRI diprediksi berdasarkan harga penutupan saham BBRI pada hari sebelumnya (lag) dengan sentimen periode ke t dan periode ke $t - 1$.
Tanpa Lag	Model E (Tanpa Sentimen)	Memprediksi harga saham BBRI tanpa menggunakan lag atau harga penutupan saham BBRI pada hari sebelumnya.
	Model F (Model Dengan Sentimen Berita Periode ke t)	Memprediksi harga saham BBRI tanpa menggunakan lag atau harga penutupan saham BBRI pada hari sebelumnya dengan sentimen periode ke t .
	Model G (Model Dengan Sentimen Berita Periode ke $t - 1$)	Memprediksi harga saham BBRI tanpa menggunakan lag atau harga penutupan saham BBRI pada hari sebelumnya dengan sentimen periode ke $t - 1$.
	Model H (Model Dengan Sentimen Berita Periode ke t dan Periode ke $t - 1$)	Memprediksi harga saham BBRI tanpa menggunakan lag atau harga penutupan saham BBRI pada hari sebelumnya dengan sentimen periode ke t dan periode ke $t - 1$.

Adapun kategori sentimen berita dalam penelitian ini dapat dinyatakan sebagai berikut.

- Berita periode ke t ini mengindikasikan bahwa sentimen berita diambil paling akhir sebelum penutupan harga saham BBRI pukul 16.00 WIB.
- Berita periode ke $t - 1$ ini mengindikasikan bahwa sentimen berita diambil paling akhir setelah penutupan harga saham BBRI pukul 16.00 WIB.

Sentimen diklasifikasikan netral jika opini berita saham tidak relevan dengan saham tertentu dan tidak menunjukkan kecenderungan sentimen positif atau negatif atau mencerminkan sentimen pasar saham yang netral. Selain itu, opini berita saham yang tidak terdapat berita pada waktu hari tertentu juga diklasifikasikan sebagai sentimen netral. Pada analisis *Support Vector Regression* dalam memprediksi harga saham BBRI dengan sentimen menggunakan hasil prediksi IndoBERT pada sentimen positif, netral, dan negatif sebagai variabel prediktornya.

Tabel 3. Variabel Penelitian

Variabel	Keterangan	Skala	Kategori
D_{1t}^*	Klasifikasi Sentimen Berita Saham BBRI Positif Periode Ke t	Nominal	D_1 bernilai 1 untuk klasifikasi positif, sedangkan klasifikasi negatif dan netral bernilai 0
D_{2t}^*	Klasifikasi Sentimen Berita Saham BBRI Negatif Periode Ke t	Nominal	D_2 bernilai 1 untuk klasifikasi negatif, sedangkan klasifikasi positif dan netral bernilai 0
D_{3t}^{**}	Klasifikasi Sentimen Berita Saham BBRI Positif Periode Ke $t - 1$	Nominal	D_3 bernilai 1 untuk klasifikasi positif, sedangkan klasifikasi negatif dan netral bernilai 0
D_{4t}^{**}	Klasifikasi Sentimen Berita Saham BBRI Negatif Periode Ke $t - 1$	Nominal	D_4 bernilai 1 untuk klasifikasi negatif, sedangkan klasifikasi positif dan netral bernilai 0
Y_t	Harga penutupan Saham BBRI Harian ke- t	Rupiah	-
Y_{t-1}	Harga penutupan Saham BBRI Harian Satu Hari Sebelumnya	Rupiah	-

*Catatan:

* digunakan model B dan F (dengan sentimen periode ke t) dan model D dan H (dengan sentimen periode ke t dan periode ke $t - 1$)

** digunakan model C dan G (dengan sentimen periode ke $t - 1$) dan model D dan H (dengan sentimen periode ke t dan periode ke $t - 1$)

Langkah analisis yang digunakan dalam menyusun penelitian ini adalah sebagai berikut.

1. Melakukan *crawling* data opini berita saham BBRI di *website* CNBC Indonesia menggunakan *scrapping package* pada *python* dengan fungsi *HyperText Markup Language* (HTML).
2. Menyimpan file opini berita saham BBRI hasil *data crawling*.
3. Melakukan *crawling* dan *cleaning* data sentimen berita saham BBRI menjadi sentimen periode ke t dan periode ke $t - 1$.
4. Melakukan tahapan *text pre-processing* data sentimen berita saham BBRI untuk sentimen periode ke t dan periode ke $t - 1$.
5. Melakukan karakteristik data pada sentimen berita saham BBRI untuk sentimen periode ke t dan periode ke $t - 1$ menggunakan *wordcloud*.
6. Melakukan tahapan analisis sentimen IndoBERT pada data sentimen berita saham BBRI periode ke t dan periode ke $t - 1$.
7. Melakukan analisis karakteristik pada harga penutupan saham BBRI dengan *time series plot*.
8. Melakukan identifikasi lag signifikan menggunakan plot ACF dan PACF.
9. Melakukan analisis peramalan harga penutupan saham BBRI menggunakan *Support Vector Regression* (SVR) dengan dan tanpa melibatkan hasil analisis sentimen sebagai variabel prediktor.
 - a) Membagi data *training* sebesar 80% (34 bulan) dan data *testing* sebesar 20% (9 bulan).
 - b) Menentukan dan membandingkan antara fungsi kernel yang akan digunakan yaitu menggunakan fungsi kernel *Radial Basis Function* (RBF) dan *linear*.
 - c) Menentukan nilai parameter kernel *linear* paling optimum yang digunakan menggunakan metode *Fruit Fly Optimization Algorithm* yaitu parameter C , ε , dan γ pada parameter kernel RBF, sedangkan parameter C dan ε untuk kernel *linear*.
10. Menentukan model SVR dengan parameter optimum FOA terbaik.
 - a) Pemilihan fungsi kernel antara *linear* dengan RBF berdasarkan *error* terkecil.
 - b) Fungsi kernel terbaik pada poin 10a digunakan untuk memilih skenario model terbaik berdasarkan *error* terkecil sekaligus melakukan evaluasi hasil model SVR menggunakan *line chart plot*.
11. Meramalkan harga penutupan saham BBRI berdasarkan model terbaik.
12. Menginterpretasikan hasil analisis.
13. Menarik kesimpulan dan saran.

4. HASIL DAN PEMBAHASAN

4.1 Analisis Sentimen Pasar Menggunakan IndoBERT

Tahapan analisis sentimen menggunakan IndoBERT meliputi *text preprocessing*, *pre-trained*, dan *fine tuning* yang dijelaskan secara rinci sebagai berikut.

1. *Text Preprocessing*

Tahapan *text preprocessing*, meliputi pelabelan menggunakan manual, kemudian melakukan *casefolding* (mengubah menjadi huruf kecil), *cleaning* (terkait hapus *url*, *link*, simbol, angka, dan lainnya), tokenisasi (terkait memecah kalimat menjadi sub kata), dan normalisasi. Selain itu, *wordcloud* ditampilkan untuk melihat kata-kata mana yang paling sering dibahas. Hasil dari *wordcloud* menunjukkan bahwa "IHSG" adalah kata yang paling sering di bahas pada berita CNBC Indonesia. Artinya, berdasarkan beberapa berita yang telah dianalisis, BBRI berkontribusi dalam fluktuasi nilai IHSG.

2. *Pre-trained* IndoBERT

Pada *pre-trained* ini menggunakan model *IndoBERT-base-p1* dengan total parameter 124.443.651 parameter yang harus dilatih. *Sequence* yang digunakan sebanyak 512 dan *batch size* sebanyak 32. Adapun *splitting* data pada *pre-trained* IndoBERT dapat dijelaskan pada Tabel 4.

Tabel 4. Splitting Data Berita Saham BBRI Periode Ke t dan Periode Ke $t - 1$

Split Data	1	2	3	Jumlah Sampel		
				1	2	3
Training	60%	70%	80%	516	602	688
Validasi	20%	15%	10%	172	173	86
Testing	20%	15%	10%	173	86	87

Berdasarkan *splitting* tersebut, uji coba *pre-trained* IndoBERT dilakukan untuk periode ke t dan periode ke $t - 1$. Tabel 4 menunjukkan bahwa masih ada teks yang tidak konsisten (tidak akurat) dalam analisis IndoBERT. Hal ini ditunjukkan nilai *softmax probabilities* yang $< 80\%$ sehingga perlu dilakukan *fine tuning* model IndoBERT untuk meningkatkan performa model agar mampu mengklasifikasikan sentimen dengan lebih baik.

3. *Fine tuning* IndoBERT

Tahapan *fine tuning* IndoBERT menggunakan *hyperparameter*, seperti *n-epoch* sebesar 4, *batch size* sebanyak 32, dan *learning rate* menggunakan Adam dengan nilai 3×10^{-5} serta model optimasi menggunakan *cuda* berdasarkan penelitian [12]. Hasil selama *fine tuning* IndoBERT menunjukkan nilai akurasi untuk data *training* secara keseluruhan baik untuk berita periode ke t maupun periode ke $t - 1$. Artinya, model IndoBERT telah dilatih dengan baik atau efektif. Selain itu, hal ini diperkuat dengan uji coba prediksi teks sentimennya didapatkan nilai *softmax probabilities* secara keseluruhan di atas 90%.

4. Klasifikasi Prediksi Sentimen IndoBERT

- *Berita Periode ke-t*

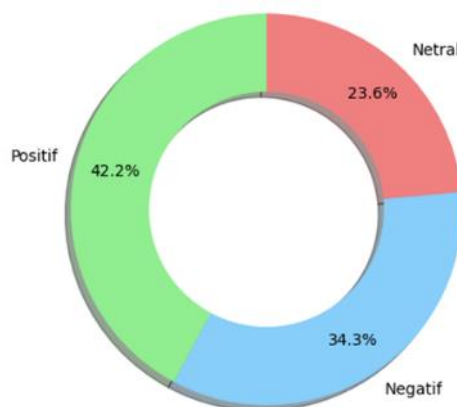
Nilai akurasi untuk data *testing* berita sentimen periode ke t disajikan pada Tabel 5. Terlihat bahwa secara umum model prediksi lebih unggul dalam memprediksi sentimen positif dibandingkan dengan sentimen lainnya. Pada Tabel 5 juga nampak bahwa split data 3 memiliki kemampuan memprediksi sentimen berita saham BBRI paling baik di antara lainnya. Gambar 3 yang menunjukkan bahwa klasifikasi sentimen cenderung didominasi oleh sentimen positif.

- *Berita Periode ke $t - 1$*

Tabel 6 menunjukkan nilai akurasi untuk data *testing* berita sentimen periode ke $t - 1$. Terlihat bahwa dari model akurasi lebih unggul dalam mengklasifikasikan sentimen negatif. Sama halnya dengan periode ke- t , split data 3 memiliki kemampuan memprediksi sentimen berita saham BBRI terbaik di antara split yang lain. Gambar 4 yang menunjukkan bahwa klasifikasi sentimen cenderung didominasi oleh sentimen positif.

Tabel 5. Nilai Akurasi Sentimen IndoBERT Periode Ke t

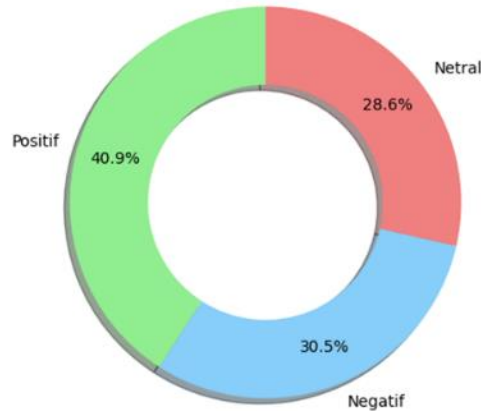
Keterangan	Split 1	Split 2	Split 3
Akurasi Positif	0,77	0,92	0,97
Akurasi Netral	0,72	0,83	0,94
Akurasi Negatif	0,89	0,81	0,90
Akurasi Keseluruhan	0,79	0,85	0,93



Gambar 3. Proporsi Prediksi Model IndoBERT Periode Ke t

Tabel 6. Nilai Akurasi Sentimen IndoBERT Periode Ke $t - 1$

Keterangan	Split 1	Split 2	Split 3
Akurasi Positif	0,74	0,88	0,97
Akurasi Netral	0,62	0,75	0,76
Akurasi Negatif	0,90	0,97	0,97
Akurasi Keseluruhan	0,72	0,87	0,91



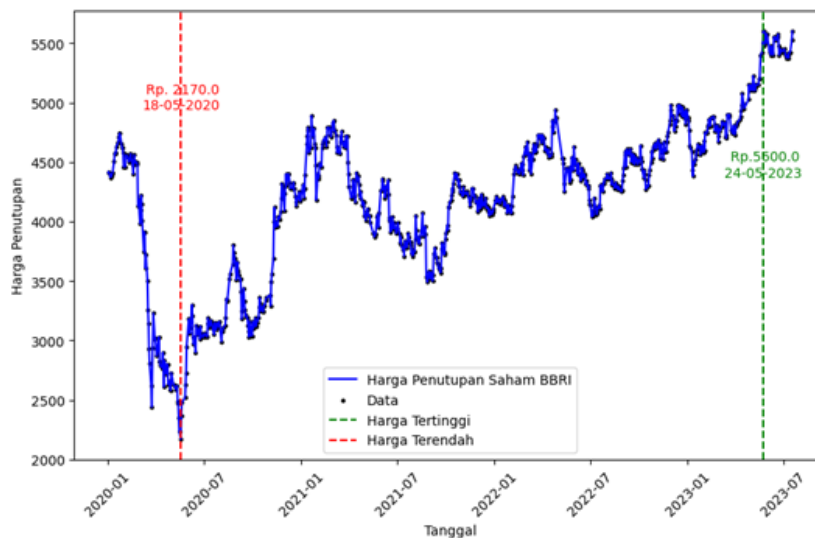
Gambar 4. Proporsi Prediksi Model IndoBERT Periode Ke $t - 1$

4.2 Analisis Harga Saham Menggunakan SVM-FOA

Setelah melakukan analisis sentimen maka langkah selanjutnya adalah analisis harga saham yang meliputi identifikasi *time series plot* dan pembentukan model prediksi harga penutupan saham BBRI menggunakan SVR-FOA.

1. *Time Series Plot* Harga Penutupan Saham BBRI

Time series plot dari harga penutupan saham BBRI mulai tanggal 02 Januari 2023 hingga 18 Juli 2023 yang dapat divisualisasikan pada Gambar 5 yang menunjukkan bahwa *plot* harga penutupan saham BBRI cenderung pola *trend* naik dan adanya fluktuansi.



Gambar 5. *Time Series Plot* Harga Penutupan Saham BBRI

2. Pemodelan Prediksi Harga Penutupan Saham BBRI

Pembentukan model diawali dengan pembagian data menjadi *training* dan *testing* dengan proporsi 80:20. Data *training* mencakup periode 02 Januari 2020 hingga 26 Oktober 2022, sedangkan *testing* adalah harga penutupan selama 27 Oktober 2022 hingga 18 Juli 2023. Selanjutnya, konfigurasi parameter untuk SVR dilakukan menggunakan FOA dengan ketentuan *smell* sebesar 100. *Pop* akan diinisiasi sebesar 20, 30, 40, 50, 60, 70, 80, 90, dan 100 sementara *gen* akan diinisiasi sebesar 100. Hasil dari pemodelan secara rinci disajikan pada Tabel 7.

Tabel 7. Hasil Konfigurasi Parameter FOA Terbaik Berdasarkan Kernel

Kernel	Model	Pop	Parameter			MAPE	
			C	ϵ	γ	Training	Testing
RBF	A	90	1001,512780	0,239994	18,034157	1,629208	11,272677
	B	30	4840,614115	1,109887	170,908936	1,232365	12,812367
	C	100	4089,182979	1,230957	247,117463	1,197299	12,688052
	D	30	135,7946169	0,047345	7,06320946	2,787245	12,63627
Linear	A	40	1001,362205	0,193271	19,887570	1,682845	1,032080
	B	100	4089,182979	1,230957	247,117463	1,679485	1,047047
	C	30	4840,614115	1,109887	170,908936	1,683137	1,034018
	D	20	136,1747012	0,042432	7,69513173	1,677970	1,049993

*Catatan: A (Model Tanpa Sentimen), B (Model Sentimen Periode Ke t), C (Model Sentimen Periode Ke $t - 1$), dan D (Model Sentimen Periode Ke t dan Periode Ke $t - 1$)

Tabel 7 menunjukkan bahwa hasil konfigurasi parameter FOA terbaik adalah SVM dengan kernel *linear* yang memiliki nilai MAPE *testing* terendah. Model ini kemudian digunakan untuk membandingkan performa prediksi dengan atau tanpa lag. Hasil perbandingan tersebut disajikan pada Tabel 8 yang menunjukkan bahwa model dengan lag menghasilkan nilai *error* peramalan yang terendah dibandingkan tanpa lag. Secara khusus, model tanpa sentimen dengan lag memiliki kemampuan prediksi paling unggul karena memiliki nilai MAPE paling kecil. Meskipun demikian, terlihat pada Tabel 8 bahwa sebenarnya kebaikan prediksi antara model dengan dan tanpa sentimen hampir sama. Hal ini secara tidak langsung senada dengan teori *Efficient Market Hypothesis* (EMH) yang menyatakan bahwa harga saham merefleksikan informasi yang dipublikasikan terkait saham tersebut [21]. Mengingat model dengan dan tanpa sentimen memiliki performa prediksi yang hampir sampai, model D dengan lag akhirnya dipilih sebagai model terbaik dengan alasan model lebih sederhana.

Tabel 8. Evaluasi Hasil Model SVR Dengan dan Tanpa Lag

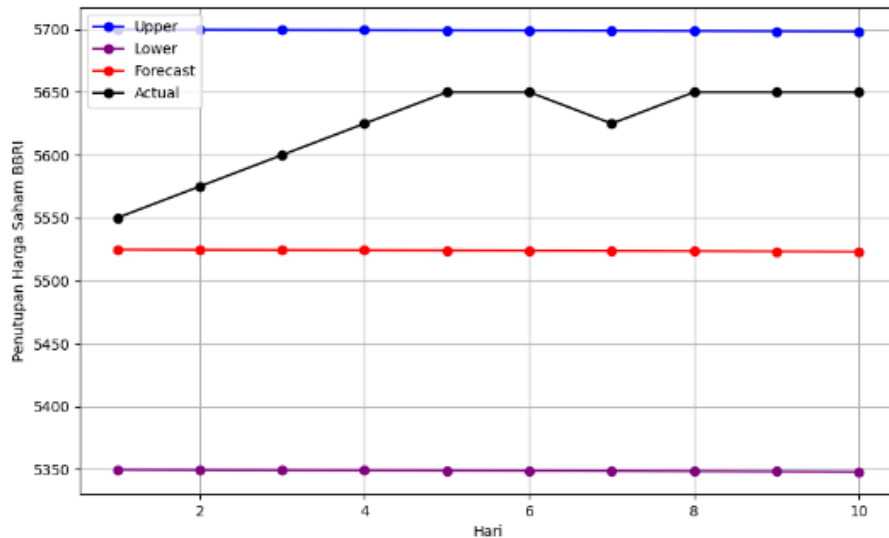
Kriteria	Model	MAPE (%)	RMSE	MAPE (%)	RMSE
		Training	Training	Testing	Testing
Dengan Lag	A	1,682845	89,369737	1,032080	65,104698
	B	1,679485	89,421102	1,047047	65,282743
	C	1,683137	89,397156	1,034018	65,102004
	D	1,677970	89,498950	1,049993	65,511578
Tanpa Lag	C	10,203392	501,483535	5,618119	378,752790
	D	10,145359	502,493346	5,488326	370,494929
	E	10,155998	500,435675	5,621596	378,656938
	F	10,066816	503,071755	5,555870	372,342178

*Catatan: A (Model Tanpa Sentimen), B (Model Sentimen Periode Ke t), C (Model Sentimen Periode Ke $t - 1$), dan D (Model Sentimen Periode Ke t dan Periode Ke $t - 1$), E (Model Tanpa Sentimen dan Tanpa Lag), F (Model Sentimen Periode Ke t dan Tanpa Lag), G (Model Sentimen Periode Ke $t - 1$ dan Tanpa Lag), dan H (Model Sentimen Periode Ke t dan Periode Ke $t - 1$, serta Tanpa Lag)

4.3 Peramalan Harga Penutupan Saham BBRI

Peramalan harga penutupan Saham BBRI dilakukan selama 10 hari kedepan menggunakan model terbaik ditunjukkan pada Gambar 6. Meskipun pola tren harga prediksi saham BBRI tidak sama dengan harga asli, namun hasil peramalan tersebut masih dianggap sama dengan harga asli karena masuk dalam selang kepercayaan

95%. Dalam kata lain, model yang terbentuk masih dapat digunakan untuk meramalkan harga penutupan saham BBRI.



Gambar 6. Plot Perbandingan Harga Penutupan Saham BBRI

5. KESIMPULAN

Kesimpulan dari hasil analisis yang telah dilakukan dalam penelitian ini adalah sebagai berikut.

1. Model prediksi sentimen berita saham BBRI menggunakan analisis IndoBERT memiliki akurasi secara keseluruhan di atas 90%.
2. Model peramalan harga penutupan saham BBRI terbaik adalah model tanpa sentimen dengan fungsi kernel *linear*.
3. Hasil peramalan harga penutupan saham BBRI dapat dikatakan sesuai dengan harga riil karena berada dalam selang kepercayaan 95%.

Model prediksi dari penelitian ini dapat digunakan sebagai referensi bagi para investor dalam keputusan transaksi jual beli di pasar saham, khususnya untuk emiten BBRI. Penelitian selanjutnya dapat mempertimbangkan penggunaan metode lainnya yang lebih *advance* untuk analisis sentimen maupun untuk membangun model untuk mengoptimalkan hasil prediksi.

DAFTAR PUSTAKA

- [1] I. CNN, "Cetak Laba Kuartal I, Saham BBRI Terbang," 3 Rabu 2023. [Online]. Available: <https://www.cnnindonesia.com/teknologi/20230501124025-303-943910/cetak-laba-rp1556-triliun-di-kuartal-i-2023-saham-bbri-terbang>.
- [2] M. R. I. Taufani, "SVB Krisis Teknologi atau Krisis Perbankan? Ini Penjelasan," 23 April 2023. [Online]. Available: <https://www.cnbcindonesia.com/research/20230423200317-128-431925/svb-krisis-teknologi-atau-krisis-perbankan-ini-penjelasan>.
- [3] C.-J. Lu, "Financial Time Series Forecasting Using Independent Component Analysis and Support Vector Regression," *Decision Support Systems*, pp. 115-125, 2009.
- [4] M. F. Rohmah, I. K. G. Putra, R. S. Hartati and A. L., "Comparison Four Kernels of SVR to Predict Consumer Price Index," *Journal of Physics: Conference Series*, pp. 2-10, 2021.
- [5] E. S. Hediarti, *Peramalan Harga Saham Dengan Menggunakan Metode Support Vector Regression*, Surabaya: ITS Press, 2019.

- [6] D. I. Purnama and S. Setianingsih, "Support Vector Regression (SVR) Model for Forecasting Number of Passengers on Domestic Flights at Sultan Hasanudin Airport Makassar," *Jurnal Matematika, Statistika, dan Komputasi*, pp. 391-403, 2020.
- [7] Mustakim, A. Buono and I. Hermadi, "Performance Comparison Between Support Vector Regression And Artificial Neural Network For Prediction of Oil Palm Production," *Jurnal Ilmu Komputer dan Informasi*, pp. 1-8, 2016.
- [8] L. W. G. Cao, "Support Vector Regression with Fruit Fly Optimization Algorithm for Seasonal Electricity Consumption Forecasting," *Energy*, pp. 734-745, 2016.
- [9] Z. Li, Y. Fan, B. Jiang, T. Lei and W. Liu, "A Survey On Sentiment Analysis and Opinion Mining for Social Multimedia," *Multimedia Tools and Applications*, pp. 6939-6967, 2019.
- [10] Mihalcea and Ignatow, *A Guidebook for the Social Sciences*, UK: SAGE Publications, 2017.
- [11] M. Turland, *Php "Architect's Guide to Web Scraping"*, Canada: Marco Tabini & Associates, 2016.
- [12] J. Devlin, M. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *2019 Conference of The North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, North American, 2019.
- [13] B. Willie, K. Vincentio, G. I. Winata, S. Cahyawijaya, X. Li, Z. Y. Lim, S. Soleman, M. R. P. Fung, S. Bahar and A. Purwati, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference Natural Language Processing*, pp. 843-857, 2020.
- [14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, "Attention is All You Need : Advances in Neural Information Processing Systems," *Nips*, pp. 5999-6009, 2017.
- [15] W. C. Hong, "Electric Load Forecasting by Support Vector Model," *Applied Mathematical Modelling Vol. 33*, pp. 2444-2454, 2008.
- [16] W. K. Haerdle and C. M. D. D & Hafner, "Support Vector Machines with Evolutionary Model Selection for Default Prediction, in *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*," eds. Racine, JS, Su, L, and Ullah, A, Oxford, pp. 346-373, 2014.
- [17] W. T. Pan, "A New Fruit Fly Optimization Algorithm: Taking The Financial Distress Model As An Example," *Knowledge-Based System, No. 26*, pp. 68-74, 2012.
- [18] W. W. Wei, *Time Series Analysis Univariate and Multivariate Methods 2nd Edition*, USA: Pearson Education, 2006.
- [19] Z. Chen, "Gaussian Process Regression Methods and Extensions for Stock Market Prediction (Doctoral Dissertation)," University of Leicester, UK, 2017.
- [20] A. Hotho, A. Numberger and G. Paab, "A Brief Survey of Text Mining," *In Ldv Forum*, pp. Vol. 20(1), p. 19-62, 2005.
- [21] G. B. Malkiel, "Efficient Market Hypothesis," in *Finance*, London, Palgrave Macmillan UK, 1989, pp. 127-134.
- [22] D. Rustiana and S. Ramadhani, "Strategi di Pasar Modal Syariah," *Jurnal Ilmu Komputer, Ekonomi, dan Manajemen (JIKEM)*, pp. 1578-1589, 2022.
- [23] T. Purwanti, "Ini Saham Sektoral yang Jadi Favorit Investasi Anak Muda," 12 April 2022. [Online]. Available: <https://www.cnbcindonesia.com/market/20220414133554-17-331731/ini-saham-sektoral-yang-jadi-favorit-investasi-anak-muda>.
- [24] A. Malik, "The Big 4 Bank Diprediksi Cemerlang, Puluhan Reksadana Borong Saham BMRI, BBRI, BBNI, dan BBKA," 24 Februari 2023. [Online]. Available: <https://www.bareksa.com/berita/saham/2023-02-24/the-big-4-bank-diprediksi-cemerlang-puluhan-reksadana-borong-saham-bmri-bbri-bbni-dan-bbca>.