

Perbandingan Kinerja Metode Regresi Linear dan *Extreme Gradient Boosting* (XGBoost) dalam Prediksi Harga Saham BUMN Sektor Perbankan

Annisa AF Suharto^{1*}, Mauliddin²

^{1,2} Universitas Hasanuddin, Indonesia

*Korespondensi: Email: annisaafs3003@gmail.com; mauliddin@unhas.ac.id

Abstrak

Pasar saham, khususnya pada sektor perbankan Badan Usaha Milik Negara (BUMN), memiliki peran penting dalam mencerminkan dinamika ekonomi Indonesia. Tingginya volatilitas harga saham mendorong kebutuhan akan metode prediksi yang akurat untuk mendukung pengambilan keputusan investasi. Penelitian ini bertujuan untuk membandingkan kinerja metode Regresi Linear dan *Extreme Gradient Boosting* (XGBoost) dalam memprediksi harga saham BUMN sektor perbankan. Penelitian dilakukan dengan pendekatan kuantitatif deskriptif-komparatif menggunakan data harga saham harian BBNI dan BBTN selama periode Oktober 2023 hingga Oktober 2024. Model dievaluasi menggunakan metrik *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE). Hasil penelitian menunjukkan bahwa XGBoost memberikan hasil prediksi yang lebih akurat dibandingkan Regresi Linear pada kedua saham yang dianalisis. Untuk saham BBNI, XGBoost menghasilkan MAPE sebesar 0,39% dan RMSE sebesar 55,08, lebih baik dibandingkan Regresi Linear dengan MAPE 1,31% dan RMSE 91,02. Demikian pula pada saham BBTN, XGBoost mencatat MAPE 0,40% dan RMSE 15,30, mengungguli Regresi Linear dengan MAPE 1,56% dan RMSE 28,22. Temuan ini menunjukkan bahwa metode XGBoost lebih efektif dalam menangkap pola kompleks data historis saham perbankan. Oleh karena itu, pemilihan metode prediksi yang tepat perlu mempertimbangkan kemampuan model dalam mengakomodasi dinamika pasar yang fluktuatif. Penelitian ini memberikan kontribusi dalam pengembangan literatur prediksi harga saham serta panduan praktis bagi investor dalam menyusun strategi analisis berbasis data. Selain itu, penelitian ini memberikan kebaruan melalui perbandingan dua pendekatan statistik dan machine learning dalam konteks saham perbankan BUMN Indonesia yang masih terbatas dieksplorasi dalam studi sebelumnya.

Kata kunci: *Prediksi Harga Saham, Regresi Linear, XGBoost, Perbankan BUMN, Machine Learning*

PENDAHULUAN

Pasar saham telah menjadi salah satu indikator utama yang mencerminkan kondisi ekonomi suatu negara, termasuk Indonesia. Dalam konteks ini, sektor perbankan, terutama yang dikelola oleh Badan Usaha Milik Negara (BUMN), memegang peranan strategis sebagai motor penggerak ekonomi nasional (Rohyati et al., 2024). Saham-saham perbankan BUMN tidak hanya menarik bagi investor karena stabilitas dan fundamental perusahaannya, tetapi juga karena dianggap merepresentasikan kinerja pasar modal secara keseluruhan (Sofi et al., 2021). Namun demikian, tingginya volatilitas harga saham menjadi tantangan tersendiri bagi para pelaku pasar dalam mengambil keputusan investasi yang tepat. Perubahan harga yang cepat dan tidak menentu sering

kali menimbulkan keraguan, sehingga diperlukan metode prediksi yang akurat dan andal (Sreeraksha M S & Bhargavi M S, 2019).

Penelitian ini secara spesifik memilih saham BBNI (Bank Negara Indonesia) dan BBTN (Bank Tabungan Negara) sebagai objek studi utama. Pemilihan kedua saham ini didasarkan pada karakteristik pasar yang berbeda namun saling melengkapi; BBNI mewakili kelompok perbankan dengan kapitalisasi pasar besar (*big cap*) yang berfokus pada layanan korporasi dan internasional, sementara BBTN mewakili bank dengan fokus segmentasi pembiayaan perumahan yang sangat sensitif terhadap dinamika suku bunga nasional. Perbedaan profil kapitalisasi dan volatilitas ini penting untuk menguji sejauh mana model prediksi tetap akurat pada berbagai tingkat risiko dan perilaku pasar yang berbeda.

Namun demikian, tingginya volatilitas harga saham menjadi tantangan tersendiri bagi para pelaku pasar. Secara teoretis, prediksi harga saham dipengaruhi oleh tiga fenomena utama: Volatilitas, di mana perubahan harga terjadi secara cepat dan tidak menentu dalam waktu singkat; Non-linearity, yakni hubungan antara variabel input dan harga saham yang tidak bersifat linear sederhana atau proporsional; serta Autocorrelation, di mana nilai harga saham saat ini memiliki ketergantungan temporal yang kuat pada nilai-nilai di masa lampau. Perubahan harga yang kompleks ini sering kali menimbulkan keraguan, sehingga diperlukan metode prediksi yang akurat dan andal (Sreeraksha M S & Bhargavi M S, 2019).

Berbagai pendekatan telah digunakan dalam upaya memprediksi harga saham. Salah satu metode yang masih banyak diterapkan adalah regresi linear karena kemudahan interpretasi dan implementasinya. Namun, regresi linear memiliki keterbatasan dalam menangani pola hubungan non-linier, yang umum terjadi pada data keuangan. Untuk mengatasi hal ini, pendekatan berbasis *machine learning* seperti *Extreme Gradient Boosting (XGBoost)* mulai banyak digunakan dalam studi prediktif (Farouk et al., 2024). XGBoost merupakan algoritma *boosting* berbasis pohon keputusan yang dirancang untuk menangani kompleksitas data dan hubungan non-linier, meskipun penggunaannya memerlukan sumber daya komputasi yang lebih tinggi serta penanganan *overfitting* yang cermat (Chen & Guestrin, 2016).

Studi-studi sebelumnya telah meneliti penggunaan regresi linear maupun XGBoost secara terpisah dalam memprediksi harga saham. Namun, belum banyak yang secara eksplisit membandingkan kedua metode tersebut dalam konteks saham BUMN sektor perbankan di Indonesia. Padahal, efektivitas metode prediksi sangat dipengaruhi oleh struktur dan karakteristik data yang dianalisis. Idealnya, model prediksi harus mampu menghasilkan tingkat galat (*error*) yang rendah, sebagaimana diukur melalui metrik seperti *Root Mean Squared Error (RMSE)* dan *Mean Absolute Percentage Error (MAPE)* (Nurfalinda et al., 2025). Dalam praktiknya, kesenjangan antara nilai prediksi

dan aktual masih sering ditemukan, terutama ketika metode yang digunakan tidak sesuai dengan pola data yang dianalisis.

Oleh karena itu, penelitian ini bertujuan untuk membandingkan kinerja Regresi Linear dan XGBoost dalam memprediksi harga saham BBNI dan BBTN. Melalui pendekatan deskriptif-komparatif, diharapkan dapat diperoleh pemahaman komprehensif mengenai keunggulan masing-masing metode. Kontribusi penelitian ini mencakup:

1. Pengembangan Literatur: Memberikan nilai kebaruan (*novelty*) melalui studi komparatif antara statistik konvensional dan *machine learning* pada sektor perbankan BUMN Indonesia.
2. Validasi Metodologis: Menguji ketangguhan model XGBoost dalam menangani masalah *non-linearity* dan *autocorrelation* melalui fitur *lag* pada data runtun waktu.
3. Panduan Praktis: Memberikan referensi bagi investor dalam memilih strategi analisis yang lebih presisi dan adaptif terhadap dinamika pasar guna meminimalkan risiko investasi di era digital.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis deskriptif-komparatif untuk mengevaluasi dua metode prediksi harga saham, yaitu Regresi Linear dan *Extreme Gradient Boosting* (XGBoost). Data berupa harga penutupan harian (*closing price*) saham BBNI dan BBTN selama periode 11 Oktober 2023 hingga 11 Oktober 2024 diperoleh dari situs Investing.com. Analisis dilakukan menggunakan Google Colaboratory dengan pustaka Python melalui beberapa tahapan sistematis.

Pra-pemrosesan Data dan Rekayasa Fitur

Tahapan analisis mencakup akuisisi dan transformasi data. Selain fitur dasar (*Open, High, Low, Price*), penelitian ini melakukan rekayasa fitur dengan menambahkan indikator teknikal seperti *Moving Average* (MA), *Relative Strength Index* (RSI), dan *Trading Volume* guna memperkuat representasi finansial model. Untuk menangkap pola historis dan mengatasi masalah *autocorrelation*, dibuat fitur *lag* ($t-1$ dan $t-2$). Sebelum pemodelan, dilakukan *Min-Max Scaling* (normalisasi) untuk menyamakan rentang fitur antara 0 hingga 1, yang sangat krusial bagi algoritma berbasis gradien seperti XGBoost guna mencegah bias skala dan mempercepat konvergensi.

Pembagian Data (*Time-Series Split*)

Untuk menjaga integritas data runtun waktu, data tidak dibagi secara acak (random) guna menghindari *look-ahead bias*. Sebagai gantinya, digunakan metode *Time-Series Split*, di mana data dibagi secara kronologis menjadi data latih (80%) dan data uji (20%). Selain itu, untuk menjamin estimasi *error* yang lebih *robust*, diterapkan prosedur *Time-Series Cross-Validation* dengan metode *expanding window*.

Model Regresi Linear

Model Regresi Linear diterapkan untuk memodelkan hubungan linear antara variabel bebas dan terikat menggunakan persamaan:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n + \varepsilon \quad (1)$$

dengan y : nilai prediksi harga saham, β_0 : konstanta (intersep), β_i : koefisien regresi, dan x_1, x_2, \dots, x_n : variabel independen (fitur harga), dan ε adalah galat (error).

Sebelum model digunakan, dilakukan serangkaian uji diagnostik asumsi klasik yang ketat. Hasil uji ditampilkan melalui tabel nilai *Variance Inflation Factor* (VIF) untuk multikolinieritas, statistik Durbin-Watson untuk autokorelasi, serta plot residual untuk memverifikasi normalitas, linearitas, dan heteroskedastisitas. Hal ini memastikan model bersifat *Best Linear Unbiased Estimator* (BLUE).

Model Extreme Gradient Boosting (XGBoost)

Model XGBoost digunakan karena kemampuannya menangani data kompleks, *non-linear*, dan multivariat melalui pembangunan pohon keputusan secara bertahap:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathcal{F} \quad (2)$$

dengan \hat{y}_i : hasil prediksi pada data ke- i , f_k : fungsi dari pohon keputusan ke- k , x_i : fitur input untuk observasi ke- i , dan \mathcal{F} : ruang fungsi semua pohon keputusan yang mungkin.

Untuk mengatasi potensi *overfitting* (terutama yang terdeteksi pada data saham BBTN), model ini tidak menggunakan parameter *default*. Dilakukan Hyperparameter Tuning melalui *Grid Search* untuk mengoptimasi parameter kunci seperti *learning rate* (eta), *max_depth*, dan *n_estimators*.

2.5 Metrik Evaluasi Performa kedua model dievaluasi menggunakan dua metrik utama: *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE).

- RMSE: Mengukur akar rata-rata kuadrat selisih antara nilai aktual dan prediksi.
- MAPE: Mengukur rata-rata persentase kesalahan absolut terhadap nilai aktual.

Nilai RMSE dan MAPE yang lebih rendah menunjukkan performa prediksi yang lebih baik. Hasil evaluasi ini digunakan untuk menentukan metode yang paling efektif serta menganalisis bagaimana karakteristik data (seperti volatilitas) memengaruhi akurasi masing-masing metode.

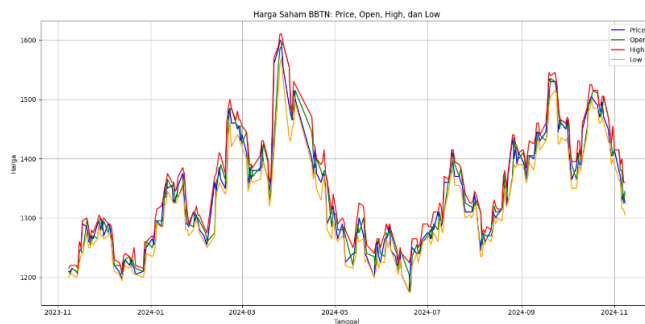
HASIL PENELITIAN DAN PEMBAHASAN

Deskripsi Data dan Transformasi

Data harga saham harian BBNI dan BBTN dikumpulkan dari situs Investing.com untuk periode 11 Oktober 2023 hingga 11 Oktober 2024. Variabel yang digunakan meliputi harga pembukaan (*open*), harga tertinggi (*high*), harga terendah (*low*), dan harga penutupan (*price*). Grafik pada Gambar 1 dan Gambar 2 memperlihatkan fluktuasi dinamis harga selama periode observasi, mencerminkan respons pasar terhadap berbagai kondisi ekonomi nasional.



Gambar 1. Grafik Data Saham BBNI untuk (Price, Open, High, Low)



Gambar 2. Grafik Data Saham BBTN untuk (Price, Open, High, Low)

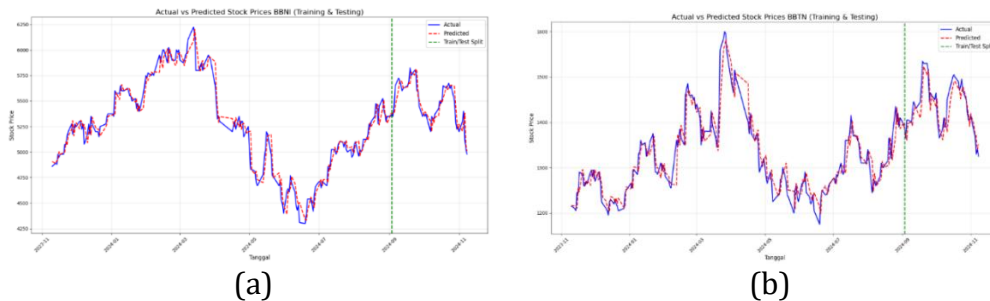
Untuk menangkap dinamika temporal, dilakukan transformasi data dengan menambahkan dua *lag* ($t-1$ dan $t-2$) pada masing-masing variabel. Penambahan *lag* ini bertujuan memodelkan ketergantungan nilai saat ini terhadap nilai historis yang berdekatan, yang dinilai paling relevan dalam prediksi jangka pendek (Leites et al., 2024). Selanjutnya, data dibagi menggunakan metode Time Series Split secara kronologis (80% *training set* dan 20% *test set*), guna menghindari *look-ahead bias* yang sering terjadi pada pembagian data secara acak. Sebelum masuk ke model, dilakukan Min-Max Scaling untuk menormalkan rentang data, yang sangat penting bagi algoritma berbasis gradien seperti XGBoost agar terhindar dari bias fitur.

Model Regresi Linear

Regresi linear digunakan sebagai model *baseline*. Meskipun uji asumsi klasik (seperti uji autokorelasi dengan Durbin-Watson) menunjukkan adanya pola

yang tidak sepenuhnya linear pada residual, model ini tetap diterapkan untuk memberikan gambaran awal. Model dilatih menggunakan delapan variabel prediktor (dua *lag* untuk *open*, *high*, *low*, dan *price*).

Visualisasi pada Gambar 3 menunjukkan bahwa model mampu mengikuti tren umum pergerakan harga saham, meskipun mengalami deviasi signifikan pada periode dengan fluktuasi tinggi. Hal ini menegaskan keterbatasan regresi linear dalam menangkap pola non-linear yang kompleks.



Gambar 3. Grafik Model Regresi Linear Saham (a) BBNI; (b) BBTN

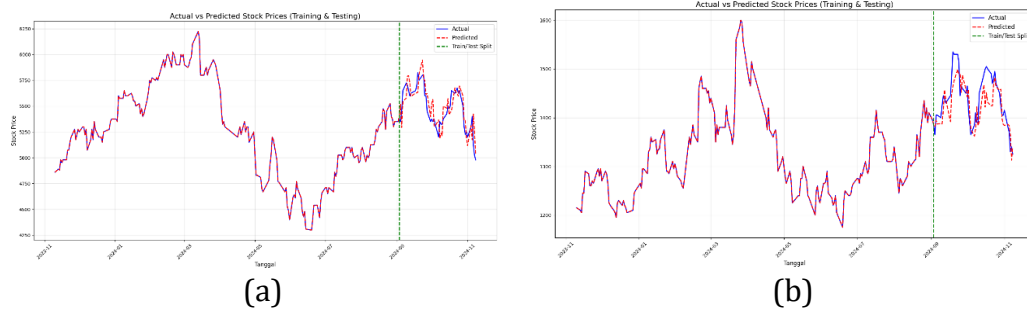
Tabel berikut menyajikan perbandingan antara harga aktual dan prediksi untuk kedua saham:

Tabel 1. Hasil Prediksi Harga Saham dengan Model Regresi Linear

Date	BBNI		BBTN	
	Actual	Predicted	Actual	Predicted
11/10/2023	4860	4907.272436	1215	1216.477138
11/13/2023	4890	4888.349771	1210	1215.609513
11/14/2023	4880	4917.482931	1205	1213.744178
11/15/2023	4980	4921.602783	1245	1210.271944
⋮	⋮	⋮	⋮	⋮
11/5/2024	5400	5282.538862	1370	1373.034734
11/6/2024	5125	5381.296103	1330	1370.335244
11/7/2024	5025	5068.291768	1340	1339.588209
11/8/2024	4980	5008.506625	1325	1351.330288

Model XGBoost

Model *Extreme Gradient Boosting* (*XGBoost*) digunakan sebagai metode pembandingan terhadap regresi linear dalam memprediksi harga saham. Berbeda dari regresi linear, *XGBoost* tidak mengharuskan pemenuhan asumsi klasik regresi, sehingga lebih fleksibel dalam menangani data dengan distribusi yang tidak normal maupun hubungan non-linear antar variabel.



Gambar 4. Grafik Model XGBoost Saham (a) BBNI; (b) BBTN

Berdasarkan hasil pada Gambar 4 performa model selama proses pelatihan menunjukkan tingkat akurasi yang sangat tinggi. Namun, saat diterapkan pada data pengujian, terlihat adanya penurunan performa yang mengindikasikan kemungkinan terjadi *overfitting*, terutama pada saham BBTN.

Model XGBoost pada saham BBNI menunjukkan kestabilan yang lebih baik antara data latih dan data uji, menandakan kemampuan generalisasi yang relatif kuat. Sebaliknya, pada saham BBTN, terdapat deviasi yang lebih mencolok antara hasil prediksi dan data aktual selama periode pengujian, yang mengisyaratkan bahwa model kurang mampu menangkap dinamika pasar secara konsisten untuk saham tersebut.

Tabel berikut menampilkan hasil prediksi harga saham menggunakan model XGBoost selama periode pengujian:

Tabel 2. Hasil Prediksi Harga Saham dengan Model XGBoost

Date	BBNI		BBTN	
	Actual	Predicted	Actual	Predicted
11/10/2023	4860	4859.731445	1215	1215.032837
11/13/2023	4890	4888.862305	1210	1209.964355
11/14/2023	4880	4880.806152	1205	1205.141602
11/15/2023	4980	4979.377441	1245	1244.957397
⋮	⋮	⋮	⋮	⋮
11/5/2024	5400	5174.159668	1370	1383.550171
11/6/2024	5125	5422.928223	1330	1364.286987
11/7/2024	5025	5261.874512	1340	1312.877563
11/8/2024	4980	5028.538574	1325	1340.114868

Evaluasi Model Regresi Linear dan XGBoost

Evaluasi dilakukan untuk membandingkan performa model regresi linear dan XGBoost dalam memprediksi harga saham. Berdasarkan hasil evaluasi, model XGBoost menunjukkan performa yang lebih unggul dibandingkan Regresi Linear, sebagaimana ditunjukkan oleh nilai *Mean Absolute Percentage Error* (MAPE) dan *Root Mean Square Error* (RMSE) yang lebih rendah. Hal ini mengindikasikan bahwa model XGBoost memiliki tingkat akurasi prediksi yang lebih tinggi, khususnya pada saham BBNI.

Tabel 3. Hasil Evaluasi Model

Model	Saham	MAPE	RMSE
Regresi Linear	BBNI	1,31161%	91,01843
	BBTN	1,55636%	28,22300
XGBoost	BBNI	0,39018%	55,08474
	BBTN	0,39925%	15,30188

Secara keseluruhan, model XGBoost terbukti lebih andal dalam menangkap pola pergerakan harga saham yang bersifat kompleks dan non-linear. Dengan demikian, model ini dapat dipertimbangkan sebagai pendekatan yang lebih efektif dalam prediksi harga saham. Hasil analisis penelitian juga menunjukkan bahwa (1) Harga saham BUMN perbankan dipengaruhi oleh banyak variabel makro yang tidak selalu linear. XGBoost melalui struktur *ensemble decision trees* mampu memodelkan hubungan kompleks ini secara lebih presisi dibandingkan garis lurus pada Regresi Linear, (2) Penggunaan fitur *lag* memungkinkan XGBoost menangkap pola ketergantungan waktu. Hal ini mendukung temuan Farouk et al. (2024) bahwa model *computational intelligence* lebih adaptif terhadap volatilitas pasar, dan (3) Karakteristik Saham: Meskipun XGBoost unggul pada keduanya, hasil BBNI lebih stabil. Hal ini dikarenakan kapitalisasi pasar BBNI yang lebih besar (*big cap*) cenderung memiliki tren yang lebih teratur dibandingkan BBTN yang lebih fluktuatif (Sreeraksha & Bhargavi, 2019).

KESIMPULAN DAN SARAN

Kesimpulan Penelitian ini menunjukkan bahwa model XGBoost memiliki kinerja prediktif yang lebih baik dibandingkan Regresi Linear dalam memprediksi harga saham BBNI dan BBTN. Hal ini dibuktikan oleh nilai MAPE yang sangat rendah (di bawah 1%), yang menunjukkan akurasi sangat tinggi. XGBoost mampu menangani pola kompleks, volatilitas, dan non-linearity pada data perbankan BUMN secara lebih efektif daripada pendekatan statistik konvensional.

Saran Untuk meningkatkan akurasi, penelitian mendatang disarankan menyertakan variabel fundamental (suku bunga) dan teknikal tambahan (RSI, Moving Average). Selain itu, penggunaan metode *Cross-Validation* berbasis *rolling window* dapat memperkuat validitas model terhadap perubahan siklus pasar jangka panjang.

REFERENSI

- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Farouk, M., Shaker, N., Abdelminaam, D., Elrashidy, O., Fathy, B., Khames, M., Mansour, M., Abdelrazeq, M., Ali, M., & Elazab, R. (2024). *Micrtsoft_Stock_Price: An Efficient Framework For Microsoft Stock Price Prediction Using Computational Intelligence. Journal of Computing and*

- Communication*, 3(1), 88–103.
<https://doi.org/10.21608/jocc.2024.339927>
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, 33(5), 2223–2273.
<https://doi.org/10.1093/rfs/hhaa009>
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts. <https://otexts.com/fpp3/>
- Leites, J., Cerqueira, V., & Soares, C. (2024). *Lag Selection for Univariate Time Series Forecasting using Deep Learning: An Empirical Study*. <http://arxiv.org/abs/2405.11237>
- Nurfalinda, Al Fiani, M., & Radzi Rathomi, M. (2025). Prediksi Temperatur Maksimum di Kota Tanjungpinang Menggunakan Model CNN-LSTM. *Jurnal Ilmiah Komputer Dan Informatika*, 14(1), 2715–7849.
<https://doi.org/10.34010/komputa.v14i1>
- Rohyati, Rokhmah, F. P. N., Syazeedah, H. N. U., Fitiyaningrum, R. I., Ramadhan, Gi., & Syahwildan, M. (2024). Tantangan dan Peluang Pasar Modal Indonesia dalam Meningkatkan Minat Investasi di Era Digital. *Jurnal Ilmiah Ekonomi Dan Bisnis*, 3, 909–918.
- Sofi, K., Sunge, A. S., Riady, S. R., & Kamalia, A. Z. (2021). PERBANDINGAN ALGORITMA LINEAR REGRESSION, LSTM, DAN GRU DALAM MEMPREDIKSI HARGA SAHAM DENGAN MODEL TIME SERIES. *SEMINASTIKA*, 3(1), 39–46.
<https://doi.org/10.47002/seminastika.v3i1.275>
- Sreeraksha M S, & Bhargavi M S. (2019). A Comparative Study of Machine Learning Models for Stock Market Rate Prediction. *International Journal of Computer Sciences and Engineering*, 7(6), 985–990.
<https://doi.org/10.26438/ijcse/v7i6.985990>
- Tandelilin, E. (2017). *Pasar Modal Manajemen Portofolio & Investasi*. Kanisius.