



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No. 4 (2025) pp: 3429-3438

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Prediksi Harga Saham PT Telkom Indonesia Tbk (TLKM) Menggunakan Linear Regression, Random Forest, dan XGBoost

Sari Lumbantoruan¹, Arnawati Duha², Tia Puji Astuti³, Ahmad Jurnaidi Wahidin⁴, Beni Rahmatullah⁵, Ika Kurniawati⁶

^{1,2,3}Informatika, Universitas Bina Sarana Informatika

^{4,5}Teknologi Informasi, Universitas Bina Sarana Informatika

⁶Sistem Informasi, Universitas Nusa Mandiri

¹sarilumbantoruan2021@gmail.com, ²arnawatiduha1110@gmail.com, ³tiapujiastuti01@gmail.com,

⁴ahmad.ajn@bsi.ac.id, ⁵beni.brh@bsi.ac.id, ⁶ika.iki@nusamandiri.ac.id

Abstrak

Prediksi harga saham merupakan salah satu aspek penting dalam pengambilan keputusan investasi, karena mampu membantu investor dalam mengantisipasi risiko dan mengoptimalkan keuntungan. PT Telkom Indonesia Tbk (TLKM) merupakan salah satu emiten blue chip di Bursa Efek Indonesia yang memiliki kapitalisasi pasar besar dan likuiditas tinggi, sehingga menarik untuk dianalisis dari sisi pergerakan harga sahamnya. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja tiga model machine learning, yaitu Linear Regression, Random Forest, dan XGBoost, dalam memprediksi harga penutupan saham TLKM berdasarkan data historis. Data yang digunakan merupakan data harga saham harian TLKM periode 7 November 2019 hingga 6 November 2024 yang diperoleh dari platform Kaggle, dengan variabel Date, Open, High, Low, Close, Adj Close, dan Volume. Tahapan penelitian meliputi pembersihan data, konversi rekayasa fitur berbasis informasi historis dan indikator teknikal seperti lag harga penutupan, moving average, volatilitas, dan return harian, serta pembagian data menjadi 80% data latih dan 20% data uji. Selanjutnya, ketiga model dibangun dan dievaluasi menggunakan metrik Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), dan koefisien determinasi (R^2). Hasil pengujian menunjukkan bahwa Linear Regression memberikan kinerja paling baik dibandingkan Random Forest dan XGBoost, dengan nilai RMSE sebesar 8,13, MAE 6,15, MAPE 0,19%, dan R^2 0,9997. Temuan ini mengindikasikan bahwa pada data saham TLKM dengan pola pergerakan yang relatif stabil, model linear sederhana masih mampu memberikan akurasi prediksi jangka pendek yang sangat tinggi dan dapat dijadikan alternatif yang efektif dibandingkan model ensemble yang lebih kompleks.

Kata kunci: Prediksi Harga Saham, PT Telkom Indonesia Tbk, Linear Regression, Random Forest, XGBoost, Machine Learning

1. Pendahuluan

Pasar modal merupakan salah satu pilar penting dalam perekonomian modern karena menyediakan sarana pendanaan bagi perusahaan sekaligus menjadi alternatif investasi bagi masyarakat. Pergerakan harga saham yang bersifat dinamis dan dipengaruhi oleh berbagai faktor ekonomi, politik, kinerja perusahaan, serta sentimen pasar membuat aktivitas prediksi harga saham menjadi topik yang terus mendapat perhatian dalam penelitian maupun praktik investasi. Berbagai teknik machine learning telah banyak dimanfaatkan untuk meningkatkan akurasi peramalan pasar saham, baik dengan memanfaatkan data teknikal maupun fundamental [1], [2]. Pendekatan deep learning dan machine learning juga dilaporkan mampu menangkap pola nonlinier pada data keuangan yang kompleks sehingga berpotensi memberikan prediksi yang lebih baik dibandingkan metode statistik tradisional [3].

Di Indonesia, PT Telkom Indonesia Tbk (TLKM) merupakan salah satu emiten blue chip dengan kapitalisasi pasar besar dan likuiditas tinggi, sehingga banyak dijadikan objek analisis oleh peneliti maupun praktisi. Sejumlah penelitian telah mengkaji prediksi harga saham TLKM menggunakan berbagai algoritma. Support Vector Machine (SVM) telah diterapkan untuk meramalkan harga penutupan saham PT Telkom Indonesia dan menunjukkan

Prediksi Harga Saham PT Telkom Indonesia Tbk (TLKM) Menggunakan Linear Regression, Random Forest, dan XGBoost

akurasi yang cukup baik [4], [5]. Linear Regression juga telah digunakan untuk memprediksi harga saham TLKM dan menghasilkan kinerja yang stabil pada data historis [6].

Penelitian terkait prediksi harga saham di luar TLKM juga sangat beragam. Kombinasi model lain seperti XGBoost yang digabungkan dengan LSTM menunjukkan bahwa pendekatan hibrida dapat meningkatkan kinerja peramalan dibandingkan model tunggal [11]. Selain itu, beberapa penelitian membandingkan kinerja Random Forest Regression dengan Support Vector Regression pada prediksi harga saham [12], serta menganalisis perbandingan antara Linear Regression dan Neural Network pada tugas prediksi harga saham [13]. Studi lain secara khusus mengkaji Linear Regression, Support Vector Regression, dan XGBoost pada saham setelah stock split dan menunjukkan keunggulan model tertentu pada kondisi data yang berbeda [14].

Penelitian terkait prediksi harga saham menggunakan Random Forest secara khusus juga menunjukkan hasil yang menjanjikan. Beberapa studi menerapkan Random Forest Regression untuk memodelkan hubungan nonlinier antara indikator teknikal dan harga penutupan saham, dan melaporkan bahwa model ini mampu mengurangi overfitting sekaligus memberikan performa yang kompetitif dibandingkan metode tradisional seperti regresi linier maupun model time series klasik [1]. Namun demikian, sebagian besar penelitian tersebut hanya berfokus pada evaluasi Random Forest sebagai satu model tunggal, tanpa membandingkannya secara sistematis dengan model lain seperti XGBoost dan Linear Regression pada instrumen saham tertentu.

Meskipun telah banyak penelitian terkait saham TLKM dan prediksi harga saham secara umum, studi yang secara khusus membandingkan kinerja tiga model yang relatif populer dan komputasionalnya efisien, yaitu Linear Regression, Random Forest, dan XGBoost, pada saham TLKM dengan menggunakan data historis harian lima tahun terakhir masih relatif terbatas. Padahal, ketiga model tersebut sering direkomendasikan sebagai baseline yang kuat untuk tugas regresi finansial serta lebih mudah diimplementasikan pada sistem pendukung keputusan investasi [1], [12]. Oleh karena itu, penelitian ini dilakukan untuk membangun dan mengevaluasi model Linear Regression, Random Forest, dan XGBoost dalam memprediksi harga penutupan saham TLKM berdasarkan data historis. Evaluasi model menggunakan metrik Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), dan koefisien determinasi (R^2). Hasil penelitian ini diharapkan dapat memberikan kontribusi empiris terhadap pemilihan model machine learning yang efektif dan efisien untuk prediksi harga saham TLKM serta menjadi referensi bagi pengembangan sistem pendukung keputusan berbasis data di pasar modal Indonesia.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan membangun model prediksi harga saham berbasis machine learning. Objek yang dianalisis adalah harga saham harian PT Telkom Indonesia Tbk (TLKM) yang diperdagangkan di Bursa Efek Indonesia. Fokus penelitian ini adalah membandingkan kinerja tiga model regresi, yaitu Linear Regression, Random Forest Regression, dan XGBoost Regression, dalam memprediksi harga penutupan saham berdasarkan data historis dan indikator teknikal [1], [6], [14].

2.1 Data Penelitian

Data yang digunakan merupakan data sekunder berupa harga saham harian TLKM yang diperoleh dari platform Kaggle, untuk periode 7 November 2019 hingga 6 November 2024. Dataset terdiri atas 1.212 observasi dengan variabel Date, Open, High, Low, Close, Adj Close, dan Volume. Saham TLKM dipilih karena termasuk kategori blue chip dengan kapitalisasi pasar besar dan likuiditas tinggi sehingga pergerakan harganya relevan untuk dianalisis [6].

Pada tahap awal, kolom Date diubah ke format tanggal (datetime) dan seluruh baris diurutkan secara kronologis. Kolom Volume yang semula menggunakan pemisah ribuan dikonversi menjadi bilangan bulat. Data diperiksa untuk memastikan tidak terdapat duplikasi tanggal maupun nilai yang tidak logis. Apabila terdapat baris dengan nilai yang tidak konsisten, baris tersebut dihapus agar kualitas data yang digunakan pada tahap pemodelan tetap terjaga [1].

2.2 Pra-pemrosesan dan Rekayasa Fitur

Setelah proses pembersihan data, dilakukan pra-pemrosesan dan rekayasa fitur (feature engineering) untuk memperkaya informasi yang diberikan kepada model. Rekayasa fitur dilakukan dengan memanfaatkan pola historis harga dan indikator teknikal yang umum digunakan dalam analisis pasar saham [9].

Beberapa fitur turunan yang dibentuk antara lain:

- **Lag harga penutupan**, yaitu harga penutupan beberapa hari sebelumnya (misalnya Close_lag1 sampai Close_lag5) untuk merepresentasikan ketergantungan harga saat ini terhadap harga masa lalu jangka pendek.
- **Moving Average (MA)**, misalnya rata-rata bergerak 7 hari (MA_7) dan 21 hari (MA_21) untuk menggambarkan tren jangka pendek dan menengah.
- **Volatilitas**, yang dihitung sebagai simpangan baku harga penutupan dalam jendela waktu tertentu untuk menggambarkan tingkat fluktuasi harga historis.
- **Return harian**, yang dihitung sebagai persentase perubahan harga penutupan dari satu hari ke hari berikutnya untuk merepresentasikan dinamika perubahan harga secara relatif.

Pembentukan indikator tersebut menyebabkan beberapa baris awal tidak memiliki nilai lengkap karena membutuhkan data beberapa hari sebelumnya. Oleh karena itu, baris yang mengandung nilai kosong akibat perhitungan fitur teknikal dihapus, dan hanya observasi dengan fitur lengkap yang digunakan pada tahap pelatihan dan pengujian model [1].

2.3 Pembagian Data Latih dan Data Uji

Pembagian data dilakukan dengan mempertimbangkan karakteristik deret waktu. Setelah data diurutkan berdasarkan tanggal dan seluruh fitur terbentuk, dataset dibagi menjadi dua bagian, yaitu data latih (training set) dan data uji (test set). Sekitar 80% observasi pertama digunakan sebagai data latih, sedangkan 20% observasi terakhir digunakan sebagai data uji.

Pendekatan ini dipilih agar model hanya belajar dari informasi masa lalu dan dievaluasi pada periode waktu yang lebih baru, sehingga skenario pengujian lebih mendekati kondisi prediksi nyata di pasar saham [12]. Data latih digunakan untuk membangun dan menyesuaikan parameter model, sedangkan data uji digunakan untuk mengukur kinerja akhir model yang telah dilatih.

2.4 Model Prediksi

Penelitian ini membandingkan tiga model machine learning yang banyak digunakan pada tugas regresi harga saham, yaitu Linear Regression, Random Forest Regression, dan XGBoost Regression [14], [13]

- Linear Regression digunakan sebagai model dasar (baseline) yang mengasumsikan hubungan linear antara fitur masukan dan harga penutupan sebagai variabel keluaran. Model ini sederhana, mudah diinterpretasikan, dan sering dijadikan acuan awal dalam penelitian prediksi harga saham [6].
- Random Forest **Regression** merupakan metode ensemble berbasis pohon keputusan yang membangun banyak pohon pada sampel data dan subset fitur yang berbeda, kemudian menggabungkan prediksi setiap pohon dengan cara rata-rata. Pendekatan ini mampu menangani hubungan nonlinier dan interaksi antar fitur serta lebih tahan terhadap overfitting dibandingkan single decision tree [12].
- XGBoost **Regression** merupakan implementasi dari teknik gradient boosting yang menggabungkan pohon-pohon keputusan secara bertahap. Setiap pohon baru dilatih untuk mengurangi kesalahan dari pohon sebelumnya. XGBoost dikenal memiliki kinerja tinggi dan efisiensi komputasi yang baik pada berbagai permasalahan regresi, termasuk prediksi harga saham [14]

Pemilihan parameter awal pada masing-masing model mengacu pada pengaturan yang umum digunakan dalam literatur, kemudian dilakukan penyesuaian terbatas (tuning sederhana) melalui beberapa percobaan pada data latih untuk memperoleh kombinasi parameter yang memberikan kinerja yang lebih baik tanpa meningkatkan kompleksitas model secara berlebihan.

2.5 Evaluasi Kinerja Model

Kinerja ketiga model dievaluasi menggunakan empat metrik, yaitu Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), dan koefisien determinasi (R^2). RMSE dan MAE mengukur besarnya kesalahan prediksi dalam skala harga saham, sedangkan MAPE mengukur rata-rata kesalahan dalam bentuk persentase sehingga memudahkan interpretasi akurasi relatif. Koefisien determinasi R^2 menunjukkan proporsi variasi harga aktual yang dapat dijelaskan oleh model [13].

Model yang baik ditandai dengan nilai RMSE, MAE, dan MAPE yang lebih kecil serta nilai R^2 yang mendekati satu. Perbandingan nilai keempat metrik pada data uji digunakan untuk menentukan model mana di antara Linear Regression, Random Forest, dan XGBoost yang memiliki performa paling baik dalam memprediksi harga penutupan saham TLKM.

2.6 Lingkungan Implementasi

Seluruh tahapan pra-pemrosesan data, rekayasa fitur, pelatihan model, dan evaluasi kinerja diimplementasikan menggunakan bahasa pemrograman Python. Pustaka yang digunakan antara lain Pandas dan NumPy untuk pengolahan data, Scikit-learn untuk implementasi Linear Regression dan Random Forest, serta pustaka XGBoost untuk implementasi model XGBoost Regression. Eksperimen dijalankan pada lingkungan notebook seperti Google Colab atau Jupyter Notebook sehingga memudahkan proses dokumentasi, replikasi, dan pengembangan lebih lanjut [3].

2.7 Deskripsi Data dan Fitur

Data historis harga saham harian PT Telkom Indonesia Tbk (TLKM) yang digunakan dalam penelitian ini mencakup periode 7 November 2019 hingga 6 November 2024 dengan total 1.212 observasi. Secara umum, pergerakan harga saham TLKM menunjukkan pola tren yang relatif stabil dengan beberapa fase kenaikan dan penurunan yang dipengaruhi kondisi pasar dan faktor makroekonomi. Nilai harga penutupan (Close) berada pada kisaran sekitar 2.500 hingga 4.800 rupiah per lembar saham, sedangkan volume perdagangan menunjukkan variasi yang cukup besar namun tetap berada pada rentang yang wajar untuk saham berkapitalisasi besar.

Setelah dilakukan pra-pemrosesan dan rekayasa fitur, terbentuk sejumlah variabel masukan berupa lag harga penutupan, moving average (MA_7 dan MA_21), volatilitas historis, serta return harian. Fitur-fitur ini digunakan sebagai input bagi ketiga model yang dibandingkan. Baris data yang tidak memiliki nilai lengkap akibat perhitungan indikator teknikal dihapus, sehingga hanya observasi dengan fitur lengkap yang digunakan pada tahap pelatihan dan pengujian. Pendekatan tersebut sejalan dengan praktik umum pada penelitian prediksi harga saham berbasis deret waktu [1], [9].

2.8 Hasil Pelatihan dan Pengujian Model

Model Linear Regression, Random Forest Regression, dan XGBoost Regression dilatih menggunakan sekitar 80% data sebagai data latih dan dievaluasi pada 20% data sisanya sebagai data uji. Hasil evaluasi kinerja ketiga model pada data uji ditunjukkan oleh nilai RMSE, MAE, MAPE, dan R^2 .

Secara ringkas, kinerja model dapat dirangkum sebagai berikut:

- Linear Regression menghasilkan RMSE sebesar 8,13, MAE 6,15, MAPE 0,19%, dan R^2 0,9997.
- Random Forest **Regression** menghasilkan RMSE sebesar 34,24, MAE 24,82, MAPE 0,81%, dan R^2 0,9945.
- XGBoost **Regression** menghasilkan RMSE sebesar 42,69, MAE 34,14, MAPE 1,11%, dan R^2 0,9915.

Berdasarkan nilai-nilai tersebut, terlihat bahwa Linear Regression memiliki RMSE, MAE, dan MAPE paling kecil serta nilai R^2 paling mendekati satu. Hal ini menunjukkan bahwa Linear Regression mampu memberikan prediksi harga penutupan saham TLKM yang paling mendekati nilai aktual dibandingkan dua model lainnya.

3. Hasil dan Diskusi

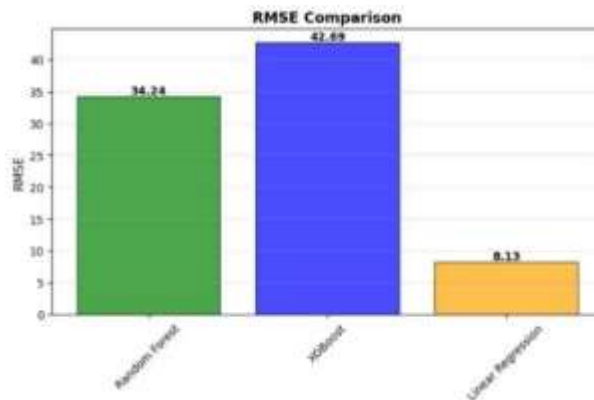
3.1 Perbandingan Kinerja Model

Perbandingan nilai RMSE dan MAE menunjukkan bahwa rata-rata selisih absolut antara nilai prediksi dan nilai aktual pada model Linear Regression jauh lebih kecil daripada Random Forest maupun XGBoost. Nilai MAPE sebesar 0,19% mengindikasikan bahwa kesalahan relatif rata-rata prediksi Linear Regression hanya sekitar nol koma sekian persen dari harga aktual, yang dapat dikategorikan sangat kecil dalam konteks prediksi harga saham harian. Sementara itu, Random Forest dan XGBoost memiliki MAPE yang masih cukup baik (di bawah 2%), tetapi tetap berada di atas Linear Regression.



Gambar 1 Pergerakan historis harga penutupan saham PT Telkom Indonesia Tbk (TLKM) periode 2019–2024.

Pergerakan historis harga penutupan saham TLKM periode 2019–2024 ditunjukkan pada Gambar 1. Terlihat beberapa fase kenaikan dan penurunan harga yang cukup tajam, namun secara umum pola pergerakan harga cenderung stabil pada rentang 2.500–4.800 rupiah.



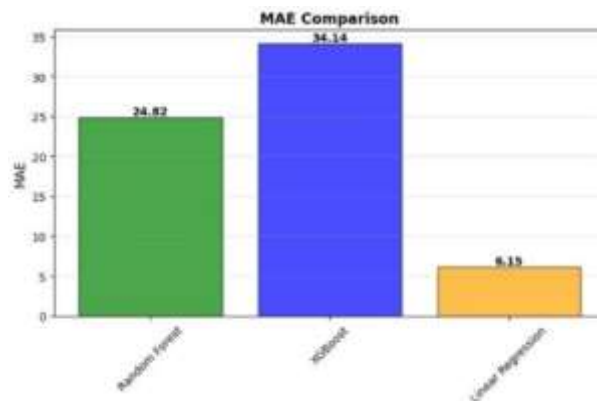
Gambar 2 Perbandingan RMSE Comparison

Perbandingan Gambar 2 menunjukkan perbandingan nilai *Root Mean Squared Error* (RMSE) dari tiga model, yaitu Random Forest, *XGBoost*, dan Linear Regression. Terlihat bahwa Linear Regression menghasilkan RMSE paling kecil yaitu 8,13, sehingga memiliki galat prediksi terkecil dan kinerja terbaik di antara ketiga model. Random Forest memiliki RMSE sebesar 34,24, sedangkan *XGBoost* memiliki RMSE terbesar yaitu 42,69, yang

DOI: <https://doi.org/10.31004/riggs.v4i4.4158>

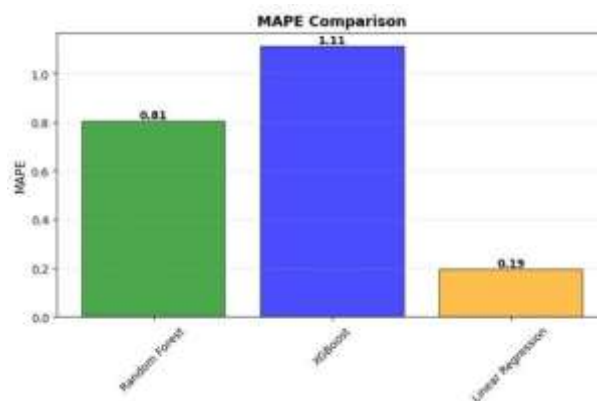
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

berarti kedua model ini menghasilkan kesalahan prediksi yang lebih besar dibandingkan Linear Regression. Secara keseluruhan, grafik ini mengindikasikan bahwa untuk data harga saham TLKM yang digunakan dalam penelitian, model Linear Regression lebih efektif dalam meminimalkan kesalahan prediksi dibandingkan model non-linier Random Forest dan *XGBoost*.



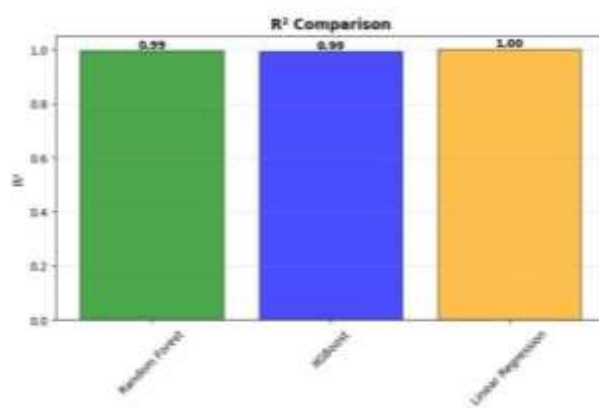
Gambar 3 Perbandingan MAE Comparison

Gambar 3 menyajikan perbandingan nilai *Mean Absolute Error* (MAE) dari tiga model, yaitu Random Forest, *XGBoost*, dan Linear Regression. Terlihat bahwa Linear Regression memiliki nilai MAE paling kecil, yaitu 6,15, yang menunjukkan bahwa rata-rata selisih absolut antara nilai prediksi dan nilai aktual paling rendah dibandingkan dua model lainnya. Random Forest menghasilkan MAE sebesar 24,82, sedangkan *XGBoost* memiliki MAE terbesar yaitu 34,14, sehingga kesalahan rata-rata prediksi keduanya lebih besar. Hasil ini menguatkan temuan sebelumnya bahwa pada data harga saham TLKM yang digunakan, Linear Regression memberikan akurasi prediksi yang lebih baik dibandingkan Random Forest maupun *XGBoost*.



Gambar 4 Perbandingan MAPE Comparison

Gambar 4 menampilkan perbandingan nilai *Mean Absolute Percentage Error* (MAPE) untuk tiga model, yaitu Random Forest, *XGBoost*, dan Linear Regression. MAPE mengukur rata-rata persentase kesalahan prediksi terhadap nilai aktual. Terlihat bahwa Linear Regression memiliki MAPE paling kecil, yaitu 0,19%, sehingga kesalahan persentase rata-rata prediksinya paling rendah dan kinerjanya paling baik. Random Forest memiliki MAPE sebesar 0,81, sedangkan *XGBoost* mencapai 1,11, yang menunjukkan bahwa kedua model ini menghasilkan kesalahan persentase yang lebih besar dibandingkan Linear Regression. Dengan demikian, berdasarkan metrik MAPE, Linear Regression kembali terbukti sebagai model yang paling akurat dalam memprediksi harga penutupan saham TLKM pada penelitian ini.



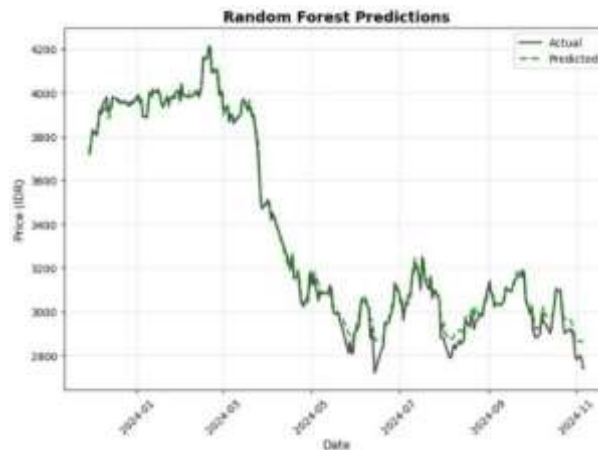
Gambar 5 Perbandingan R² Comparison

Gambar 5 menampilkan perbandingan nilai koefisien determinasi (R^2) untuk tiga model, yaitu Random Forest, *XGBoost*, dan Linear Regression. Ketiga model sama-sama menghasilkan nilai R^2 yang sangat tinggi, yakni 0,99 untuk Random Forest, 0,99 untuk *XGBoost*, dan 1,00 untuk Linear Regression. Nilai R^2 yang mendekati 1 menunjukkan bahwa hampir seluruh variasi harga penutupan saham TLKM pada data uji dapat dijelaskan oleh variabel-variabel input yang digunakan. Meskipun perbedaannya relatif kecil, Linear Regression memiliki R^2 paling tinggi (1,00) sehingga secara teoritis mampu menjelaskan proporsi variansi data sedikit lebih baik dibandingkan Random Forest dan *XGBoost*, yang menguatkan hasil sebelumnya bahwa Linear Regression merupakan model dengan kinerja terbaik pada penelitian ini.



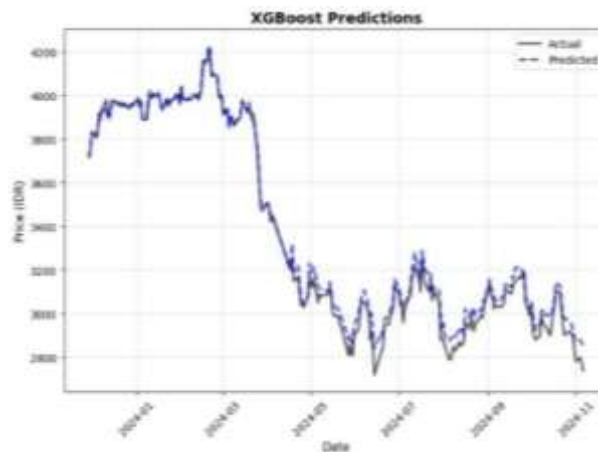
Gambar 6 Perbandingan Liner Regression

Gambar 6 menunjukkan perbandingan antara harga penutupan aktual dan hasil prediksi *model* Linear Regression untuk saham TLKM pada periode data uji tahun 2024. Sumbu horizontal menampilkan waktu (*Date*), sedangkan sumbu vertikal menunjukkan harga saham dalam Rupiah (*Price/IDR*). Garis hitam merepresentasikan data aktual, sementara garis putus-putus kuning menunjukkan nilai prediksi. Terlihat bahwa kedua garis hampir selalu saling tumpang tindih, baik pada fase tren naik di awal tahun, penurunan tajam di pertengahan tahun, maupun fluktuasi harga pada periode berikutnya. Kedekatan pola ini mengindikasikan bahwa model Linear Regression mampu mengikuti pergerakan harga saham dengan sangat baik, sehingga kesalahan prediksinya relatif kecil dan konsisten dengan nilai RMSE, MAE, MAPE, dan R^2 yang menunjukkan kinerja tinggi.



Gambar 7 Perbandingan Random Forest Predictions

Gambar 7 menampilkan perbandingan antara harga penutupan aktual dan hasil prediksi model Random Forest untuk saham TLKM pada data uji tahun 2024. Sumbu horizontal menunjukkan waktu (*Date*), sedangkan sumbu vertikal merepresentasikan harga saham dalam Rupiah (*Price/IDR*). Garis hitam menggambarkan harga aktual, sementara garis putus-putus hijau menunjukkan nilai prediksi Random Forest. Secara umum, Random Forest mampu mengikuti pola pergerakan harga—baik tren penurunan tajam di pertengahan tahun maupun fluktuasi pada periode berikutnya—yang tampak dari bentuk kurva prediksi yang relatif searah dengan data aktual. Namun, pada beberapa titik, terutama di sekitar puncak dan lembah harga, garis prediksi terlihat tidak sepenuhnya menumpuk dengan garis aktual, sehingga masih muncul selisih yang cukup jelas. Hal ini konsisten dengan nilai RMSE, MAE, dan MAPE yang lebih besar dibandingkan Linear Regression, yang mengindikasikan bahwa meskipun Random Forest dapat menangkap tren umum, akurasi prediksinya masih lebih rendah dibandingkan model terbaik dalam penelitian ini.



Gambar 8 Perbandingan XGBoost Predictions

Gambar 8 menampilkan perbandingan antara harga penutupan aktual dan hasil prediksi model *XGBoost* untuk saham TLKM pada data uji tahun 2024. Sumbu horizontal menunjukkan waktu (*Date*), sedangkan sumbu vertikal menunjukkan harga saham dalam Rupiah (*Price/IDR*). Garis hitam merepresentasikan harga aktual, sementara garis putus-putus biru menunjukkan nilai prediksi *XGBoost*. Terlihat bahwa *XGBoost* mampu mengikuti pola tren utama—penurunan harga di pertengahan tahun serta fluktuasi pada periode berikutnya—namun garis prediksi sering berada di atas garis aktual, terutama pada beberapa puncak dan lembah harga. Pola ini mengindikasikan adanya kecenderungan model untuk *overestimate* harga saham, sehingga selisih antara nilai prediksi dan aktual relatif lebih besar. Hal tersebut selaras dengan nilai RMSE, MAE, dan MAPE yang paling tinggi di antara ketiga

DOI: <https://doi.org/10.31004/riggs.v4i4.4158>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

model, sehingga dapat disimpulkan bahwa *XGBoost* adalah model dengan akurasi terendah dalam penelitian ini meskipun masih mampu menangkap bentuk umum pergerakan harga.

Dari sisi R^2 , ketiga model sama-sama menghasilkan nilai yang tinggi di atas 0,99, yang menunjukkan bahwa hampir seluruh variasi harga penutupan pada data uji dapat dijelaskan oleh variabel-variabel masukan. Namun demikian, Linear Regression masih unggul dengan nilai R^2 tertinggi, sehingga secara keseluruhan dapat disimpulkan bahwa model ini memberikan kinerja terbaik pada data yang dianalisis.

Hasil ini menarik karena menunjukkan bahwa pada data saham TLKM yang cenderung memiliki pola pergerakan harga yang relatif stabil, model linear yang sederhana justru dapat mengungguli model ensemble yang lebih kompleks seperti Random Forest dan *XGBoost*. Temuan ini sejalan dengan hasil penelitian yang menyatakan bahwa pada beberapa jenis saham dengan pola tren yang tidak terlalu volatil, model linear masih dapat memberikan performa yang sangat kompetitif [6], [13]. Di sisi lain, penelitian lain menunjukkan bahwa Random Forest, *XGBoost*, maupun model deep learning cenderung lebih unggul ketika pola data bersifat sangat nonlinier dan volatil [12], [14], [7].

3.2 Pembahasan

Beberapa faktor dapat menjelaskan keunggulan Linear Regression dalam penelitian ini. Pertama, fitur-fitur yang digunakan—seperti lag harga penutupan, moving average, dan return harian—secara teoritis memiliki hubungan yang relatif linear dengan harga penutupan saat ini. Dalam kondisi seperti ini, model Linear Regression mampu merepresentasikan hubungan tersebut secara efektif tanpa memerlukan struktur yang kompleks. Kedua, karakteristik saham TLKM sebagai saham blue chip cenderung lebih stabil dibandingkan saham berkapitalisasi kecil, sehingga pola historis harga dan indikator teknikalnya lebih teratur dan mudah ditangkap oleh model linear. Stabilitas ini menyebabkan keuntungan tambahan dari penggunaan model nonlinier dan ensemble seperti Random Forest dan *XGBoost* tidak selalu memberikan peningkatan kinerja yang signifikan dibandingkan Linear Regression.

Temuan ini konsisten dengan beberapa penelitian yang melaporkan bahwa dalam konteks saham tertentu, Linear Regression dapat menghasilkan performa yang sangat baik dan bahkan mengungguli model yang lebih kompleks [6], [13]. Namun, hasil penelitian ini tidak serta-merta meniadakan pentingnya model ensemble dan deep learning. Pada saham dengan volatilitas tinggi, horizon prediksi yang lebih panjang, atau kombinasi variabel teknikal dan fundamental yang lebih kompleks, model Random Forest, *XGBoost*, maupun LSTM dan CNN-LSTM tetap berpotensi memberikan keunggulan kinerja [7], [8], [9].

Secara praktis, hasil penelitian ini memberikan indikasi bahwa untuk prediksi harga penutupan harian saham TLKM pada periode dan fitur yang digunakan, Linear Regression dapat dijadikan pilihan utama karena mampu memberikan akurasi yang sangat tinggi dengan komputasi yang lebih sederhana dan interpretasi yang lebih mudah. Model ini dapat menjadi baseline kuat yang dapat dikembangkan lebih lanjut, misalnya dengan pengayaan fitur atau pengujian pada horizon prediksi yang berbeda sebelum mempertimbangkan penggunaan model yang lebih kompleks.

4. Kesimpulan

Berdasarkan hasil pengujian pada data uji, dapat disimpulkan bahwa proses pra-pemrosesan data dan rekayasa fitur yang meliputi pembentukan lag harga penutupan, moving average, volatilitas, dan return harian telah berhasil menghasilkan himpunan fitur yang representatif sebagai masukan model prediksi harga penutupan saham TLKM. Ketiga model yang digunakan, yaitu Linear Regression, Random Forest Regression, dan *XGBoost* Regression, sama-sama menunjukkan kinerja yang sangat baik dengan nilai R^2 di atas 0,99, sehingga sebagian besar variasi harga penutupan dapat dijelaskan oleh variabel-variabel input yang digunakan; hal ini mengonfirmasi bahwa pendekatan machine learning berbasis data historis dan indikator teknikal cukup efektif untuk tugas prediksi harga saham harian. Di antara ketiganya, Linear Regression memberikan kinerja terbaik dengan RMSE sebesar 8,13, MAE 6,15, MAPE 0,19%, dan R^2 0,9997, yang menunjukkan bahwa pada periode dan konfigurasi fitur yang digunakan, hubungan antara variabel masukan dan harga penutupan saham TLKM dapat dimodelkan secara sangat baik oleh model linear yang relatif sederhana. Sementara itu, Random Forest Regression dan *XGBoost* Regression tetap menghasilkan performa yang baik dengan kesalahan yang relatif rendah dan nilai R^2 yang tinggi, tetapi masih berada di bawah Linear Regression, sehingga mengindikasikan bahwa pada saham TLKM yang memiliki pola

pergerakan harga cenderung stabil, peningkatan kompleksitas model tidak selalu berbanding lurus dengan peningkatan akurasi dibandingkan model linear. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa Linear Regression dapat dijadikan model acuan (*baseline*) yang kuat untuk prediksi harga penutupan harian saham TLKM pada konfigurasi data dan fitur yang digunakan. Temuan ini juga memberikan implikasi bahwa pemilihan model untuk prediksi harga saham perlu mempertimbangkan karakteristik data dan tujuan penggunaannya. Untuk pengembangan penelitian selanjutnya, disarankan untuk mengeksplorasi kombinasi variabel teknikal dan fundamental, memperluas objek penelitian ke beberapa saham lain dalam satu sektor, menguji horizon prediksi yang berbeda, serta membandingkan hasil model machine learning dengan pendekatan deep learning sehingga dapat diperoleh gambaran yang lebih komprehensif mengenai kinerja berbagai model pada kondisi pasar yang beragam.

Referensi

1. M. M. Kumbure, C. Lohrmann, P. Luukka, and J. Porras, "Machine learning techniques and data for stock market forecasting: A literature review," *Expert Syst. Appl.*, vol. 197, no. December 2021, p. 116659, 2022, doi: 10.1016/j.eswa.2022.116659.
2. Raavihemalatha, GodiAmulya, and C. S. N. Sailalitha, "Deep Learning for Stock Market Prediction," *2024 2nd Int. Conf. Comput. Charact. Tech. Eng. Sci. IC3TES 2024*, 2024, doi: 10.1109/IC3TES62412.2024.10877452.
3. Rhoda Adura Adeleye, Tula Sunday Tubokirifuruar, Binaebi Gloria Bello, Ndubuisi Leonard Ndubuisi, Onyeka Franca Asuzu, and Oluwaseyi Rita Owolabi, "Machine Learning for Stock Market Forecasting: a Review of Models and Accuracy," *Financ. Account. Res. J.*, vol. 6, no. 2, pp. 112–124, 2024, doi: 10.51594/farj.v6i2.783.
4. D. Rahmadini, N. D. F. Salsabila, A. Kunaefi, and D. Rolliawati, "Analisis Peramalan Harga Penutupan Saham PT. Telekomunikasi Indonesia dengan Support Vector Machine (SVM)," *J. Ilm. Educic Pendidik. dan Inform.*, vol. 11, no. 1, pp. 22–33, 2024, doi: 10.21107/educic.v11i1.22120.
5. W. R. U. Fadilah, D. Agfiannisa, and Y. Azhar, "Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode Support Vector Machine," *Fountain Informatics J.*, vol. 5, no. 2, p. 45, 2020, doi: 10.21111/fij.v5i2.4449.
6. S. Daruyani, O. D. Purbiyanti, and M. F. Hardeva, "Implementasi Machine Learning
7. Menggunakan Algoritma Linear Regression Untuk Memprediksi Harga Saham Pt. Telkom Indonesia Tbk," *J. Inf. Syst. Informatics Comput. Issue Period.*, vol. 9, no. 1, pp. 81–91, 2025, doi: 10.52362/jisicom.v9i1.1910.
8. F. R. Pratama, B. Santoso, and S. Kacung, "Prediksi Harga Saham Pt Telkom Menggunakan Metode Cnn-Lstm," *J. Inf. Syst. Manag.*, vol. 7, no. 1, pp. 66–70, 2025, doi: 10.24076/joism.2025v7i1.2087.
9. M. F. Hadi and D. Priyanto, "Perbandingan Kinerja RNN , LSTM , dan GRU dalam," no. September 2025, pp. 265–269, 2024.
10. Furizal, A. Ritonga, A. Ma'arif, and I. Suwarno, "Stock Price Forecasting with Multivariate Time Series Long Short-Term Memory: A Deep Learning Approach," *J. Robot. Control*, vol. 5, no. 5, pp. 1322–1335, 2024, doi: 10.18196/jrc.v5i5.22460.
11. S. J. Pipin, R. Purba, and H. Kurniawan, "Prediksi Saham Menggunakan Recurrent Neural Network (RNN-LSTM) dengan Optimasi Adaptive Moment Estimation," *J. a. Comput. Syst. Informatics*, vol. 4, no. 4, pp. 806–815, 2023, doi: 10.47065/josyc.v4i4.4014.
12. 10.47065/josyc.v4i4.4014.
13. P. H. Vuong, T. T. Dat, T. K. Mai, P. H. Uyen, and P. T. Bao, "Stock-price forecasting based on XGBoost and LSTM," *Comput. Syst. Sci. Eng.*, vol. 40, no. 1, pp. 237–246, 2022, doi: 10.32604/CSSE.2022.017685.
14. M. Y. Urrochman, H. Asy'ari, and F. A. Hizham, "Performance Comparison of Random Forest Regression, Svr Models in Stock Price Prediction," *J. Pilar Nusa Mandiri*, vol. 21, no. 1, pp. 16–23, 2025, doi: 10.33480/pilar.v21i1.6072.
15. E. C. Wibowo and A. D. Cahyono, "Analisis Perbandingan Algoritma Regresi Linier dengan Neural Network untuk Prediksi Harga Saham Comparative Analysis of Linear Regression and Neural Network Algorithms for Stock Price Prediction," *Sist. J. Sist. Inf.*, vol. 14, no. 4, pp. 2540–9719, 2025, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
16. M. Y. Andrika and M. Rahardi, "Comparative Study of Linear Regression, SVR, and XGBoost for Stock Price Prediction After a Stock Split," *J. Appl. Informatics Comput.*, vol. 9, no. 4, pp. 1817–1824, 2025, doi: 10.30871/jaic.v9i4.10220.
17. M. Parker, "Stock Price Prediction Using a Stacked Heterogeneous Ensemble," 2025.