

PREDICTION OF BANK NEO COMMERCE STOCK PRICES USING A MULTIPLE LINEAR REGRESSION ALGORITHM

PREDIKSI HARGA SAHAM BANK NEO COMMERCE DENGAN MENGUNAKAN ALGORITMA REGRESI LINIER BERGANDA

Ulil Amri¹, Sriani²

¹Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara,
Kec. Pancur Batu, Kabupaten Deli Serdang, Sumatera Utara 20353, Indonesia
email: uamriharahap@gmail.com, sriani@uinsu.ac.id

Abstract – This study addresses the challenge of forecasting Bank Neo Commerce (BBYB) daily prices, where volatility and data leakage often bias results. We build a leakage-free pipeline to predict next-day Adjusted Close using multiple linear regression (OLS) with $t-1$ predictors lags and technical indicators: AdjClose_{t-1} , AdjClose_{t-5} , Vol_{t-1} , SMA5_{t-1} , EMA5_{t-1} , and RSI14_{t-1} . Daily BBYB.JK data from Yahoo Finance (12 March 2019–12 March 2025) are evaluated with a 5-fold rolling time-series split, and metrics are reported as mean $\pm$ SD. The goal is to assess OLS accuracy and its practical value against a persistence baseline. OLS attains MAE 25.15 ± 18.24 , MSE $2,344.41 \pm 2,956.78$, R^2 0.94 ± 0.06 , while the baseline AdjClose_{t-1} yields MAE 22.40 ± 17.21 , MSE $1,894.99 \pm 2,420.42$, R^2 0.96 ± 0.04 . A walk-forward long-only backtest (0.1% fee) delivers Final Value 1.04 versus 0.72 for buy-and-hold, with lower volatility and drawdown. The approach is interpretable, reproducible, and ready for extensions (feature reduction/regularization, non-linear models, return/volatility features).

Keywords - multiple linear regression, machine learning, stock price prediction, Bank Neo Commerce, Python.

Abstrak – Penelitian ini menanggapi masalah prediksi harian harga saham Bank Neo Commerce (BBYB) yang volatil dan rawan kebocoran data. Kami membangun pipeline bebas kebocoran untuk memprediksi Adjusted Close $t-1$ menggunakan regresi linier berganda (OLS) dengan prediktor $t-1$ berupa *lag* dan indikator teknikal (AdjClose_{t-1} , AdjClose_{t-5} , Vol_{t-1} , SMA5_{t-1} , EMA5_{t-1} , RSI14_{t-1}). Data harian BBYB.JK (Yahoo Finance, 12 Maret 2019–12 Maret 2025) dievaluasi memakai time-series split *rolling* 5-lipatan dan metrik dilaporkan sebagai mean $\pm$ SD. Tujuan studi adalah menilai akurasi OLS serta nilai praktisnya dibanding baseline persistensi. Hasil menunjukkan OLS mencapai MAE $25,15 \pm 18,24$, MSE $2.344,41 \pm 2.956,78$, R^2 $0,94 \pm 0,06$, sedangkan baseline AdjClose_{t-1} lebih baik: MAE $22,40 \pm 17,21$, MSE $1.894,99 \pm 2.420,42$, R^2 $0,96 \pm 0,04$. *Walk-forward backtest* long-only (biaya 0,1%) memberi Final Value 1,04 (vs 0,72 *buy-and-hold*) dengan volatilitas/*drawdown* lebih rendah. Pendekatan ini terinterpretasi, reproduibel, dan siap dikembangkan (reduksi fitur, regularisasi, fitur berbasis return/volatilitas).

Kata Kunci - regresi linier berganda (OLS), machine learning, prediksi harga saham, Bank Neo Commerce, Python.

I. PENDAHULUAN

Perkembangan teknologi informasi yang semakin pesat mendorong kebutuhan akan teknik pengolahan data yang efisien agar dapat menghasilkan informasi yang bermanfaat. Salah satu teknologi yang berkembang adalah machine learning, yang memungkinkan komputer mempelajari pola dari data sehingga dapat digunakan untuk melakukan prediksi. Dengan hadirnya teknologi digital, kegiatan berinvestasi menjadi lebih mudah karena masyarakat dapat berinvestasi melalui aplikasi yang disediakan perusahaan efek di bawah naungan Bursa Efek Indonesia (BEI). Hal ini meningkatkan partisipasi investor ritel, namun juga menimbulkan tantangan karena harga saham sering berfluktuasi secara signifikan [8]. Saham merupakan salah satu instrumen investasi yang populer. Pemilik saham berarti memiliki sebagian kepemilikan perusahaan (emiten). Namun, fluktuasi harga saham membuat prediksi pergerakan harga menjadi sulit. Harga dipengaruhi oleh permintaan dan penawaran pasar, kondisi ekonomi, serta faktor eksternal lainnya. Oleh karena itu, dibutuhkan metode prediktif yang dapat membantu meminimalisasi risiko. Salah satu pendekatan yang potensial adalah penerapan machine learning untuk prediksi harga saham [14], [9].

Machine learning bekerja dengan memanfaatkan data historis untuk menemukan pola dan membangun model prediksi. Secara umum terdapat tiga metode pembelajaran, yaitu supervised learning, unsupervised learning, dan reinforcement learning. Dalam penelitian ini digunakan pendekatan supervised learning, di mana model dilatih menggunakan data yang sudah dilabeli. Salah satu algoritma yang banyak digunakan adalah regresi linier, karena mampu menghitung hubungan antar variabel independen dengan variabel dependen secara jelas dan terukur [4]. Beberapa penelitian sebelumnya telah menunjukkan keberhasilan regresi linier dalam berbagai kasus prediksi. Himawan dkk. [5] memprediksi harga saham BRI menggunakan regresi linier dan memperoleh nilai RMSE 27.780. Putra dkk. [13] menggunakan regresi linier untuk memprediksi harga saham dengan hasil MAPE 13,75%. Nuriyah dkk. [10] memanfaatkan regresi linier berganda untuk prediksi harga saham BCA dengan hasil MAPE 0,55%. Sementara itu, Pebrelia [11] menerapkan regresi linier berganda untuk analisis curah hujan dan menghasilkan MSE 14,02. Hasil-hasil tersebut menunjukkan potensi besar regresi linier, namun penelitian khusus untuk saham perbankan digital seperti Bank Neo Commerce (BBYB) masih jarang dilakukan.

Berdasarkan kondisi tersebut, penelitian ini bertujuan membangun model prediksi harga saham Bank Neo Commerce (BBYB) menggunakan algoritma regresi linier berganda berbasis machine learning. Data historis diperoleh dari Yahoo Finance (kode saham BBYB.JK) dengan periode 12 Maret 2019 hingga 12 Maret 2025. Variabel dependen yang diprediksi adalah Close (harga penutupan saham), sedangkan variabel independen berasal dari data historis berupa *lagged features* (misalnya Close $t-1$, Close $t-5$, Volume $t-1$) serta indikator teknikal (Simple Moving Average/SMA, Exponential Moving Average/EMA, dan Relative Strength Index/RSI) yang dihitung berdasarkan data penutupan hari sebelumnya. Dataset kemudian dibagi menggunakan pendekatan *time series split* untuk mengevaluasi performa model secara lebih valid. Kontribusi penelitian ini adalah menyediakan model prediksi saham BBYB yang akurat, yang diharapkan dapat membantu investor maupun pelaku pasar modal dalam mengambil keputusan investasi yang lebih tepat dan terinformasi.

II. SIGNIFIKANSI STUDI

2.1 Studi Literatur

Machine learning merupakan salah satu cabang dari kecerdasan buatan (*artificial intelligence*) yang berfokus pada kemampuan komputer untuk belajar dari data dan memperbaiki performanya secara otomatis tanpa arahan eksplisit [15]. Inti dari machine learning adalah membangun model yang mampu menemukan pola data, sehingga dapat digunakan untuk tugas klasifikasi maupun prediksi. Secara umum, machine learning terbagi menjadi tiga jenis, yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning* [16]. Dalam penelitian ini digunakan pendekatan *supervised learning* dengan algoritma regresi linier berganda karena sesuai untuk memodelkan hubungan variabel dalam prediksi harga saham.

Prediksi (*prediction*) adalah proses memperkirakan nilai data di masa depan berdasarkan pola dari data historis. Dalam konteks pasar saham, prediksi berperan penting karena harga saham dipengaruhi oleh banyak faktor dan pergerakannya sulit ditebak secara pasti. Oleh karena itu, dibutuhkan metode analisis yang mampu memanfaatkan data historis untuk menghasilkan estimasi yang mendekati kondisi aktual [15]. Salah satu metode prediksi yang sering digunakan adalah regresi linier berganda. Metode ini bertujuan untuk memodelkan hubungan antara variabel independen (misalnya harga pembukaan, tertinggi, terendah, dan volume perdagangan) dengan variabel dependen (harga penutupan).

Bank Neo Commerce (BBYB) merupakan salah satu bank digital di Indonesia yang sebelumnya bernama Bank Yudha Bhakti. Transformasi menjadi bank digital dilakukan pada tahun 2020, dan sejak itu BBYB mencatat pertumbuhan pengguna yang pesat. Pada tahun 2021, BBYB menjadi bank digital kedua dengan jumlah pengguna terbanyak, mencapai 2,3 juta nasabah. Fenomena menarik terjadi ketika harga saham BBYB mengalami kenaikan signifikan hingga lebih dari 700% sepanjang tahun 2021, meskipun laporan keuangan perusahaan pada periode yang sama justru menunjukkan kerugian. Kondisi ini menimbulkan kontradiksi antara performa fundamental perusahaan dengan harga saham di pasar, sehingga menarik untuk diteliti lebih lanjut menggunakan pendekatan machine learning.

Beberapa penelitian terdahulu telah menerapkan algoritma regresi linier dalam berbagai konteks. Himawan dkk. [5] menggunakan regresi linier untuk prediksi harga saham BRI dan memperoleh nilai RMSE 27.780. Putra dkk. [13] juga meneliti saham BRI dengan regresi linier dan menghasilkan MAPE 13,75%. Penelitian lain oleh Nuriyah dkk. [10] menerapkan regresi linier berganda untuk prediksi harga saham BCA dan memperoleh MAPE 0,55%, menunjukkan tingkat akurasi yang tinggi. Selain pada saham, Hasibuan dkk. [4] memprediksi harga mobil bekas dengan tingkat akurasi 76,22%, sementara Pebralia [11] menggunakan regresi linier berganda untuk analisis curah hujan dengan MSE 14,02 dan MAE 2,23. Di sisi lain, beberapa penelitian membandingkan regresi linier dengan metode lain. Maulana dan Kumalasari [9] membandingkan regresi linier dengan *neural network*, SVM, *Gaussian Process*, dan regresi polinomial untuk prediksi harga saham GGRM, sedangkan Mahendra dkk. [8] membandingkan regresi linier dengan *Artificial Neural Network* (ANN) dan menunjukkan bahwa ANN dapat menghasilkan error lebih kecil pada interval tertentu. Riyandi dkk. [14] menggunakan regresi linier untuk prediksi harga saham dengan RMSE 2.731, sedangkan Sidik dan Ansawarman [16] memanfaatkan regresi linier untuk prediksi jumlah kendaraan bermotor dengan nilai R^2 sebesar 0,98, menandakan hasil prediksi sangat sesuai dengan data aktual.

Penelitian lebih mutakhir juga menunjukkan arah perkembangan yang menarik. Kulkarni dkk. [17] menegaskan bahwa performa model prediksi harga saham sangat bergantung pada konteks dataset, sehingga metode non-linier seperti *deep learning* perlu dipertimbangkan. Saberironaghi dkk. [18] dalam kajian pustakanya menyebutkan LSTM, CNN, dan SVM sebagai metode populer dengan akurasi tinggi pada data saham. Guennioui dkk. [19] mengembangkan model hibrida LightGBM–XGBoost yang terbukti efektif dalam strategi investasi praktis, sedangkan Szydlowski dan Chudziak [20] memperkenalkan arsitektur Transformer (Hidformer) yang unggul untuk data runtun waktu kompleks. Lebih lanjut, Dave dkk. [21] membuktikan bahwa pendekatan multi-agent berbasis LSTM, GRU, dan *attention* dapat meningkatkan akurasi prediksi pasar saham. Secara umum, penelitian-penelitian tersebut menunjukkan bahwa regresi linier, khususnya regresi linier berganda, mampu memberikan hasil prediksi yang cukup akurat baik pada saham maupun bidang lain. Namun, penelitian terkait prediksi harga saham pada sektor perbankan digital seperti Bank Neo Commerce (BBYB) masih jarang dilakukan. Hal ini menjadi celah penelitian yang diisi dalam studi ini.

2.2 DATA

Data utama dalam penelitian ini diperoleh dari Yahoo Finance dengan kode saham BBYB.JK yang merepresentasikan saham Bank Neo Commerce (BBYB). Dataset memiliki frekuensi harian dan mencakup periode 12 Maret 2019 hingga 12 Maret 2025. Atribut awal yang tersedia adalah Date, Open, High, Low, Adjusted Close, dan Volume.

Variabel yang digunakan adalah:

1. Variabel dependen (target): Adjusted Close (harga penutupan tersesuaian).
2. Variabel independen (prediktor): fitur yang diturunkan dari data historis dan dihitung hingga hari sebelumnya ($t-1$), meliputi lagged features *AdjustedClose_{t1}*, *AdjustedClose_{t5}*, *Volume_{t1}* serta indikator teknikal *SMA5_{t1}*, *EMA5_{t1}*, *RSI14_{t1}*.

Atribut Date digunakan sebagai penanda waktu (indeks deret waktu) dan tidak dimasukkan langsung ke dalam pemodelan. Seluruh waktu dinyatakan dalam Asia/Jakarta (UTC+7).

2.3 SUMBER DAN PERIODE DATA

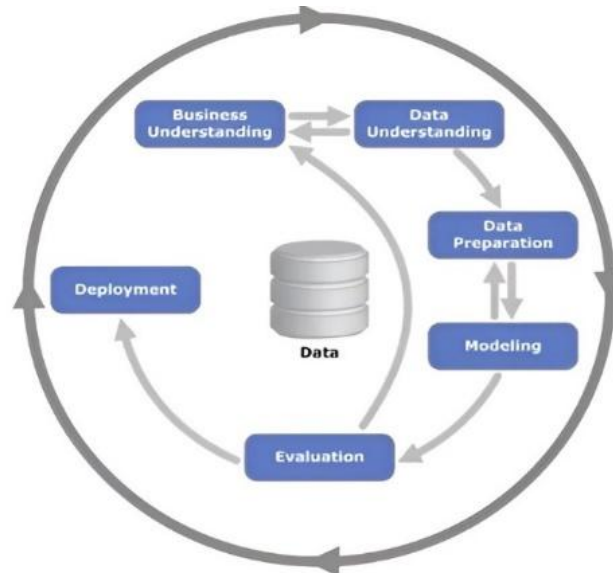
Penelitian berfokus pada saham BBYB yang diperdagangkan di Bursa Efek Indonesia dengan ticker BBYB.JK. Seluruh data harga dan volume diunduh dari Yahoo Finance, berfrekuensi harian, mencakup periode 12 Maret 2019 – 12 Maret 2025, dan dinyatakan dalam zona waktu Asia/Jakarta (UTC+7). Harga yang digunakan dalam analisis adalah Adjusted Close, sehingga aksi korporasi (dividen/stock split) telah terefleksi pada seri harga yang dianalisis. Variabel turunan berupa lagged features dan indikator teknikal dihitung sampai $t-1$ untuk mencegah *data leakage* sebelum digunakan pada pemodelan regresi linier berganda.

2.4 METODE PENELITIAN DAN EVALUASI

Metode penelitian ini akan dilakukan melalui beberapa tahapan sebagai berikut :

- **Identifikasi dan Perumusan Masalah** Tahap awal penelitian adalah mengidentifikasi permasalahan fluktuasi harga saham Bank Neo Commerce (BBYB) yang sulit diprediksi karena dipengaruhi oleh kondisi ekonomi, volume perdagangan, dan sentimen investor. Berdasarkan identifikasi tersebut, tujuan penelitian dirumuskan untuk membangun model prediksi harga saham BBYB menggunakan algoritma regresi linier berganda berbasis *machine learning* sebagai alat bantu pengambilan keputusan.

- **CRISP-DM (Cross Industry Standard Process for Data Mining)** Metodologi penelitian ini mengadopsi kerangka kerja CRISP-DM yang terdiri atas enam tahapan utama, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Kerangka ini dipilih karena memberikan struktur penelitian yang sistematis dan memungkinkan evaluasi model secara berulang.



Gambar 1. CRISP-DM (Cross Industry Standard Process for Data Mining)

- **Business Understanding** Tahap ini berfokus pada pemahaman konteks penelitian dan tujuan yang ingin dicapai. Permasalahan yang diangkat adalah sulitnya memprediksi harga saham Bank Neo Commerce (BBYB) akibat fluktuasi pasar yang dipengaruhi berbagai faktor. Tujuan penelitian ini adalah membangun model prediksi harga saham BBYB menggunakan algoritma regresi linier berganda, dengan kriteria keberhasilan yang diukur berdasarkan nilai *Mean Squared Error (MSE)*, *Mean Absolute Error (MAE)*, dan koefisien determinasi (R^2).
- **Data Understanding** Tahap ini mencakup pengumpulan serta eksplorasi awal (*exploratory data analysis/EDA*) atas data historis BBYB.JK dari Yahoo Finance berfrekuensi harian pada periode 12 Maret 2019 – 12 Maret 2025, dengan zona waktu Asia/Jakarta (UTC+7). Atribut awal yang tersedia ialah Date, Open, High, Low, Adjusted Close, dan Volume; Date digunakan sebagai indeks deret waktu.
- **Data Preparation**, Tahap ini mempersiapkan data mentah agar siap dianalisis dengan menata indeks waktu (Asia/Jakarta, UTC+7), memastikan kualitas data (mengatasi nilai hilang/duplikasi), menggunakan Adjusted Close agar efek dividen/split sudah tercermin, dan mencegah kebocoran dengan hanya membentuk prediktor dari informasi hingga $t-1$ (lag dan indikator teknikal), sementara nilai Open/High/Low/Close pada hari yang sama tidak digunakan sebagai prediktor. Baris *warm-up* yang belum memiliki fitur lengkap dibuang. Pemisahan data dilakukan berdasarkan waktu menggunakan *TimeSeriesSplit*, dan setiap transformasi yang diperlukan di-fit hanya pada data latih lalu diterapkan ke validasi/pengujian, sehingga pelatihan dan evaluasi berlangsung terpisah dan valid untuk prediksi *one-step-ahead*.
- **Modeling** Model utama adalah Regresi Linier Berganda (OLS) untuk memprediksi Adjusted Close harian satu langkah ke depan. Seluruh prediktor hanya memakai informasi sampai $t-1$ (lag & indikator teknikal) agar tidak terjadi data leakage; nilai Open/High/Low/Close pada hari yang sama tidak digunakan sebagai prediktor.

$$y_t = \alpha + \sum_{k=1}^p \beta_k x_{k,t} + \varepsilon_t \quad (1)$$

$$\hat{y}_t = \alpha + \beta_1 \text{AdjClose}_{t-1} + \beta_2 \text{AdjClose}_{t-5} + \beta_3 \text{Vol}_{t-1} + \beta_4 \text{SMA5}_{t-1} + \beta_5 \text{EMA5}_{t-1} + \beta_6 \text{RSI14}_{t-1} \quad (2)$$

$\hat{y}_t$ = prediksi Adjusted Close pada hari t (satuan rupiah).

α (intercept) = nilai dasar prediksi.

β_k (koefisien) = perubahan pada $\hat{y}$ untuk setiap +1 unit fitur ke-k ($\beta_k > 0$ naikan prediksi; $\beta_k < 0$ turunkan).

AdjClose_{t-1} , AdjClose_{t-5} = harga tersesuaikan hari ke-(t-1) dan (t-5)

Vol_{t-1} = volume perdagangan hari ke-(t-1)

SMA5_{t-1} , EMA5_{t-1} = rata-rata bergerak 5-hari (sederhana & eksponensial) hingga t-1

RSI14_{t-1} = momentum 14-hari hingga t-1 (overbought/oversold).

- **Evaluation** Tahap evaluasi model bertujuan untuk menilai kinerja prediksi dan mengukur seberapa baik model mampu menyesuaikan diri dengan data. Pada penelitian ini, evaluasi dilakukan menggunakan tiga metrik utama, yaitu Mean Squared Error (MSE), Mean Absolute Error (MAE), dan koefisien determinasi (R^2).

MSE adalah rata-rata dari kuadrat selisih antara nilai aktual dan nilai prediksi:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

MAE adalah mengukur rata-rata selisih absolut antara nilai aktual dengan nilai prediksi:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

R-squared (atau koefisien determinasi) menunjukkan proporsi variasi dalam variabel dependen yang dapat dijelaskan oleh model:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

Di mana:

dengan keterangan:

n = jumlah data pengujian,

y_i = nilai aktual harga penutupan,

$\hat{y}_i$ = nilai prediksi harga penutupan,

$\bar{y}_i$ = rata-rata nilai aktual.

Evaluasi dilakukan dengan pendekatan *time series split* menggunakan skema *rolling window*, sehingga performa model dilaporkan dalam bentuk rata-rata dan simpangan baku dari beberapa fold pengujian.

- **Deployment**, Hasil model disajikan dalam bentuk tabel dan grafik yang membandingkan harga penutupan aktual dengan nilai prediksi. Implementasi dilakukan pada Jupyter Notebook menggunakan bahasa pemrograman Python dengan pustaka pandas, scikit-learn, dan matplotlib.

III. HASIL DAN PEMBAHASAN

3.1 DATA SELECTION

Data yang dianalisis bersumber dari Yahoo Finance (ticker BBYB.JK) berfrekuensi harian pada periode 12 Maret 2019–12 Maret 2025 dengan zona waktu Asia/Jakarta (UTC+7). Target prediksi ditetapkan Adjusted Close agar dampak aksi korporasi (dividen/split) sudah terefleksi. Seluruh prediktor diturunkan dari informasi hingga $t-1$ untuk mencegah *data leakage*: AdjClose $t-1$, AdjClose $t-5$, Vol $t-1$, SMA5 $t-1$, EMA5 $t-1$, dan RSI14 $t-1$; sementara nilai Open/High/Low pada hari yang sama tidak digunakan sebagai prediktor.

Prapemrosesan mencakup pembersihan duplikasi, penanganan nilai hilang, pemeriksaan outlier pada return, serta pembuangan baris *warm-up* akibat perhitungan lag/indikator. Gambar 2 menampilkan cuplikan matriks fitur–target yang digunakan dalam seluruh eksperimen dan evaluasi *walk-forward*, sehingga pemodelan one-step-ahead terhadap y_t =Adjusted Closet dapat dinilai secara adil dengan TimeSeriesSplit.

	AdjClose	AdjClose_t1	AdjClose_t5	Vol_t1	SMA_5_t1	EMA_5_t1	RSI_14_t1
Date							
2019-04-02 07:00:00	250.82	252.63	256.24	583,908.00	252.99	252.73	61.70
2019-04-03 07:00:00	250.82	250.82	252.63	42,727.00	251.91	252.09	59.57
2019-04-04 07:00:00	249.02	250.82	252.63	0.00	251.54	251.67	58.70
2019-04-05 07:00:00	247.21	249.02	250.82	19,703.00	250.82	250.79	42.86
2019-04-08 07:00:00	243.60	247.21	252.63	32,101.00	250.10	249.60	25.00

Gambar 2. Cuplikan subset data hasil seleksi (BBYB.JK)

3.2 Kinerja Model Utama (Regresi Linier Berganda)

Ringkasan kinerja rata-rata lintas lima *fold* pada target Adjusted Close (1-day-ahead) ditunjukkan pada gambar 3. Regresi Linier (OLS) mencapai $MAE = 25,15 \pm 18,24$, $MSE = 2.344,41 \pm 2.956,78$, dan $R^2 = 0,94 \pm 0,06$. Sebagai tolak ukur minimal, baseline naïf ($\hat{y}_t = y_{t-1}$) menghasilkan $MAE = 22,40 \pm 17,21$, $MSE = 1.894,99 \pm 2.420,42$, dan $R^2 = 0,96 \pm 0,04$. Hasil ini menunjukkan bahwa pada horizon harian baseline sederhana sangat kompetitif, sementara OLS tetap menjadi model utama studi karena alur pemodelan bebas *data leakage* dan tervalidasi waktu.

	Model	MAE(mean)	MAE(sd)	MSE(mean)	MSE(sd)	R2(mean)	R2(sd)
0	Baseline_Naive(AdjClose_t-1)	22.40	17.21	1,894.99	2,420.42	0.96	0.04
1	Linear(OLS)	25.15	18.24	2,344.41	2,956.78	0.94	0.06

Gambar 3. Hasil kinerja model OLS

3.3 Perbandingan Model

Evaluasi model dilakukan dengan validasi TimeSeriesSplit rolling 5-fold, menargetkan Adjusted Close, dan menggunakan fitur t-1. pada gambar 4 menunjukkan Model baseline naïf memberikan hasil terbaik dengan MAE 22,40; MSE 1.894,99; dan R^2 0,96. Sementara itu, model Lasso, Ridge, dan OLS memiliki performa yang serupa (MAE 24,6–25,3; MSE 2.200–2.344; R^2 0,94). Model Random Forest dan SVR (RBF) menunjukkan performa yang jauh lebih buruk. Hasil ini menunjukkan bahwa pada horizon prediksi 1 hari, baseline naïf sangat kuat, dan model lain belum mampu melampauinya, kemungkinan disebabkan oleh multikolinearitas fitur dan sifat pasar harian.

	MSE (sd)	R2 (mean)	R2 (sd)
0	2,420.42	0.96	0.04
1	2,887.56	0.93	0.07
2	372,638.12	0.32	0.66
3	351,239.84	-47.04	88.23
4	459,807.50	-30.82	55.50
5	128,567,645,340,994.25	-10,873,312,705.11	13,697,616,054.48

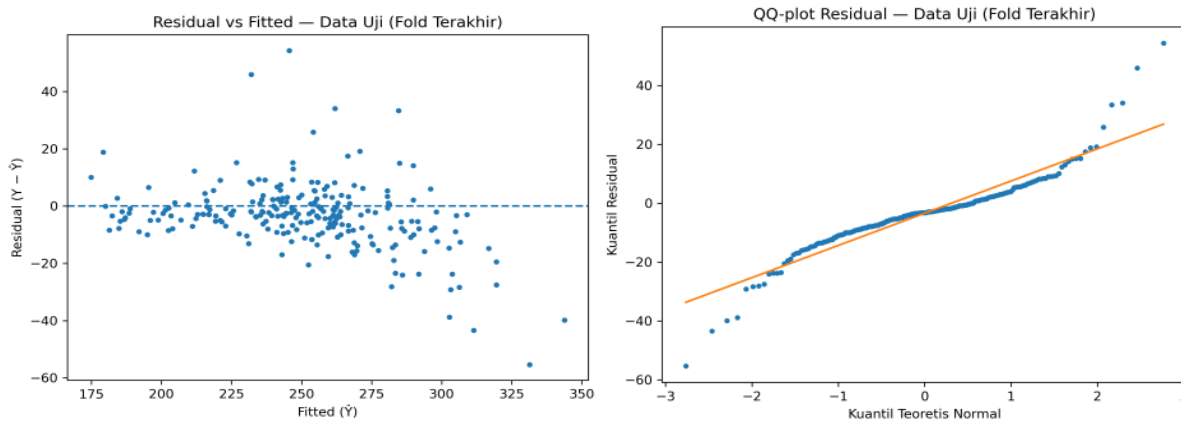
Gambar 4. Perbandingan kinerja model untuk prediksi Adjusted Close t+1 BBYB.JK

3.4 Diagnostik dan Validitas Model

Diagnostik OLS pada data uji menunjukkan multikolinearitas sangat kuat pada fitur level/rata-rata bergerak (Gambar 5; mis. VIF: $SMA_{t-1}=637,52$, $EMA_{t-1}=282,54$, $AdjClose_{t-1}=244,08$), sementara fitur lain rendah ($\approx 1,30$ – $1,66$). Uji Breusch Pagan signifikan ($p < 0,001$) sehingga varians error tidak konstan, dan Durbin Watson = 2.552 tidak menunjukkan autokorelasi positif yang berarti. Pola kipas pada Residual vs Fitted (Gambar 6(a)) konsisten dengan heteroskedastisitas, sedangkan QQ-plot (Gambar 6(b)) memperlihatkan penyimpangan ekor dari garis acuan normal. Implikasinya, koefisien OLS kurang stabil untuk inferensi. mitigasi yang disarankan meliputi reduksi fitur redundan (pilih salah satu SMA/EMA atau beralih ke fitur return/volatilitas), regulerisasi (Ridge/Lasso) dengan penskalaan prediktor, serta robust standard errors pada uji koefisien.

```
VIF:
SMA5_lag1      637.52
EMA12_lag1     282.54
Close_lag1     244.08
ret_lag1        1.66
RSI14_lag1      1.57
Vol_lag1        1.30
Name: VIF, dtype: float64
Breusch-Pagan p-value: 0.0000
Durbin-Watson: 2.552
```

Gambar 5. Diagnostik OLS: VIF, p-value Breusch–Pagan, dan statistik Durbin–Watson.



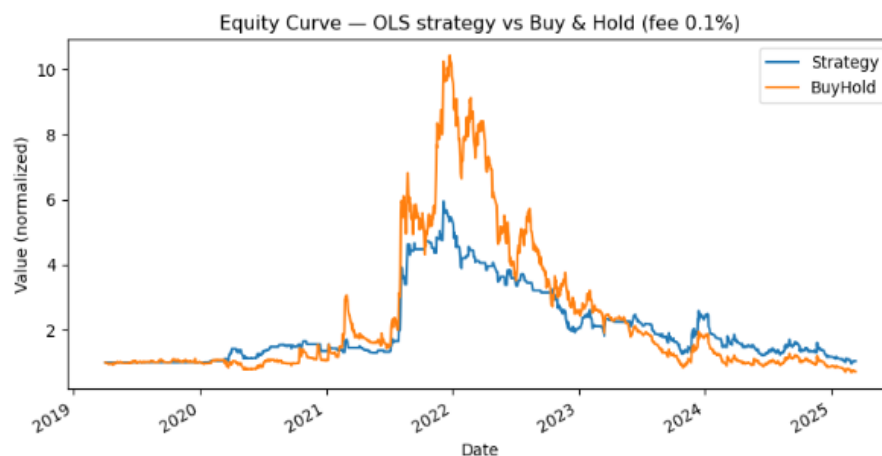
(a)

(b)

Gambar 6. (a). Residual vs Fitted (b) QQ-plot residual

3.5 Backtest Strategi

Aturan long-only jika $\hat{y}_{t+1} > y_t$ menghasilkan Final Value 1,04 ($\approx +4\%$) dengan Max Drawdown 84% dan Volatilitas 0,53, sedangkan buy-and-hold berakhir di 0,72 ($\approx -28\%$) dengan Max Drawdown 93% dan Volatilitas 0,74. Kurva ekuitas menunjukkan strategi OLS cenderung mengurangi penurunan besar pascahype 2021, meskipun tidak menyalip kenaikan ekstrem saat pasar bullish. Secara praktis, model berguna untuk manajemen risiko (eksposur lebih rendah), namun keunggulan absolut terhadap baseline persistensi belum konsisten.



Gambar 7. Equity curve strategi OLS long-only vs buy-and-hold (biaya 0,1%).

```

Model backtest: Linear(OLS)
      Final Value  Max Drawdown  Volatility
Strategy         2.40         -0.77         0.54
BuyHold           0.72         -0.93         0.74

```

Gambar 8. Ringkasan *backtest*

KESIMPULAN

Penelitian ini memprediksi Adjusted Close BBYB satu hari ke depan menggunakan fitur $t-1$ (lag & indikator teknikal) dan validasi time-series split *rolling* 5-fold yang bebas kebocoran. Model utama Regresi Linier (OLS) mencapai MSE $2,344.41 \pm 2,956.78$, MAE 25.15 ± 18.24 , dan R^2 0.94 ± 0.06 ; sedangkan baseline naif (AdjClose_{t-1}) lebih baik dengan MSE $1,894.99 \pm 2,420.42$, MAE 22.40 ± 17.21 , R^2 0.96 ± 0.04 . Ini menegaskan bahwa pada horizon harian, baseline persistensi sangat kompetitif meski pipeline OLS sudah ketat dan reproduibel. Diagnostik menunjukkan multikolinearitas sangat tinggi pada fitur rata-rata bergerak (contoh $\text{VIF} \approx 638$ untuk $\text{SMA}_{t-1} \approx 283$ untuk EMA_{t-1} serta $\text{AdjClose}_{t-1} \approx 244$, adanya heteroskedastisitas (Breusch–Pagan $p < 0.001$), dan Durbin–Watson = 2.552 (tidak ada autokorelasi positif kuat). *Walk-forward backtest* long-only (biaya 0.1%) menghasilkan Final Value 1.04, Max Drawdown 84%, Volatilitas 0.53; sedangkan buy and hold: 0.72, 93%, 0.74 menunjukkan penurunan volatilitas dan drawdown pada strategi OLS meski keunggulan kumulatif terbatas. Ke depan, perbaikan diarahkan pada reduksi fitur redundan (pilih salah satu SMA/EMA atau beralih ke fitur return/volatilitas), regularisasi (Ridge/Lasso) dengan penskalaan, evaluasi horizon >1 hari, dan uji stabilitas antar-rezim pasar.

REFERENSI

- [1] B. N. Commerce, “Tentang Kami,” n.d.
- [2] D. A. Dzulhijjah, M. B. Herlambang, M. Haifan, P. Studi, P. Insinyur, dan T. Selatan, “Implementasi Framework CRISP-DM untuk Proses Data Mining Aplikasi Credit Scoring PT. XYZ,” *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis*, vol. 1, no. 1, pp. 238–251, 2023.
- [3] J. Hao dan T. K. Ho, “Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language,” *Journal of Educational and Behavioral Statistics*, vol. 44, no. 3, pp. 348–361, 2019, doi: 10.3102/1076998619832248.
- [4] E. Hasibuan, S. Informasi, F. Ilmu Informasi, U. Gunadarma, J. Margonda, R. No, P. Cina, dan D. Jawa, “Implementasi Machine Learning untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear berbasis Web,” *Jurnal Ilmiah Komputasi*, vol. 21, no. 4, pp. 595–602, 2022, doi: 10.32409/jikstik.21.4.3327.
- [5] I. Himawan, O. Nurdiawan, G. Dwilestari, et al., “Prediksi Harga Saham dengan Algoritma Regresi Linier dengan Rapidminer,” *JURSIMA*, vol. 10, no. 3, pp. 239–247, 2022. [Online]. Available: <https://ejournal.indobarunasional.ac.id/index.php/jursima/article/view/475>
- [6] H. Huo, Z. Rong, O. Kononova, W. Sun, T. Botari, T. He, V. Tshitoyan, dan G. Ceder, “Semi-supervised machine-learning classification of materials synthesis procedures,” *NPJ Computational Materials*, vol. 5, no. 1, pp. 1–7, 2019, doi: 10.1038/s41524-019-0204-1.
- [7] T. Jaelani, “Machine Learning untuk Prediksi Produksi Gula Nasional,” *JMPM (Jurnal Material dan Proses Manufaktur)*, vol. 6, no. 1, pp. 31–36, 2022, doi: 10.18196/jmpm.v6i1.14897.
- [8] B. H. Mahendra, L. Chaerani, dan G. Gumay, “Analisis Perbandingan Prediksi Harga Saham menggunakan Algoritma Artificial Neural Network dan Linear Regression,” *Jurnal Ilmiah Komputasi*, vol. 22, no. 2, pp. 303–312, 2023, doi: 10.32409/jikstik.22.2.3357.

- [9] R. Maulana dan D. Kumalasari, “Analisis dan Perbandingan Algoritma Data Mining dalam Prediksi Harga Saham GGRM,” *Jurnal Informatika Kaputama (JIK)*, vol. 3, no. 1, pp. 22–28, 2019, doi: 10.59697/jik.v3i1.393.
- [10] A. S. Nuriyah, R. Saladini, K. N. Fujiah, dan F. Indrayatna, “Prediksi Harga Saham Bank BCA Menggunakan Metode Regresi Linier Berganda,” *Seminar Nasional Statistika Aktuaria II*, 2023. [Online]. Available: <https://prosidingnsa.statistics.unpad.ac.id>
- [11] J. Pebralia, “Analisis Curah Hujan Menggunakan Machine Learning Metode Regresi Linier Berganda Berbasis Python dan Jupyter Notebook,” *Jurnal Ilmu Fisika dan Pembelajarannya (JIFP)*, vol. 6, no. 2, pp. 23–30, 2022, doi: 10.19109/jifp.v6i2.13958.
- [12] K. Puteri dan A. Silvanie, “Machine Learning untuk Model Prediksi Harga Sembako,” *Jurnal Nasional Informatika*, vol. 1, no. 2, pp. 82–94, 2020.
- [13] J. S. Putra, R. D. Ramadhani, dan A. Burhanuddin, “Prediksi Harga Saham Bank BRI Menggunakan Algoritma Linear Regression sebagai Strategi Jual Beli Saham,” *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, vol. 2, no. 1, pp. 1–10, 2022, doi: 10.20895/dinda.v2i1.273.
- [14] A. Riyandi, I. Nur Ardiansyah, dan R. Dany, “Analisis Data Mining untuk Prediksi Harga Saham: Perbandingan Metode Regresi Linier dan Pola Historis,” *Jurnal Teknologi Sistem Informasi (JTSI)*, vol. 4, no. 2, pp. 278–288, 2023.
- [15] E. Setiadi dan A. Wibowo, “Uji Kelayakan Konstruksi Balok PT Daeyung dari Aplikasi SAP2000 Menggunakan Machine Learning dengan Model Linier Regression,” *Jurnal Rekayasa Sipil dan Arsitektur*, vol. 5, no. 1, pp. 799–814, 2024.
- [16] A. D. Sidik dan A. Ansawarman, “Prediksi Jumlah Kendaraan Bermotor Menggunakan Machine Learning,” *Formosa Journal of Multidisciplinary Research*, vol. 1, no. 3, pp. 559–568, 2022, doi: 10.55927/fjmr.v1i3.745.
- [17] A. Kulkarni, S. Patil, dan S. Bhosale, “A Quest for Context-Specific Stock Price Prediction: Comparing Time-Series, Machine Learning, and Deep Learning Models for BSE Stocks,” *SN Computer Science*, vol. 6, no. 2, pp. 1–15, 2025.
- [18] P. Saberironaghi, M. Zandieh, A. Yazdani, dan A. Hematian, “Stock Market Prediction Using Machine Learning and Deep Learning Techniques: A Review,” *AI*, vol. 5, no. 3, pp. 144–163, 2025.
- [19] Y. Guennioui, K. El Yacoubi, dan H. El Bouchti, “Machine Learning-Driven Stock Price Prediction for Enhanced Investment Strategy,” *Journal of Big Data and Analytics in Finance*, vol. 6, no. 4, pp. 55–66, 2024.
- [20] Ł. Szydlowski dan M. Chudziak, “Hidformer: Transformer-Style Neural Network in Stock Price Forecasting,” *arXiv preprint arXiv:2412.19932*, Dec. 2024.
- [21] R. Dave, K. Patel, dan V. Shah, “Multi-Agent Stock Prediction Systems: An Ensemble of LSTM, GRU, and Attention Mechanisms,” *arXiv preprint arXiv:2502.15853*, Feb. 2025.