

Peramalan Harga Reksa Dana di Indonesia dengan Pendekatan Linear dan Hibrida

Forecasting Mutual Fund Prices in Indonesia Using Linear and Hybrid Approaches

Gymnastiar Al Khoarizmy^{1*}, Feryadi Yulius¹, dan Elok Fiola¹

¹Program Studi Sains Data, Fakultas Sains, Institut Teknologi Sumatera, Lampung Selatan, Indonesia

*Penulis korespondensi: gymnastiar.122450096@student.itera.ac.id

Abstract

Post-pandemic, the Indonesian capital market has exhibited dynamic fluctuations, driving retail investment growth in equity mutual funds such as the Sucorinvest Equity Fund. Historically, price predictions for financial assets often default to complex deep learning architectures; however, whether non-linear deep learning provides statistically significant advantages on persistent financial series dominated by strong autocorrelation remains an open question. This study rigorously evaluates the efficiency of linear regularized regression (ElasticNet) enhanced with autoregressive lag features and Principal Component Analysis (PCA) against a hybrid ElasticNet-LSTM residual architecture. The baseline ElasticNet model achieved remarkable forecasting accuracy ($R^2 = 0,9672$, RMSE = 25,3143), predominantly driven by lag-1 price inertia. In contrast, incorporating an LSTM model to capture residual non-linear patterns yielded an identical R^2 of 0,9672 and an RMSE of 25,3052. The Diebold-Mariano test ($p = 0,2659$) confirmed that the hybrid model's improvement was not statistically significant, indicating that residual errors closely resemble stochastic white noise. Consequently, this study substantiates Occam's razor in financial time-series forecasting: for high-persistence daily price levels, a parsimonious and interpretable linear approach offers superior practical efficiency and computational economy over complex deep learning architectures.

Keywords: Autocorrelation, Diebold-Mariano Test, ElasticNet, Hybrid Model, LSTM, Mutual Funds, Occam's Razor, Time-Series Forecasting

Abstrak

Pasca pandemi COVID-19, pasar modal Indonesia menunjukkan fluktuasi yang dinamis, mendorong peningkatan minat investasi pada reksa dana saham seperti Sucorinvest Equity Fund. Prediksi harga aset finansial sering kali langsung mengandalkan arsitektur *deep learning* yang kompleks; namun, efektivitas penambahan kompleksitas non-linear pada data yang didominasi oleh autokorelasi tinggi masih memerlukan pembuktian kritis. Penelitian ini bertujuan mengevaluasi efisiensi pendekatan regresi linear teregularisasi (ElasticNet) yang diperkuat dengan fitur kelambatan waktu (*lag*) dan Principal Component Analysis (PCA) dibandingkan dengan model hibrida ElasticNet-LSTM yang memodelkan sisaan (*residual*). Hasil evaluasi menunjukkan bahwa model linear dasar mampu mencapai akurasi peramalan yang sangat tinggi ($R^2 = 0,9672$, RMSE = 25,3143) yang didominasi oleh inersia harga lag-1. Sementara itu, penambahan model LSTM pada residual hanya menghasilkan RMSE 25,3052 dengan R^2 identik (0,9672). Uji Diebold-Mariano menghasilkan nilai $p = 0,2659$, membuktikan bahwa penambahan LSTM tidak memberikan perbaikan kinerja yang signifikan secara statistik karena sisaan telah mendekati derau putih (*white noise*). Studi ini menegaskan validitas prinsip Occam's razor dalam peramalan deret waktu finansial: untuk horizon prediksi harian berpersistensi tinggi, pendekatan linear yang sederhana, hemat komputasi, dan mudah diinterpretasikan terbukti jauh lebih efisien dan praktis dibandingkan arsitektur *deep learning* yang kompleks.

Kata Kunci: Autokorelasi, ElasticNet, LSTM, Model Hibrida, Occam's Razor, Peramalan Deret Waktu, Reksa Dana, Uji Diebold-Mariano

PENDAHULUAN

Pasar modal memegang peranan krusial sebagai sumber mobilisasi tabungan dan penyediaan kredit jangka panjang yang berkontribusi pada pertumbuhan ekonomi di Indonesia [1]. Dalam ekosistem ini, reksa dana telah menjadi salah satu instrumen investasi yang semakin populer di kalangan masyarakat, menawarkan keunggulan diversifikasi portofolio dan manajemen profesional yang didukung oleh regulasi Otoritas Jasa Keuangan (OJK) [2]. Reksa dana berfungsi sebagai wahana investasi yang mengagregasi dana dari banyak investor, mengarahkannya ke dalam berbagai kelas aset seperti saham, obligasi, dan instrumen pasar uang, sehingga menjadi pilihan menarik bagi masyarakat yang ingin berpartisipasi di pasar modal secara terukur.

Di tengah dinamika pasar, terutama pada periode pasca-pandemi COVID-19 (2022–2025), kinerja reksa dana saham sangat dipengaruhi oleh berbagai faktor makroekonomi

dan sentimen sektoral. Studi empiris menunjukkan bahwa variabel makroekonomi seperti laju inflasi dan suku bunga acuan Bank Indonesia (BI Rate) memiliki dampak signifikan terhadap pergerakan Nilai Aktiva Bersih (NAB) per unit penyertaan [3]. Selain itu, fluktuasi nilai tukar Rupiah terhadap Dolar AS (kurs JISDOR) dan indeks saham sektoral terbukti menjadi penentu utama pergerakan aset dasar [4]. Kondisi pasar yang volatil ini menciptakan tantangan analitis yang besar bagi manajer investasi dan investor ritel dalam memprediksi valuasi reksa dana secara akurat demi memitigasi risiko penurunan pasar (*drawdown*).

Seiring dengan perkembangan teknologi komputasi, penerapan pembelajaran mesin (*machine learning*) dan pembelajaran mendalam (*deep learning*) dalam analisis finansial telah berkembang pesat. Berbagai penelitian terkini telah mengeksplorasi arsitektur jaringan saraf berulang seperti Long Short-Term Memory (LSTM) dan Gated Recurrent

Unit (GRU) untuk memprediksi harga aset keuangan [5]. Namun, penerapan model *deep learning* yang sangat kompleks pada data pasar sering kali dilakukan secara langsung (*black-box*) tanpa mengevaluasi struktur dasar data secara kritis. Pada data deret waktu finansial berfrekuensi harian, karakteristik autokorelasi (ketergantungan harga pada periode sebelumnya) umumnya sangat dominan, di mana tingkat persistensi momentum sering kali menjelaskan sebagian besar varians pergerakan harga.

Menutupi kesenjangan penelitian tersebut, keunikan (*uniqueness*) dan kebaruan utama dari penelitian ini terletak pada pengujian kritis terhadap prinsip parsimoni (Occam's razor) dalam peramalan reksa dana di Indonesia. Alih-alih langsung mengasumsikan model *deep learning* pasti lebih unggul, penelitian ini membandingkan secara langsung antara model regresi linear teregularisasi yang efisien (ElasticNet) dengan arsitektur hibrida dua tahap (ElasticNet + LSTM pada residual), kemudian memvalidasi signifikansi peningkatannya menggunakan Uji Diebold-Mariano [8]. Alasan pemilihan ElasticNet sebagai model dasar adalah keunggulannya dalam menggabungkan penalti L1 (Lasso) dan L2 (Ridge), yang memungkinkan seleksi fitur otomatis sekaligus mengatasi masalah multikolinearitas antar indikator ekonomi secara stabil. Sementara itu, LSTM dipilih untuk memodelkan sisaan (*residual*) karena kemampuannya mendeteksi dependensi temporal non-linear jangka panjang. Penggunaan Principal Component Analysis (PCA) [7] pada fitur eksternal bertujuan mereduksi dimensi dan menghilangkan korelasi antar indikator, sehingga model terhindar dari kutukan dimensi (*curse of dimensionality*). Penelitian ini secara spesifik menguji hipotesis ilmiah: apakah penambahan *deep learning* non-linear pada residual memberikan peningkatan akurasi yang signifikan secara statistik setelah komponen autokorelasi linear yang dominan dimodelkan secara optimal.

METODE PENELITIAN

Pendekatan metodologi yang digunakan dalam penelitian ini dirancang secara sistematis melalui lima tahapan utama: (1) pengumpulan dan kurasi dataset deret waktu harian pasar modal dan makroekonomi pasca-pandemi; (2) pra-pemrosesan data mencakup penanganan nilai hilang, pembuatan fitur kelambatan waktu (*lag*), penormalan skala (*scaling*), dan reduksi dimensi ortogonal dengan PCA; (3) perancangan dan pelatihan model linear teregularisasi ElasticNet sebagai model dasar (*baseline*); (4) pemodelan sisaan (*residual*) menggunakan arsitektur *deep learning* LSTM untuk mendeteksi pola non-linear; dan (5) evaluasi komparatif komprehensif menggunakan metrik RMSE, MAE, R^2 , Akurasi Arah (*Directional Accuracy*), serta Uji Hipotesis Diebold-Mariano [8].

DATASET DAN PERIODE ANALISIS

Data yang digunakan dalam penelitian ini merupakan data deret waktu harian dari produk Reksa Dana Saham Sucorinvest Equity Fund (kode variabel: Sucorinvest_Close) seba-

gai variabel target, yang diintegrasikan dengan 12 variabel eksternal mencakup saham konstituen utama portofolio, indeks pasar, kurs valuta asing, dan indikator makroekonomi domestik. Pemilihan produk Sucorinvest Equity Fund didasarkan pada rekam jejaknya sebagai salah satu reksa dana saham dengan dana kelolaan (AUM) besar, likuiditas transaksi tinggi, serta sensitivitas yang kuat terhadap dinamika pasar ekuitas Indonesia pasca-pandemi. Tabel 1 menyajikan rincian lengkap mengenai definisi dan sumber dari masing-masing variabel yang digunakan.

Periode analisis yang dipilih mencakup rentang waktu 6 Januari 2022 hingga 13 Juni 2025, menghasilkan 897 observasi hari bursa setelah pembersihan data. Justifikasi pemilihan horizon waktu ini adalah untuk secara spesifik menangkap rezim pasar pasca-pandemi yang diwarnai oleh normalisasi kebijakan moneter, volatilitas inflasi global, serta penyesuaian suku bunga tinggi Bank Indonesia. Data dibagi secara kronologis (tanpa pengacakan guna menjaga integritas temporal) menjadi 80% data latih (717 observasi, periode 6 Januari 2022 hingga 19 September 2024) dan 20% data uji (180 observasi, periode 20 September 2024 hingga 13 Juni 2025), mengikuti proporsi baku pembagian data deret waktu [6].

PRA-PEMROSESAN DATA DAN REKAYASA FITUR (DATA PREPROCESSING AND FEATURE ENGINEERING)

Untuk mempersiapkan data sebelum pemodelan, dilakukan beberapa langkah rekayasa dan pra-pemrosesan fitur yang krusial.

Rekonstruksi Fitur Kelambatan Waktu (Lag Features)

Mengingat sifat data deret waktu finansial yang memiliki persistensi dan autokorelasi tinggi, dibuat tiga fitur kelambatan waktu (*lag_1*, *lag_2*, *lag_3*) dari variabel target penutupan harga reksa dana. Fitur *lag_1*, *lag_2*, dan *lag_3* merepresentasikan harga penutupan $t - 1$, $t - 2$, dan $t - 3$ hari bursa sebelumnya. Penyertaan fitur kelambatan waktu ini memungkinkan model menangkap inersia harga dan momentum jangka pendek yang merupakan sinyal prediktif terkuat pada data pasar.

Penormalan Skala (Scaling) dan Reduksi Dimensi dengan PCA

Untuk menangani perbedaan skala dan potensi multikolinearitas di antara 12 fitur eksternal, diterapkan pendekatan dalam dua tahap:

Pemisahan Fitur dan Penormalan Skala (Min-Max Scaling)

Fitur-fitur prediktor dipisahkan secara ketat menjadi dua kelompok: fitur internal kelambatan waktu (*lag*) dan 12 fitur eksternal makroekonomi/pasar. Metode penormalan skala Min-Max Scaling diterapkan khusus pada 12 fitur eksternal untuk mentransformasikan nilainya ke rentang $[0, 1]$. Langkah ini krusial untuk mencegah variabel berskala nominal besar (seperti kurs JISDOR yang bernilai belasan ribu rupiah) mendominasi proyeksi analisis komponen utama dibandingkan variabel berformat persentase (seperti laju inflasi dan BI Rate).

Tabel 1. Deskripsi variabel yang digunakan

Variabel	Deskripsi
<i>Variabel Target</i> Sucorinvest Close	Harga penutupan harian dari Reksa Dana Saham Sucorinvest, sebagai variabel y yang akan diprediksi.
<i>Fitur Internal (Dependensi Temporal)</i> lag1, lag2, lag3	Harga penutupan Sucorinvest dari 1, 2, dan 3 hari sebelumnya. Digunakan untuk menangkap momentum dan autokorelasi.
<i>Fitur Eksternal (Konteks Pasar dan Ekonomi)</i> ASII, BBRI, BBTN, BMRI, PGAS Close	Harga penutupan harian dari saham-saham utama yang relevan dengan portofolio, merepresentasikan kinerja aset dasar.
LQ45, IDX Finance Close	Indeks pasar saham utama di Indonesia, merepresentasikan sentimen pasar secara umum dan sektoral.
JISDOR Close	Kurs referensi Rupiah terhadap Dolar AS (harian), merefleksikan risiko nilai tukar dan sentimen pasar global.
Inflation Rate, BI Rate	Indikator makroekonomi utama (bulanan dan sesuai kebijakan), merefleksikan kondisi ekonomi domestik.
CPI Financial, Consumer Confidence	Indikator sentimen konsumen dan inflasi sektoral (bulanan), memberikan gambaran kondisi ekonomi dari sisi permintaan.

Principal Component Analysis (PCA) untuk Dekorelasi Fitur

PCA diterapkan pada 12 fitur eksternal yang telah dinormalisasi. Sebagai teknik statistik transformasi ortogonal linier [7], PCA bertujuan mengonversi serangkaian variabel yang saling berkorelasi menjadi variabel baru yang saling bebas linier (komponen utama/*principal components*). Berdasarkan analisis varians kumulatif, 9 komponen utama (PC1 hingga PC9) dipilih karena mampu menjelaskan 97,66% dari total varians data eksternal, sekaligus mengeliminasi multikolinearitas dan mempercepat konvergensi pelatihan *neural network*. Matriks fitur akhir yang digunakan dalam pemodelan terdiri dari gabungan 3 fitur lag asli dan 9 komponen PCA ortogonal.

Matriks fitur akhir yang digunakan untuk pemodelan terdiri dari gabungan 3 fitur lag asli dan 9 komponen PCA.

PERANCANGAN MODEL (MODEL DESIGN)

Penelitian ini membandingkan dua arsitektur pemodelan, yaitu Model A (Model Dasar ElasticNet) dan Model B (Model Hibrida ElasticNet + LSTM pada Residual). ElasticNet dipilih sebagai model dasar linier karena memadukan penalti regularisasi L1 (Lasso) dan L2 (Ridge). Formulasi fungsi objektif optimasi penalti ElasticNet disajikan pada Persamaan (1):

$$\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^n (y_i - X_i \beta)^2 + \alpha \rho \sum_{j=1}^p |\beta_j| + \frac{\alpha(1-\rho)}{2} \sum_{j=1}^p \beta_j^2 \right\} \quad (1)$$

di mana α adalah parameter intensitas regularisasi total dan ρ adalah rasio penalti L1 terhadap L2 ($l1_ratio$). Penalti L1 mendorong seleksi fitur secara tegas (*sparsity*), sedangkan penalti L2 mempertahankan stabilitas koefisien saat terdapat variabel yang berkorelasi tinggi. Penalaan hiperparameter (*hyperparameter tuning*) dilakukan secara otomatis melalui validasi silang deret waktu 5-lipat (ElasticNetCV),

yang menghasilkan nilai parameter optimal $\alpha = 0,1842$ dan $l1_ratio = 0,10$.

Model B dibangun berdasarkan paradigma dekomposisi deret waktu linier dan non-linier. Asumsi dasar pendekatan ini adalah bahwa hubungan antara variabel ekonomi dan harga aset finansial mengandung komponen linier dominan yang dapat ditangkap oleh ElasticNet, serta komponen non-linear tersembunyi yang tersisa pada sisaan (*residual*). Sisaan e_t dihitung sebagai selisih antara nilai aktual y_t dan prediksi linier $\hat{y}_{linear}$. Jika sisaan tersebut masih mengandung dependensi temporal terstruktur, maka model *deep learning* akan mampu mengekstrak pola tersebut untuk mengoreksi kesalahan prediksi linier.

Arsitektur LSTM dirancang khusus untuk memodelkan sisaan (*residual*) dengan input sekuensial berupa 9 komponen PCA selama jendela historis 15 hari bursa (*lookback window* = 15). Arsitektur jaringan saraf terdiri dari lapisan pertama LSTM dengan 50 unit (`return_sequences=True`), lapisan Dropout (0,2) untuk regularisasi, lapisan kedua LSTM dengan 25 unit, lapisan Dropout (0,2), dan diakhiri dengan lapisan Dense (1 neuron) sebagai penaksir koreksi residual. Model dilatih menggunakan optimizer Adam dan fungsi rugi Mean Squared Error (MSE). Prediksi final model hibrida diperoleh melalui penjumlahan: $\hat{y}_{final} = \hat{y}_{ElasticNet} + \hat{e}_{LSTM}$.

Alur kerja dari skema perbandingannya dapat dilihat pada Gambar 1 dan sebagai berikut:

1. Prediksi linear dibuat menggunakan model ElasticNet yang telah dilatih (Model A).
2. Residual, yaitu selisih antara harga aktual dan prediksi linear, dihitung.
3. Sebuah model Long Short-Term Memory (LSTM) dilatih untuk memprediksi residual ini. Input untuk LSTM adalah sekuens dari 9 komponen PCA, dengan panjang sekuens 15 hari.

- Prediksi akhir dari model hibrida diperoleh dengan menjumlahkan prediksi dari ElasticNet dan prediksi residual dari LSTM.

METRIK EVALUASI DAN UJI SIGNIFIKANSI STATISTIK DIEBOLD-MARIANO

Kinerja kedua model dievaluasi pada data uji (180 hari bursa) menggunakan metrik evaluasi baku: Root Mean Squared Error (RMSE) pada Persamaan (2), Mean Absolute Error (MAE) pada Persamaan (3), koefisien determinasi (R^2), serta Akurasi Arah Pergerakan (*Directional Accuracy*). Untuk membuktikan apakah selisih akurasi antara model dasar ElasticNet dan model Hibrida signifikan secara statistik atau hanya variasi acak, dilakukan Uji Hipotesis Diebold-Mariano (DM test) [8]. Uji ini menguji hipotesis nol (H_0) bahwa kedua model memiliki akurasi peramalan yang setara ($\bar{d} = 0$) berdasarkan fungsi kerugian kuadrat residual (*MSE loss differential*).

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

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

di mana y_i adalah nilai aktual, $\hat{y}_i$ adalah nilai prediksi, dan n adalah jumlah sampel. R^2 dan *Directional Accuracy* juga digunakan. Untuk validasi statistik, digunakan Uji Diebold-Mariano [8].

HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dari eksperimen komparatif, dimulai dengan analisis data eksploratif, diikuti oleh perbandingan kuantitatif kinerja model, dan diakhiri dengan validasi statistik.

ANALISIS DATA EKSPLORATIF

Sebelum pemodelan, dilakukan analisis eksploratif untuk memahami karakteristik data. Gambar 2 menampilkan pergerakan harga historis dari reksa dana Sucorinvest beserta pembagian data latih dan uji yang digunakan dalam penelitian ini. Terlihat adanya tren dan volatilitas yang bervariasi sepanjang periode analisis.

Untuk memahami hubungan antar variabel, dapat dilihat heatmap korelasi dari fitur-fitur eksternal asli yang ditampilkan pada Gambar 3. Visualisasi ini menunjukkan adanya korelasi cukup kuat di antara beberapa variabel, seperti antara indeks pasar dan harga saham sektor keuangan, yang mengindikasikan potensi multikolinearitas. Temuan ini memperkuat justifikasi penggunaan PCA sebagai teknik reduksi dimensi untuk menghasilkan representasi fitur yang lebih stabil dan bebas korelasi linier.

PRINCIPAL COMPONENT ANALYSIS (PCA)

Hasil bobot komponen utama (*loadings*) PCA pada Tabel 4 menunjukkan bahwa dimensi utama data berhasil memisahkan dinamika ekonomi secara tematis. PC1 didom-

inasi oleh CPI Financial (0,6333) dan BI Rate (0,4923), merefleksikan rezim inflasi dan kebijakan moneter domestik. PC2 mencerminkan pergerakan saham konglomerasi dan perbankan besar (ASII sebesar 0,6510 dan BBRI sebesar 0,4212). PC3 dan PC4 menangkap variasi dari saham energi dan perbankan (PGAS sebesar 0,7473 dan BMRI sebesar 0,6703). Sementara itu, PC7 didominasi oleh Inflation Rate (0,8615) dan PC9 oleh Consumer Confidence (0,7853). Dekomposisi ini membuktikan bahwa PCA berhasil mereduksi redundansi antar indikator pasar tanpa kehilangan struktur informasi esensial.

PERFORMA MODEL KOMPARATIF

Kinerja dari model ElasticNet (*baseline*) dan model Hibrida (ElasticNet + LSTM) dievaluasi pada data uji untuk menilai akurasi dan keandalannya. Tabel 3 menyajikan rangkuman metrik performa utama dalam dua skala: skala asli (Rupiah) untuk interpretasi bisnis, dan skala yang dinormalisasi [0, 1] untuk perbandingan akademis.

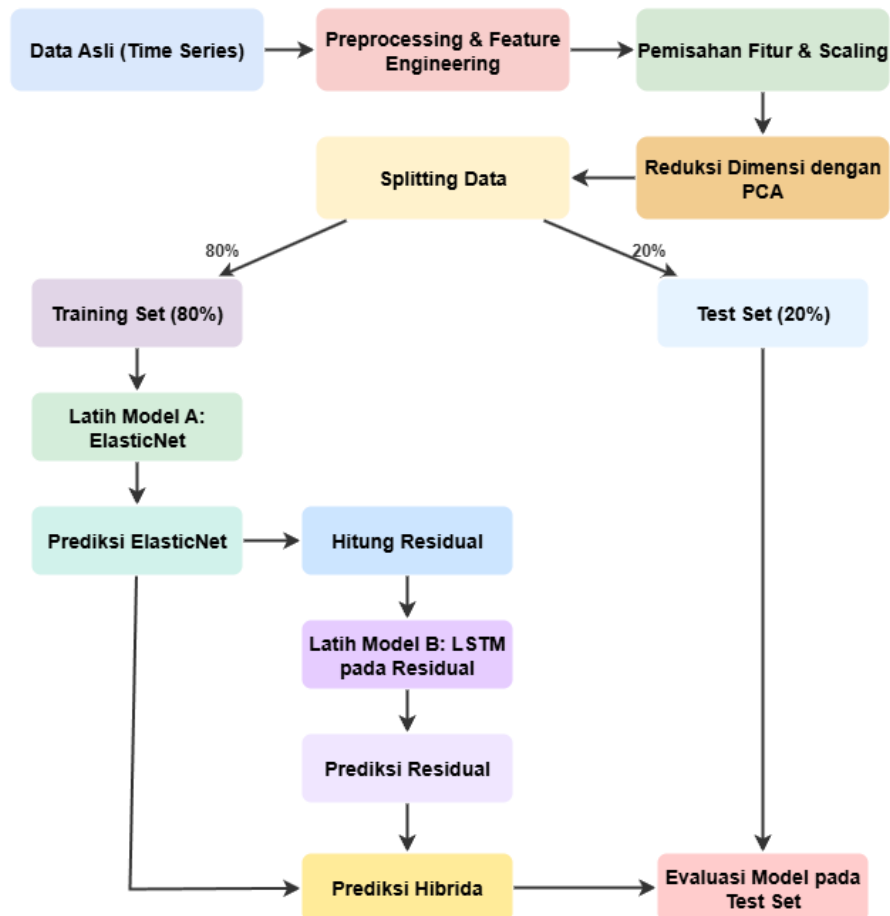
Hasil evaluasi kuantitatif pada data uji (Tabel 3) menunjukkan bahwa kedua model mencapai performa peramalan yang sangat presisi. Model dasar ElasticNet menghasilkan nilai RMSE sebesar 25,3143 dan MAE sebesar 19,0425 pada skala Rupiah asli, dengan koefisien determinasi R^2 mencapai 0,9672. Akurasi arah pergerakan harian (*Directional Accuracy*) tercatat sebesar 67,04%. Sementara itu, Model Hibrida (ElasticNet + LSTM) menghasilkan nilai RMSE sebesar 25,3052, MAE sebesar 19,0318, serta nilai R^2 identik sebesar 0,9672 dengan Akurasi Arah sebesar 67,04%. Selisih penurunan RMSE antara kedua model hanya sebesar 0,0091 poin (perbaikan marginal sebesar 0,036%).

Untuk memungkinkan perbandingan langsung dengan literatur akademis seperti studi referensi [?], kami juga menyajikan metrik *error* yang dinormalisasi. Model Hibrida kami mencapai RMSE dinormalisasi sebesar 0,0267, yang secara signifikan lebih baik daripada RMSE 0,0547 yang dilaporkan pada studi referensi tersebut. Hal ini menunjukkan bahwa pendekatan kami, terutama dengan penyertaan fitur lag, secara fundamental lebih akurat dalam menangkap dinamika harga. Visualisasi pada Gambar 4 lebih lanjut mengilustrasikan perbedaan kinerja yang sangat tipis antara kedua pendekatan yang kami uji.

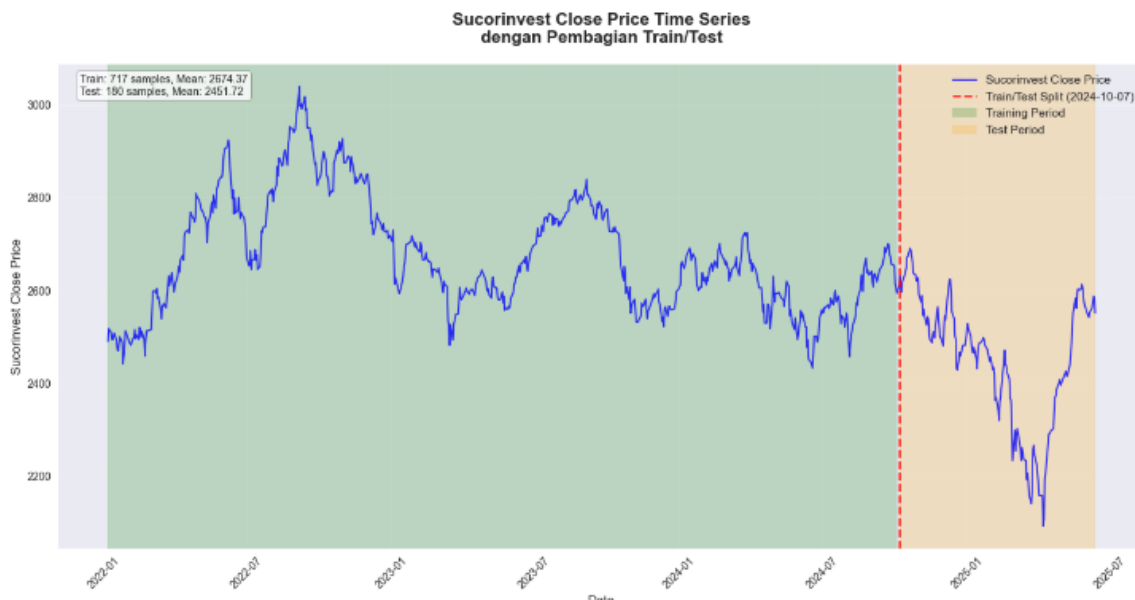
Visualisasi pada Gambar 4 menyajikan perbandingan metrik ini dalam bentuk grafik bar, yang mengilustrasikan perbedaan kinerja yang sangat tipis antara kedua pendekatan.

ANALISIS PREDIKSI VISUAL

Untuk memahami bagaimana kedua model melacak pergerakan harga aktual, dibuat plot prediksi dari kedua model terhadap nilai aktual pada data uji (Gambar 5). Secara visual, garis prediksi dari model ElasticNet dan model Hibrida hampir sepenuhnya tumpang tindih, menegaskan kembali bahwa kontribusi dari komponen LSTM pada residual sangat minimal dalam mempengaruhi prediksi akhir.



Gambar 1. Diagram alur kerja dari metodologi penelitian



Gambar 2. Grafik *time-series* harga Sucorinvest dengan pemisahan data latih dan uji

VALIDASI STATISTIK

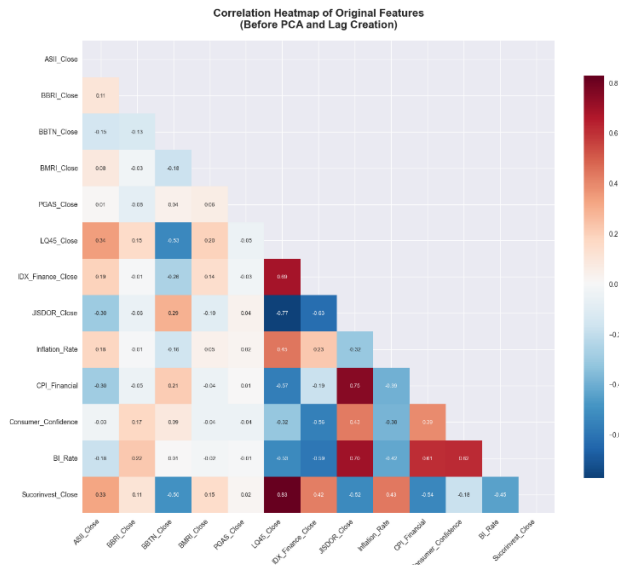
Untuk menguji apakah perbedaan kinerja yang sangat kecil antara kedua model ini signifikan secara statistik, kami melakukan Uji Diebold-Mariano.

Tabel 5 menyajikan hasil Uji Diebold-Mariano pada data uji. Dengan nilai statistik DM sebesar 1,1125 dan p-

value sebesar 0,2659 (jauh melampaui ambang signifikansi $\alpha = 0,05$), kami secara tegas gagal menolak hipotesis nol (H_0). Temuan statistik ini membuktikan bahwa penambahan arsitektur *deep learning* LSTM pada residual tidak memberikan peningkatan akurasi peramalan yang signifikan secara statistik dibandingkan model linear ElasticNet tunggal.

Tabel 2. Varians yang dijelaskan oleh setiap komponen utama (PC)

Komponen	Variance Explained	Cumulative Variance	Var. Explained (%)	Cum. Var. (%)
PC1	0,3668	0,3668	36,6836	36,6836
PC2	0,1260	0,4929	12,6029	49,2865
PC3	0,1182	0,6110	11,8164	61,1029
PC4	0,1030	0,7140	10,2967	71,3996
PC5	0,0834	0,7974	8,3430	79,7426
PC6	0,0662	0,8636	6,6206	86,3632
PC7	0,0424	0,9060	4,2382	90,6014
PC8	0,0404	0,9464	4,0369	94,6383
PC9	0,0303	0,9766	3,0252	97,6635

**Gambar 3.** Heatmap korelasi fitur-fitur asli**Tabel 3.** Perbandingan performa model pada test set

Metrik	ElasticNet	Hybrid (ElasticNet+LSTM)
RMSE	25,3143	25,3052
RMSE (Dinormalisasi)	0,0267	0,0267
MAE	19,0425	19,0318
MAE (Dinormalisasi)	0,0201	0,0201
R ²	0,9672	0,9672
Directional Accuracy	67,04%	67,04%

Perbedaan performa yang sangat tipis pada metrik *error* semata-mata merupakan fluktuasi acak.

PEMBAHASAN

Hasil penelitian ini memberikan temuan ilmiah yang sangat penting mengenai dinamika peramalan deret waktu finansial berautokorelasi tinggi. Tingginya akurasi model ElasticNet ($R^2 = 0,9672$) didominasi secara absolut oleh bobot koefisien fitur lag_1 (0,9464). Hal ini menegaskan bahwa inersia harga penutupan hari kemarin merupakan determinan paling dominan dalam penentuan harga reksa dana harian satu langkah ke depan (*one-day-ahead forecast*).

Mengapa Pola Non-Linear Tidak Muncul pada Residual LSTM?

Kegagalan komponen LSTM dalam memberikan peningkatan akurasi yang signifikan secara statistik ($p =$

0,2659) dapat dijelaskan melalui empat faktor fundamental:

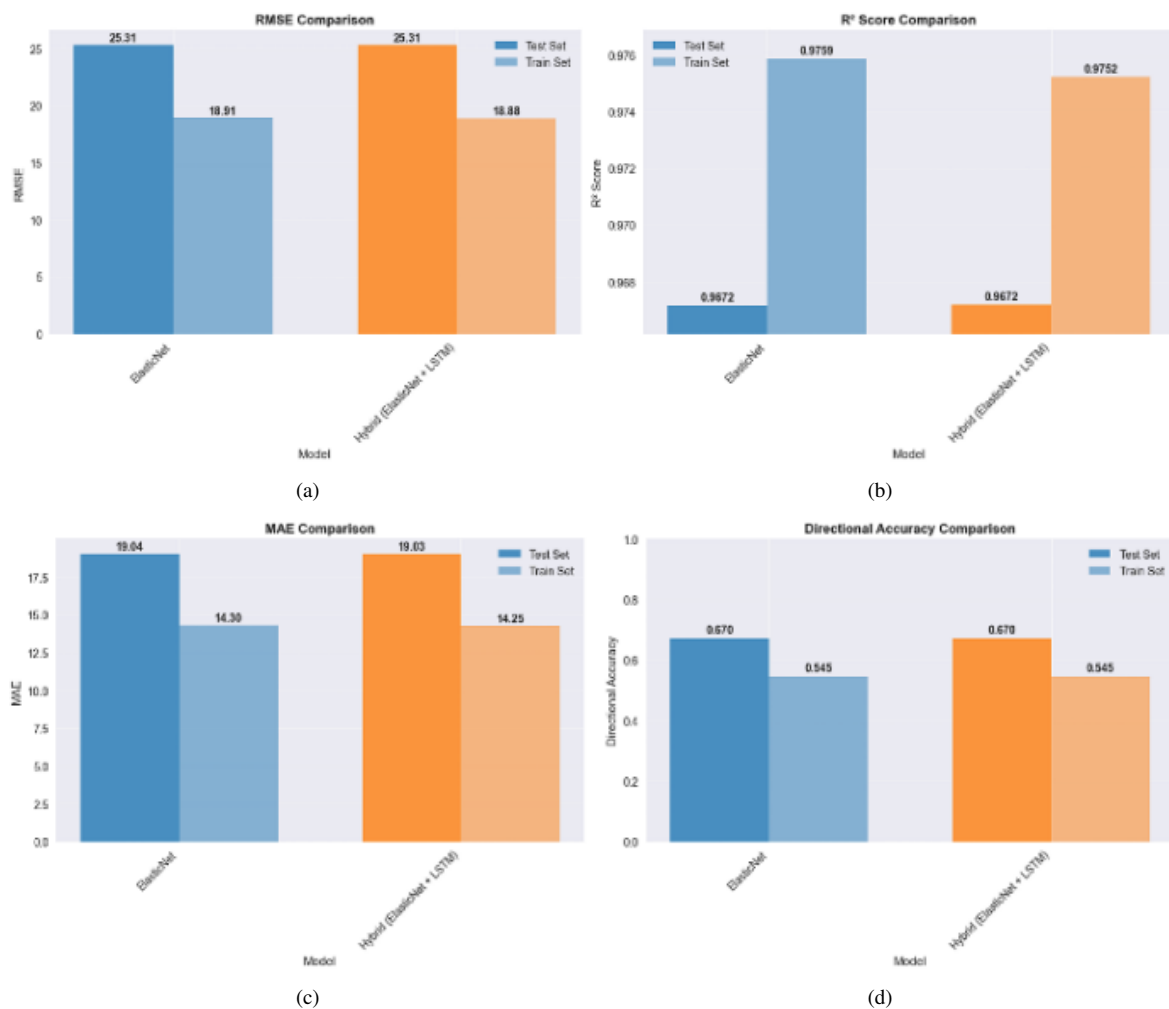
- Sifat Martingale dan Hipotesis Pasar Efisien (EMH):** Dalam horizon harian, setelah komponen tren linier dan ketergantungan kelambatan waktu (lag) dimodelkan oleh ElasticNet, sisaan (residual) harga yang tersisa mendekati proses derau putih stokastik murni (*stochastic white noise*) atau *martingale difference sequence*. Sisaan ini tidak lagi menyimpan struktur dependensi temporal deterministik yang dapat diekstraksi secara teratur oleh LSTM.
- Ketidakesesuaian Frekuensi dan Rasio Sinyal terhadap Derau (Signal-to-Noise Ratio):** Variabel makroekonomi (Inflasi, BI Rate, Indeks Keyakinan Konsumen) dirilis secara periodik (bulanan), sehingga nilainya bersifat konstan selama berminggu-minggu dalam data harian. Ketika dimasukkan ke LSTM melalui komponen PCA untuk memprediksi fluktuasi sisaan harian, rasio sinyal terhadap derau menjadi sangat rendah. LSTM justru berisiko memodelkan derau acak (*noise fitting*) ketimbang pola ekonomi riil.
- Karakteristik Pemodelan Level Harga vs. Klaster Volatilitas:** Model *deep learning* non-linear umumnya lebih unggul dalam menangkap heteroskedastisitas bersyarat (*volatility clustering* pada *return* kuadrat) dibandingkan memprediksi tingkat harga nominal absolut yang secara intrinsik didominasi oleh sinyal autoregresif linier.
- Bukti Empiris Prinsip Occam's Razor:** Temuan ini mengonfirmasi secara empiris prinsip Occam's razor dalam peramalan finansial [14], di mana model dengan kompleksitas terendah yang mampu menjelaskan data secara memadai harus diprioritaskan dibandingkan arsitektur berlapis yang memakan biaya komputasi tinggi tanpa *benefit* akurasi yang nyata.

Untuk beberapa studi peramalan harga saham terkini di pasar modal Indonesia, seperti yang disajikan pada Tabel 6. Dapat dilihat bahwa nilai R^2 sebesar 0,9672 yang dicapai oleh model kami sangat kompetitif dan sebanding dengan model-model berkinerja tinggi lainnya seperti yang dilaporkan oleh Diqi et al. (2024) [9] yang menggunakan Stacked GRU ($R^2=0,978$) dan Fahmi et al. (2024) [10] yang menggunakan Random Effect Model ($R^2=0,958$). Meskipun beberapa model seperti CNN-BiLSTM yang di-

Tabel 4. Bobot Komponen Utama (*Loadings*) dari fitur eksternal terhadap 9 komponen PCA

Fitur	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
ASII Close	−0,2321	0,6510	0,1211	−0,2810	0,6266	−0,1231	0,0129	0,0773	−0,1114
BBRI Close	0,0156	0,4212	−0,1519	−0,1128	−0,3903	0,6403	−0,0952	0,4321	0,0016
BBTN Close	0,0820	−0,1695	−0,0430	−0,0972	0,1110	−0,2022	−0,1699	0,7093	0,1274
BMRI Close	−0,0634	0,2696	0,6132	0,6703	−0,1777	−0,1621	−0,0182	0,1972	0,0085
PGAS Close	0,0012	−0,2161	0,7473	−0,5901	−0,1446	0,1458	−0,0470	−0,0384	0,0294
LQ45 Close	−0,2167	0,1074	0,0244	0,0808	−0,0168	0,2272	−0,0034	−0,3044	0,2043
IDX Finance Close	−0,1721	−0,0667	0,0580	0,1715	0,2253	0,4078	−0,2064	−0,1619	0,0185
JISDOR Close	0,3324	−0,0273	0,0291	−0,0368	0,0569	−0,1251	0,2839	0,1968	−0,1328
CPI Financial	−0,1993	−0,0355	0,0179	0,0035	0,0273	0,1357	0,8615	0,1144	0,3612
Consumer Confidence	0,6333	−0,1035	0,1366	0,1981	0,4987	0,4026	−0,0047	−0,0036	0,1709
BI Rate	0,2546	0,3012	−0,0451	−0,1279	−0,1511	−0,2607	−0,2029	−0,1620	0,7853
BBTN Close	0,4923	0,3625	0,0123	−0,0920	−0,2525	−0,0669	0,2295	−0,2587	−0,3664

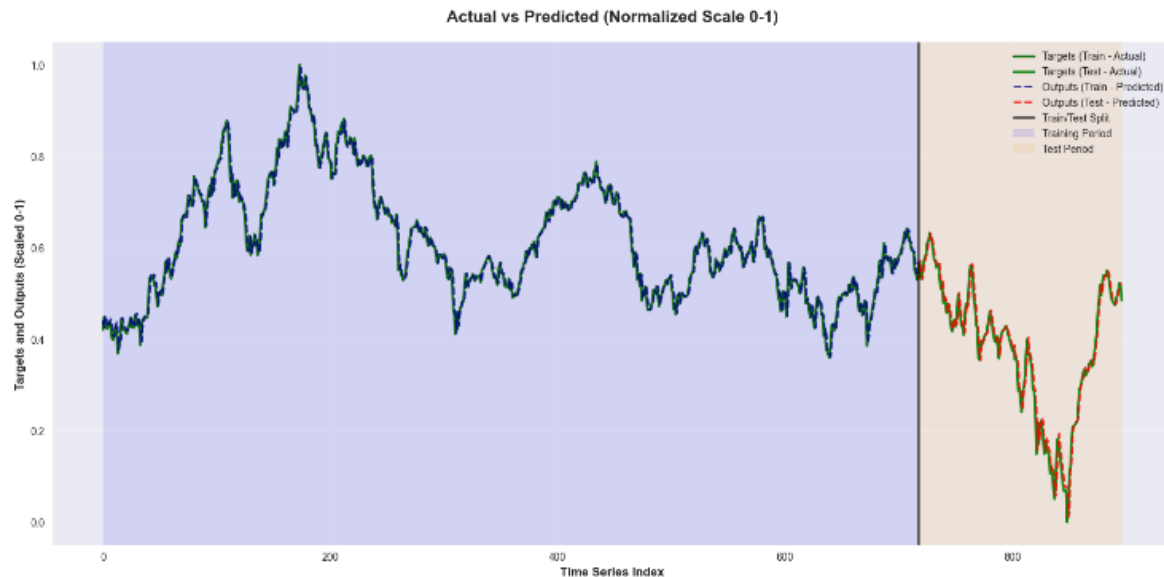
Sumber: hasil pengolahan penulis. Baris terakhir pada naskah asli berlabel “BBTN Close”, padahal label itu sudah muncul sebelumnya — perlu dikonfirmasi ke penulis (lihat catatan editor).

**Gambar 4.** (a) RMSE Comparison; (b) R^2 Score Comparison; (c) MAE Comparison; (d) Directional Accuracy Comparison**Tabel 5.** Hasil uji Diebold-Mariano

Perbandingan Model	Statistik DM	P-value
ElasticNet vs. Hybrid (ElasticNet+LSTM)	1,1125	0,2659

laporkan oleh Navastara et al. (2023) [11] menunjukkan RMSE yang berbeda (92,26), hal ini kemungkinan besar

disebabkan oleh perbedaan skala data target dan ketiadaan fitur lag yang dominan. Keunggulan utama dari pendekatan kami terletak pada pencapaian R^2 yang sangat tinggi melalui model linear yang relatif sederhana, yang menggarisbawahi temuan inti kami: pada data dengan autokorelasi kuat, model yang mampu menangkap dependensi temporal secara efisien akan menunjukkan performa superior.



Gambar 5. Perbandingan harga aktual dan prediksi model pada train dan test set

Tabel 6. Perbandingan kinerja prediksi dengan literatur terkait

Referensi	Subjek/Tujuan	Model	RMSE	R ²
Diqi et al. (2024) [9]	10 saham teratas BEI	Stacked GRU	0,0059*	0,978
Fahmi et al. (2024) [10]	Prediksi harga saham BEI	Random Effect Model	–	0,958
Navastara et al. (2023) [11]	Prediksi harga saham harian BEI	CNN-BiLSTM	92,26	–
Mariyani et al. (2025) [12]	Prediksi <i>return</i> saham teknologi	Regresi Linear	–	0,413
Penelitian Ini	Prediksi harga reksa dana	Hybrid (ElasticNet+LSTM)	25,3052	0,9672

*Nilai RMSE dinormalisasi pada skala [0, 1].

Implikasi Praktis dan Keterbatasan Penelitian (Validasi Lintas Produk)

Temuan ini memiliki implikasi praktis yang sangat bernilai bagi manajer investasi dan praktisi kuantitatif. Model regresi linear teregularisasi seperti ElasticNet terbukti jauh lebih efisien, memiliki kecepatan inferensi instan, hemat sumber daya komputasi, dan yang paling krusial, transparan serta mudah diinterpretasikan koefisiennya oleh pemangku kepentingan dibandingkan model *deep learning black-box*.

Terkait keterbatasan penelitian, studi ini berfokus pada satu produk instrumen, yaitu Reksa Dana Saham Sucor-invest Equity Fund pada rentang 2022–2025. Meskipun produk ini sangat representatif untuk kategori reksa dana ekuitas aktif berkapitalisasi besar, fokus pada satu produk membatasi generalisasi temuan lintas kelas aset. Reksa dana saham memiliki profil volatilitas tinggi yang sangat sensitif terhadap dinamika IHSG. Sebaliknya, reksa dana pendapatan tetap (obligasi) didominasi oleh pergerakan *yield* surat utang negara dan suku bunga acuan dengan volatilitas yang lebih rendah, sedangkan reksa dana pasar uang menunjukkan pergerakan linier yang hampir monotonik. Oleh karena itu, penelitian lanjutan sangat disarankan untuk melakukan pengujian komparatif lintas produk instrumen—meliputi reksa dana obligasi, pasar uang, campuran, serta reksa dana indeks (ETF)—guna menguji apakah keunggulan model linear parsimonius dan fenomena residual *white*

noise ini berlaku universal di seluruh spektrum risiko pasar modal Indonesia.

KESIMPULAN

Penelitian ini menyajikan evaluasi komparatif yang kritis antara model regresi linear teregularisasi ElasticNet dan model hibrida ElasticNet-LSTM dalam memprediksi harga penutupan harian reksa dana saham di Indonesia. Temuan empiris menunjukkan bahwa model linear dasar yang dilengkapi fitur kelambatan waktu (lag) dan reduksi dimensi PCA mencapai akurasi yang sangat tinggi ($R^2 = 0,9672$, RMSE = 25,3143). Penambahan arsitektur *deep learning* LSTM pada residual hanya menghasilkan peningkatan akurasi marginal (RMSE = 25,3052) yang dibuktikan tidak signifikan secara statistik oleh Uji Diebold-Mariano ($p = 0,2659$). Hal ini mengindikasikan bahwa setelah komponen inersia harga ditangkap secara linier, sisaan yang ada bersifat stokastik derau putih tanpa struktur temporal teratur. Studi ini menegaskan bahwa untuk peramalan deret waktu finansial dengan persistensi tinggi, pendekatan linier yang sederhana, cepat, transparan, dan teratur merupakan pilihan yang jauh lebih optimal dan praktis dalam operasional investasi dunia nyata.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih yang sebesar-besarnya kepada Program Studi Sains Data, Fakultas Sains, Insti-

tut Teknologi Sumatera (ITERA) atas dukungan akademis, fasilitas laboratorium, dan lingkungan riset yang kondusif selama pelaksanaan penelitian ini. Penulis juga mengapresiasi panitia penyelenggara kompetisi Datathon atas penyediaan tantangan analitis yang menginisiasi studi ini, serta kepada Otoritas Jasa Keuangan (OJK), Bank Indonesia, PT Bursa Efek Indonesia, dan PT Sucorinvest Asset Management atas ketersediaan publikasi data historis pasar modal Indonesia yang memungkinkan penelitian ini terlaksana.

DAFTAR PUSTAKA

- [1] P. K. D. Lubis, C. L. N. Manalu, A. A. Lubis, L. M. Tambunan, and F. Saputra, “The Role of the Capital Market in Increasing Economic Growth in Indonesia,” *Indonesian Journal of Interdisciplinary Research in Science and Technology*, vol. 2, no. 5, pp. 557–568, May 2024, doi: 10.55927/marcopolo.v2i5.9322.
- [2] S. M. Rizky, B. Andini, M. Hasda, M. Syaipudin, and W. F. R. Sudirman, “Development of Mutual Funds in The Indonesian Capital Market,” *Journal of Financial and Business*, vol. 1, no. 1, pp. 11–22, Jul. 2024, doi: 10.69693/jfb.v1i1.15.
- [3] Z. Makarim and M. Anas, “The Effect of Macroeconomics and JSEC on Mutual Fund NAV During the Pandemic,” *International Student Conference on Business, Education, Economics, Accounting, and Management (ISC-BEAM)*, vol. 2, pp. 581–596, Jul. 2024, doi: 10.21009/ISC-BEAM.012.37.
- [4] L. H. Ningrum, T. Marwa, I. Asngari, and S. Andaiyani, “Evaluating the Impact of Inflation and Exchange Rates on the Net Active Value (NAV) of Conventional Mutual Funds in Indonesia,” *KnE Social Sciences*, pp. 780–792, May 2024, doi: 10.18502/kss.v9i13.15984.
- [5] A. Boonprasope and K. Y. Tipayawong, “Predicting Healthcare Mutual Fund Performance Using Deep Learning and Linear Regression,” *International Journal of Financial Studies*, vol. 12, no. 1, Art. no. 23, Mar. 2024, doi: 10.3390/ijfs12010023.
- [6] V. R. Joseph, “Optimal ratio for data splitting,” *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 15, no. 4, pp. 531–538, Aug. 2022, doi: 10.1002/sam.11583.
- [7] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York, NY, USA: Springer-Verlag, 2002.
- [8] F. X. Diebold and R. S. Mariano, “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 253–263, Jul. 1995, doi: 10.1080/07350015.1995.10524599.
- [9] M. Diqi, M. E. Hiswati, and N. Wijaya, “Stacked Gated Recurrent Units and Indonesian Stock Predictions: A New Approach to Financial Forecasting,” *Jurnal Ilmu Komputer dan Teknologi (IKOMTI)*, vol. 6, no. 1, pp. 1–9, Feb. 2024.
- [10] R. N. Fahmi, R. S. Ramadhan, and N. E. A. A. Trihapsari, “Indonesian Stock Price Prediction Using Random Effect Model with Technical Indicators in Panel Data,” in *Proc. 2024 Beyond Technology Summit on Informatics International Conference (BTS-I2C)*, Jember, Indonesia: IEEE, 2024, pp. 89–94, doi: 10.1109/BTS-I2C62621.2024.10667822.
- [11] D. A. Navastara, F. R. Ananta, and H. A. Y. Wisesa, “Prediction of Indonesian Stock Price Using Combination of CNN and BiLSTM Model,” in *Proc. 2023 International Conference on Information Technology and Computing (ICITCOM)*, Yogyakarta, Indonesia: IEEE, 2023, pp. 307–312, doi: 10.1109/ICITCOM59616.2023.10433722.
- [12] N. P. Mariyani, P. W. Saitri, and N. P. Y. Mendra, “Predicting Stock Returns in Indonesian Technology Companies,” *International Journal of Accounting & Finance in Asia Pasific (IJAFAP)*, vol. 8, no. 1, pp. 94–110, Feb. 2025, doi: 10.32535/ijafap.v8i1.3639.
- [13] C. Mingard, H. Rees, G. Valle-Pérez, and A. A. Louis, “Deep neural networks have an inbuilt Occam’s razor,” *Nature Communications*, vol. 14, no. 1, Art. no. 5740, Sep. 2023, doi: 10.1038/s41467-023-41398-3.
- [14] G. P. Zhang, “Time series forecasting using a hybrid ARIMA and neural network model,” *Neurocomputing*, vol. 50, pp. 159–175, Jan. 2003, doi: 10.1016/S0925-2312(01)00702-0.