Analisis Perbandingan *Tree-Based Machine Learning* dan *Deep Learning* dalam Prediksi Harga Saham Berbasis *Multivariate Time Series*

Delon Bernady[1)], Tomas Leonardo[2)]

[1)2)]Jurusan Teknik Informatika, Fakultas Teknik, Universitas Palangka Raya
Jl. Yos Sudarso, Palangka Raya 73112
[1)]delonbernady80@mhs.eng.upr.ac.id
[2)]tomasleonardo@it.upr.ac.id

Abstrak

Penelitian ini membandingkan XGBoost dan PatchTST dalam memprediksi harga saham PT Medco Energi Internasional Tbk (MEDC) menggunakan pendekatan multivariate time series. Data yang digunakan meliputi riwayat harga saham MEDC, Indeks Harga Saham Gabungan (IHSG), harga minyak mentah WTI, indikator teknikal, dan sentimen berita. Mutual Information digunakan untuk melakukan seleksi fitur, sedangkan kinerja model dievaluasi menggunakan eight-fold walk-forward validation dengan rolling window. Metrik evaluasi yang digunakan meliputi RMSE, MAE, MAPE, R^2, dan CV, serta didukung oleh analisis feature importance, SHAP, dan Wilcoxon Signed-Rank Test dengan koreksi Bonferroni. Secara umum, PatchTST menunjukkan kinerja dan stabilitas yang lebih baik dibandingkan XGBoost. Konfigurasi PatchTST FS8 menghasilkan nilai RMSE terendah sebesar 53,56, MAE sebesar 41,20, MAPE sebesar 5,19%, dan R^2 sebesar 0,8710. Sementara itu, XGBoost FS8 menghasilkan RMSE sebesar 90,89, MAE sebesar 67,69, MAPE sebesar 6,82%, dan R^2 sebesar 0,6184. Namun, hasil pengujian Wilcoxon menunjukkan bahwa perbedaan kinerja kedua model tersebut tidak signifikan secara statistik setelah penerapan koreksi Bonferroni. Analisis fitur mengidentifikasi Terakhir_MEDC_Lag_1 sebagai fitur yang secara konsisten memiliki pengaruh dominan, sedangkan sentimen berita memberikan kontribusi yang relatif rendah.

Kata kunci: Prediksi Harga Saham, Tree-Based Machine Learning, Deep Learning, PatchTST, XGBoost, Multivariate Time Series.

Abstract

This study compares XGBoost and PatchTST for predicting the stock price of PT Medco Energi Internasional Tbk (MEDC) using a multivariate time series approach. The data include historical MEDC stock prices, the Jakarta Composite Index (IHSG), WTI crude oil prices, technical indicators, and news sentiment. Mutual Information was used for feature selection, while model performance was evaluated using eight-fold walk-forward validation with a rolling window. Evaluation metrics included RMSE, MAE, MAPE, R^2, and CV supported by feature importance, SHAP, and the Wilcoxon Signed-Rank Test with Bonferroni correction. PatchTST generally showed better performance and stability than XGBoost. The PatchTST FS8 configuration achieved the lowest RMSE of 53.56, MAE of 41.20, MAPE of 5.19%, and R^2 of 0.8710, while XGBoost FS8 achieved an RMSE of 90.89, MAE of 67.69, MAPE of 6.82%, and R^2 of 0.6184. However, the Wilcoxon test showed that the performance differences were not statistically

DOI:https://doi.org/10.47111/jointecoms.v6i3 *Received: 01-09-2026*
Accepted: 10-09-2026

significant after Bonferroni correction. Feature analysis identified Terakhir_MEDC_Lag_1 as a consistently dominant feature, while news sentiment had a relatively low contribution.

Keywords: *Stock Price Prediction, Tree-Based Machine Learning, Deep Learning, PatchTST, XGBoost, Multivariate Time Series*.

1. PENDAHULUAN

Perkembangan teknologi informasi memudahkan masyarakat dalam melakukan transaksi saham melalui perangkat digital dan mendorong meningkatnya partisipasi investor di pasar modal. Harga saham menjadi salah satu indikator penting dalam menilai kinerja perusahaan dan menjadi pertimbangan dalam pengambilan keputusan investasi. Namun, harga saham memiliki pola yang dinamis dan dipengaruhi oleh berbagai faktor, seperti kondisi makroekonomi, kondisi pasar, peristiwa sosial dan ekonomi, serta kebijakan perusahaan [1], [2].

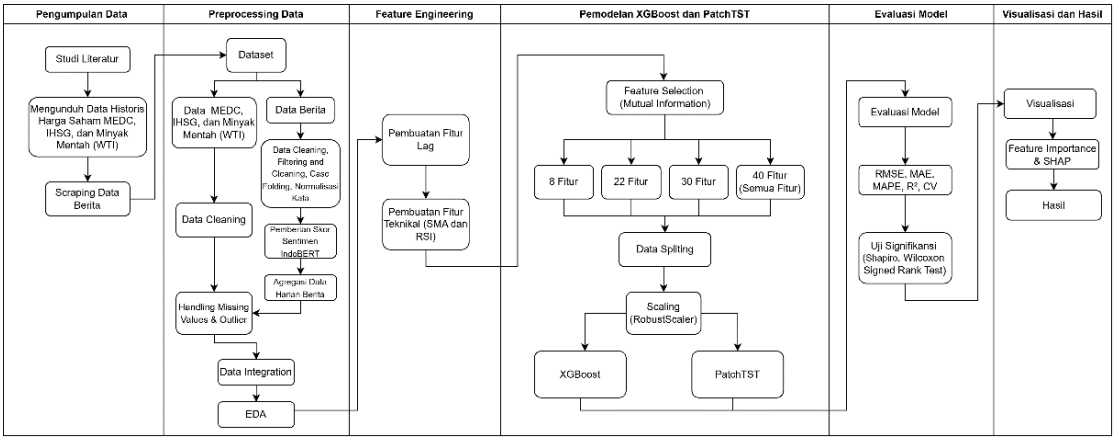
Prediksi harga saham tidak hanya bergantung pada data historis harga saham itu sendiri (*univariate*), tetapi juga dapat memanfaatkan berbagai variabel eksternal dalam pendekatan *multivariate time series*, seperti kondisi pasar, faktor ekonomi, dan sentimen berita [3]. Indeks Harga Saham Gabungan (IHSG) dapat digunakan untuk merepresentasikan kondisi pasar modal secara keseluruhan, sedangkan harga minyak mentah dunia (West Texas Intermediate/WTI) relevan digunakan pada perusahaan sektor energi karena fluktuasi harga minyak dapat berkaitan dengan pendapatan dan biaya operasional perusahaan [4]. Selain faktor pasar dan ekonomi, sentimen berita juga dapat memengaruhi pergerakan harga saham karena informasi yang disampaikan melalui berita dapat membentuk persepsi dan ekspektasi investor terhadap kondisi perusahaan maupun prospek industri di masa mendatang [5]. Karena hal tersebut, prediksi harga saham diperlukan untuk membantu memahami pola pergerakan harga dan menyediakan informasi yang dapat mendukung pengambilan keputusan investasi.

Pendekatan yang dibandingkan mewakili dua karakter pemodelan. XGBoost mempelajari hubungan nonlinier dari fitur tabular [6], sedangkan Transformer (PatchTST) menangkap pola dan dependensi pada data deret waktu melalui mekanisme *self-attention* dan *patching* [7], [8]. Meskipun berbagai penelitian telah menerapkan metode *machine learning* dan *deep learning* untuk prediksi harga saham, kajian yang membandingkan XGBoost dan PatchTST secara langsung pada saham dengan memanfaatkan data multivariat dan beberapa konfigurasi fitur masih terbatas. Selain itu, penggunaan konfigurasi fitur yang berbeda dapat menghasilkan informasi masukan yang berbeda bagi model, sehingga pengaruh pemilihan fitur terhadap kinerja prediksi perlu dikaji lebih lanjut.

Berdasarkan permasalahan tersebut penelitian ini bertujuan untuk membangun dan membandingkan model XGBoost dan PatchTST dalam memprediksi harga saham MEDC menggunakan data *multivariate time series*. Perbandingan dilakukan pada beberapa konfigurasi fitur untuk mengevaluasi pengaruh pemilihan fitur terhadap kinerja masing-masing model berdasarkan metrik RMSE, MAE, MAPE, R^2, dan CV. Selain itu, penelitian ini menganalisis kontribusi fitur terhadap hasil prediksi menggunakan *feature importance* dan SHAP untuk mengidentifikasi fitur yang dominan serta memahami kontribusinya terhadap prediksi masing-masing model.

2. METODOLOGI PENELITIAN

Metodologi penelitian adalah pendekatan sistematik yang diterapkan untuk merencanakan, melaksanakan, dan menganalisis data dalam suatu penelitian. Metodologi penelitian dapat dilihat pada Gambar 1.



Gambar 1. Metodologi Penelitian

2.1. Pengumpulan Data

Data Data penelitian berupa data harian yang diperoleh dari Investing.com, meliputi data historis saham PT Medco Energi Internasional Tbk (MEDC), Indeks Harga Saham Gabungan (IHSG), harga minyak mentah West Texas Intermediate (WTI), serta data sentimen berita yang diperoleh melalui proses *scraping* berita pada periode 2016 hingga Maret 2026. Data MEDC digunakan sebagai objek utama penelitian, sedangkan IHSG, WTI, dan sentimen berita digunakan sebagai variabel eksternal dalam membentuk data *multivariate time series*. Target prediksi dalam penelitian ini adalah harga penutupan saham MEDC yang direpresentasikan oleh variabel Terakhir_MEDC. Setiap dataset memiliki jumlah observasi dan atribut yang berbeda sesuai dengan ketersediaan data masing-masing sumber. Rincian dataset yang digunakan dalam penelitian ditunjukkan pada Tabel 1.

Tabel 1. Dataset Penelitian

Dataset	Jumlah Baris	Kolom/Fitur
MEDC	2.471	Date, Open, High, Low, Close, Volume, Return
IHSG	2.473	Date, Open, High, Low, Close, Volume, Return
WTI	2.694	Date, Open, High, Low, Close, Volume, Return
Sentimen Berita	149.905	Date, Title, Publisher, Url_asli, Isi_berita

Data MEDC, IHSG, dan sentimen berita memiliki jumlah observasi yang relatif serupa, sedangkan WTI memiliki jumlah observasi lebih banyak karena data harga minyak tidak mengikuti kalender perdagangan saham Indonesia. Seluruh dataset selanjutnya digabungkan berdasarkan tanggal yang sesuai dan diurutkan secara kronologis untuk mempertahankan urutan temporal. Proses tersebut dilakukan sebelum pembentukan fitur dan pemodelan agar informasi dari periode masa depan tidak digunakan dalam proses pelatihan.

2.2. Preprocessing Data

Prapemrosesan dilakukan melalui pembersihan dan penyelarasan data berdasarkan tanggal. Data historis MEDC, IHSG, dan WTI dibersihkan dengan menyeragamkan format tanggal, mengonversi tipe data numerik, menangani format angka dan satuan volume, serta menyesuaikan nama kolom untuk membedakan masing-masing sumber data [9], [10]. Data berita dibersihkan dari duplikasi berdasarkan URL dan disaring untuk menghapus berita yang tidak relevan menghasilkan 45.230 baris data. Selanjutnya, dilakukan *text cleaning* untuk menghapus karakter atau elemen teks yang tidak diperlukan, *case folding* untuk mengubah seluruh teks menjadi huruf kecil, serta normalisasi bahasa untuk menyeragamkan bentuk kata yang tidak baku ke bentuk yang sesuai [11], [12], [13].

Klasifikasi sentimen dilakukan menggunakan model pre-trained IndoBERT, yaitu *dwili/bert-bahasa-sentimen-berita* dari Hugging Face. Hasil klasifikasi kemudian dikonversi menjadi

sentiment score dan diagregasi berdasarkan tanggal untuk memperoleh nilai sentimen harian. Berita yang dipublikasikan pada hari non-perdagangan disesuaikan dengan tanggal perdagangan berikutnya agar informasi berita tetap dapat dikaitkan dengan observasi pasar yang relevan. Selanjutnya, data harga, indikator pasar, dan sentimen berita diselaraskan berdasarkan tanggal perdagangan untuk membentuk dataset multivariat yang digunakan dalam pemodelan [14]. Selanjutnya, data harga, indikator pasar, dan skor sentimen berita diselaraskan berdasarkan tanggal perdagangan untuk membentuk dataset multivariat yang digunakan dalam pemodelan.

2.3. Feature Engineering

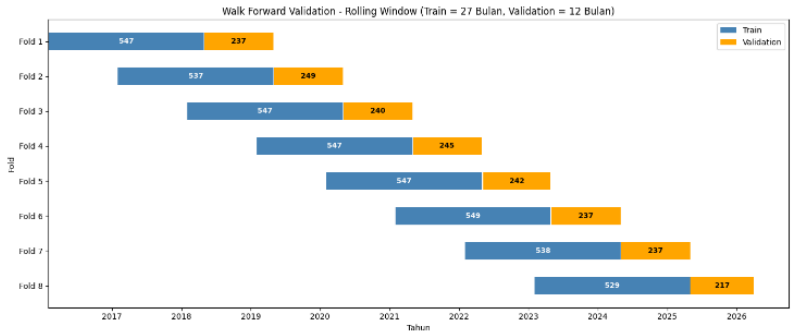
Fitur dibentuk menggunakan nilai lag historis pada MEDC, IHSG, dan WTI, serta indikator teknikal berupa *Simple Moving Average* (SMA) dengan periode 5 dan 10 hari dan *Relative Strength Index* (RSI). SMA 5 dan 10 hari digunakan untuk merepresentasikan kecenderungan pergerakan harga dalam jangka pendek, sedangkan RSI digunakan untuk merepresentasikan kondisi momentum harga. Selain itu, fitur lag digunakan untuk menyediakan informasi historis sebagai masukan model sekaligus mencegah nilai target pada waktu yang sama digunakan sebagai prediktor, sehingga risiko *data leakage* dapat diminimalkan[15], [16]. Rincian fitur yang dibentuk disajikan pada Tabel 2.

Tabel 2 Hasil *Feature Engineering*

Fitur	Fitur Lag	Indikator Teknikal
MEDC	OHLCV, Return pada lag 1; Terakhir pada lag 3, 5, 10, dan 20	SMA 5, SMA 10, RSI 14
IHSG	OHLCV, Return pada lag 1; Terakhir pada lag 3, 5, 10, dan 20	SMA 5, SMA 10, RSI 14
Minyak Mentah WTI	OHLCV, Return pada lag 1; Terakhir pada lag 3, 5, 10, dan 20	SMA 5, SMA 10, RSI 14
Sentimen Berita	Sentimen hari sebelumnya untuk prediksi hari berikutnya	-

2.4. Pembagian Data dan Pemodelan

Feature selection dilakukan menggunakan *Mutual Information* (MI) dan menghasilkan empat konfigurasi fitur, yaitu seluruh 40 fitur, FS8, FS22, dan FS30 [17]. Data dibagi berdasarkan urutan waktu dan dievaluasi menggunakan 8-fold *walk-forward validation* dengan skema *rolling window* [18]. Pada setiap *fold*, model dilatih menggunakan 27 bulan data dan dievaluasi pada 12 bulan data berikutnya. Skema ini mempertahankan urutan temporal sehingga data pada periode validasi selalu berada setelah data latih dan informasi dari masa depan tidak digunakan dalam proses pelatihan [19] . Periode pembagian dapat dilihat pada Gambar 2.



Gambar 2. Pembagian Data

RobustScaler diterapkan secara terpisah pada data latih di setiap *fold* untuk mencegah kebocoran informasi [20]. *Extreme Gradient Boosting* (XGBoost) digunakan sebagai model *tree-based machine learning* untuk mempelajari hubungan nonlinier antara fitur masukan dan harga saham MEDC. XGBoost membangun sejumlah pohon keputusan secara bertahap, di mana setiap pohon digunakan untuk memperbaiki kesalahan prediksi dari pohon sebelumnya. Pendekatan

tersebut memungkinkan model menangkap hubungan kompleks antarfitur tanpa mengharuskan hubungan antara variabel masukan dan target bersifat linear [21]. Sementara itu, *Patch Time Series Transformer* (PatchTST) digunakan sebagai model *deep learning* berbasis Transformer untuk mempelajari pola deret waktu multivariat. PatchTST membagi rangkaian data menjadi beberapa *patch* yang berisi sejumlah titik waktu berurutan sebelum memprosesnya menggunakan mekanisme *self-attention*. Pembentukan *patch* memungkinkan model mempelajari pola lokal dalam deret waktu sekaligus memanfaatkan hubungan antarpola melalui mekanisme *self-attention* [22] .

Kedua model dilatih dan dievaluasi menggunakan konfigurasi fitur dan pembagian fold yang sama sehingga setiap model memperoleh informasi masukan dan periode evaluasi yang setara. Proses pelatihan dan evaluasi dilakukan menggunakan random seed 0 untuk menjaga konsistensi hasil eksperimen. Dengan kondisi tersebut, perbedaan hasil evaluasi dapat dianalisis berdasarkan karakteristik model dan konfigurasi fitur yang digunakan, bukan akibat perbedaan pembagian data atau kondisi eksperimen.

2.5. Hasil Evaluasi dan Visualisasi

Evaluasi Kinerja model dievaluasi menggunakan hasil dari rata-rata delapan fo*ld walk-forward validation* dengan skema *rolling window*, matriks evaluasi yang digunakan RMSE, MAE, MAPE, dan R^2 sebagai metrik utama, sedangkan *coefficient of variation* (CV) digunakan sebagai ukuran tambahan untuk menilai konsistensi kinerja antar-*fold* [23] . Nilai rata-rata dan standar deviasi dari delapan *fold* digunakan untuk merangkum kinerja dan variasi model sepanjang periode validasi. Perbedaan kinerja antara XGBoost dan PatchTST diuji menggunakan *Wilcoxon Signed-Rank Test* dengan koreksi Bonferroni [24], [25]. Hasil kontribusi fitur divisualisasikan untuk menggambarkan tingak kepentingan atau *feature importance* dan SHapley Additive exPlanations (SHAP) digunakan untuk menganalisis kontribusi fitur terhadap hasil prediksi masing-masing model [26].

3. HASIL DAN PEMBAHASAN

Pengujian dilakukan untuk mengevaluasi kinerja XGBoost dan PatchTST dalam memprediksi harga saham MEDC menggunakan data multivariat dengan beberapa konfigurasi fitur. Evaluasi dilakukan melalui walk-forward validation untuk melihat kinerja dan variasi prediksi pada periode yang berbeda. Selain itu, pengujian statistik digunakan untuk menilai signifikansi perbedaan kinerja antarmodel, sedangkan analisis fitur digunakan untuk mengidentifikasi fitur yang berperan dalam proses prediksi. Hasil pengujian dan analisis tersebut dibahas berdasarkan kinerja model, variasi antar-fold, signifikansi perbedaan kinerja, serta kontribusi fitur.

3.1 Kinerja Model dan Konfigurasi Fitur

Kinerja XGBoost dan PatchTST dievaluasi pada empat konfigurasi fitur, yaitu Semua, FS8, FS22, dan FS30, menggunakan delapan *fold walk-forward validation*. Hasil pada setiap *fold* diringkas dalam bentuk nilai rata-rata dan standar deviasi untuk RMSE, MAE, MAPE, dan R^2. Nilai rata-rata menunjukkan kinerja model secara keseluruhan selama periode validasi, sedangkan standar deviasi menunjukkan variasi kinerja antar-*fold*. Hasil evaluasi untuk masing-masing model dan konfigurasi fitur disajikan pada Tabel 3.

Tabel 3. Hasil Rata-Rata *Fold Walk Forward Validation*

Model	**Fitur**	**Mean ± Std**			
		RMSE	**MAE**	**MAPE (%)**	**R^2**
XGBoost	Semua	108.55 ± 115.45	83.37 ± 94.13	8.78 ± 8.27	0.4700 ± 0.8265
	FS 8	90.89 ± 105.51	67.69 ± 85.80	6.82 ± 7.62	0.6184 ± 0.6842
	FS 22	110.57 ± 113.04	84.93 ± 92.76	9.09 ± 8.21	0.4466 ± 0.8128

	FS 30	113.80 ± 115.35	88.55 ± 94.53	9.65 ± 8.43	0.4009 ± 0.8501
PatchTST	Semua	66.09 ± 21.76	51.01 ± 16.83	6.25 ± 2.29	0.8101 ± 0.0641
	FS 8	**53.56** ± **14.17**	**41.20** ± **10.16**	**5.19** ± **2.40**	**0.8710** ± **0.0505**
	FS 22	59.40 ± 30.64	46.41 ± 24.87	5.44 ± 2.31	0.8535 ± 0.0688
	FS 30	70.81 ± 32.13	55.56 ± 27.28	6.93 ± 3.56	0.7765 ± 0.1143

Berdasarkan hasil *walk-forward validation*, penggunaan *feature selection* memberikan pengaruh terhadap kinerja kedua model. Pada XGBoost, konfigurasi FS8 menghasilkan kinerja terbaik dengan RMSE 90,89, MAE 67,69, MAPE 6,82%, dan R^2 0,6184. Pengurangan fitur dari 40 menjadi 8 fitur meningkatkan kinerja dibandingkan penggunaan seluruh fitur, tetapi penambahan fitur kembali pada FS22 dan FS30 justru diikuti penurunan kinerja. Pola yang serupa terlihat pada PatchTST, dengan FS8 menghasilkan kinerja terbaik, yaitu RMSE 53,56, MAE 41,20, MAPE 5,19%, dan R^2 0,8710. Hasil ini menunjukkan bahwa penggunaan fitur yang lebih ringkas dapat mempertahankan informasi yang relevan sekaligus mengurangi fitur yang kurang informatif bagi proses prediksi.

Dari sisi kestabilan, PatchTST menunjukkan variasi kinerja antar-fold yang lebih rendah dibandingkan XGBoost. Pada konfigurasi FS8, simpangan baku RMSE PatchTST sebesar 14,17, sedangkan XGBoost sebesar 105,51. Pola serupa terlihat pada MAE dan R^2. Perbedaan tersebut menunjukkan bahwa kinerja XGBoost lebih berubah antarperiode rolling window, sedangkan PatchTST cenderung mempertahankan kinerja yang lebih konsisten. Hal ini juga terlihat pada fold dengan perubahan kinerja yang besar, ketika XGBoost mengalami penurunan kinerja sementara PatchTST tetap menghasilkan R^2 positif.

Konfigurasi FS8 memberikan hasil terbaik pada kedua model, dengan PatchTST menghasilkan nilai error yang lebih rendah dan R^2 yang lebih tinggi dibandingkan XGBoost. Hasil tersebut menunjukkan bahwa pada dataset penelitian ini, PatchTST mampu mempertahankan kinerja prediksi yang lebih konsisten ketika diterapkan pada periode data yang berbeda, tetapi belum menunjukkan kestabilan terhadap variasi random seed karena seluruh eksperimen menggunakan seed 0. Namun, perbedaan kinerja tersebut selanjutnya perlu dilihat dari hasil pengujian statistik untuk mengetahui apakah perbedaan antara kedua model dan konfigurasi tersebut signifikan secara statistik.

3.2 Variasi Kinerja Antar-Fold

Selain nilai rata-rata dan standar deviasi, variasi relatif kinerja antar-*fold* dianalisis menggunakan *coefficient of variation* (CV). CV digunakan sebagai ukuran tambahan untuk membandingkan konsistensi kinerja model terhadap nilai rata-ratanya pada setiap metrik. Semakin rendah nilai CV, semakin kecil variasi relatif kinerja antar-fold. Hasil perhitungan CV untuk setiap model dan konfigurasi fitur disajikan pada Tabel 4.

Tabel 4. Coefficient of Variation (CV)

Model	Fitur	RMSE CV	MAE CV	MAPE CV	R^2 CV
XGBoost	Semua	1.0636	1.1290	0.9419	1.7585
	FS 8	1.1609	1.2675	1.1173	1.1064
	FS 22	1.0223	1.0922	0.9032	1.8199
	FS 30	1.0136	1.0675	0.8736	2.1205
PatchTST	Semua	0.3293	0.3299	**0.3664**	0.0791
	FS 8	**0.2646**	**0.2466**	0.4624	**0.0580**
	FS 22	0.5158	0.5359	0.4246	0.0806
	FS 30	0.4537	0.4910	0.5137	0.1472

Nilai CV menunjukkan adanya perbedaan variasi relatif kinerja antar-fold antara XGBoost dan PatchTST. Pada konfigurasi FS8, nilai CV PatchTST lebih rendah daripada XGBoost pada RMSE, MAE, MAPE, dan R^2. Pola tersebut menunjukkan bahwa perubahan periode evaluasi menghasilkan variasi kinerja yang relatif lebih kecil pada PatchTST dibandingkan XGBoost. Hasil ini sejalan dengan standar deviasi yang menunjukkan variasi antar-fold PatchTST lebih rendah, terutama pada konfigurasi FS8. Dengan demikian, CV digunakan sebagai ukuran tambahan untuk melihat konsistensi relatif, sedangkan RMSE, MAE, MAPE, dan R^2 tetap menjadi ukuran utama kinerja prediksi.

Sehingga, nilai CV memperkuat temuan bahwa PatchTST memiliki konsistensi kinerja antarperiode yang lebih tinggi dibandingkan XGBoost, terutama pada konfigurasi FS8. Hal ini menunjukkan bahwa perubahan periode evaluasi pada *rolling window* memberikan pengaruh relatif lebih kecil terhadap kinerja PatchTST.

3.3 Analisis Signifikansi Perbedaan Kinerja

Untuk mengetahui apakah perbedaan kinerja antara XGBoost dan PatchTST bersifat signifikan secara statistik, dilakukan Wilcoxon Signed-Rank Test pada setiap metrik evaluasi dan konfigurasi fitur. Pengujian dilakukan dengan membandingkan nilai kinerja kedua model pada delapan *fold* yang sama, sehingga pasangan pengamatan berasal dari periode validasi yang identik. Karena terdapat empat perbandingan yang dilakukan pada empat metrik dan empat konfigurasi fitur, tingkat signifikansi disesuaikan menggunakan koreksi Bonferroni menjadi 0,0125. Hasil pengujian signifikansi disajikan pada Tabel 5.

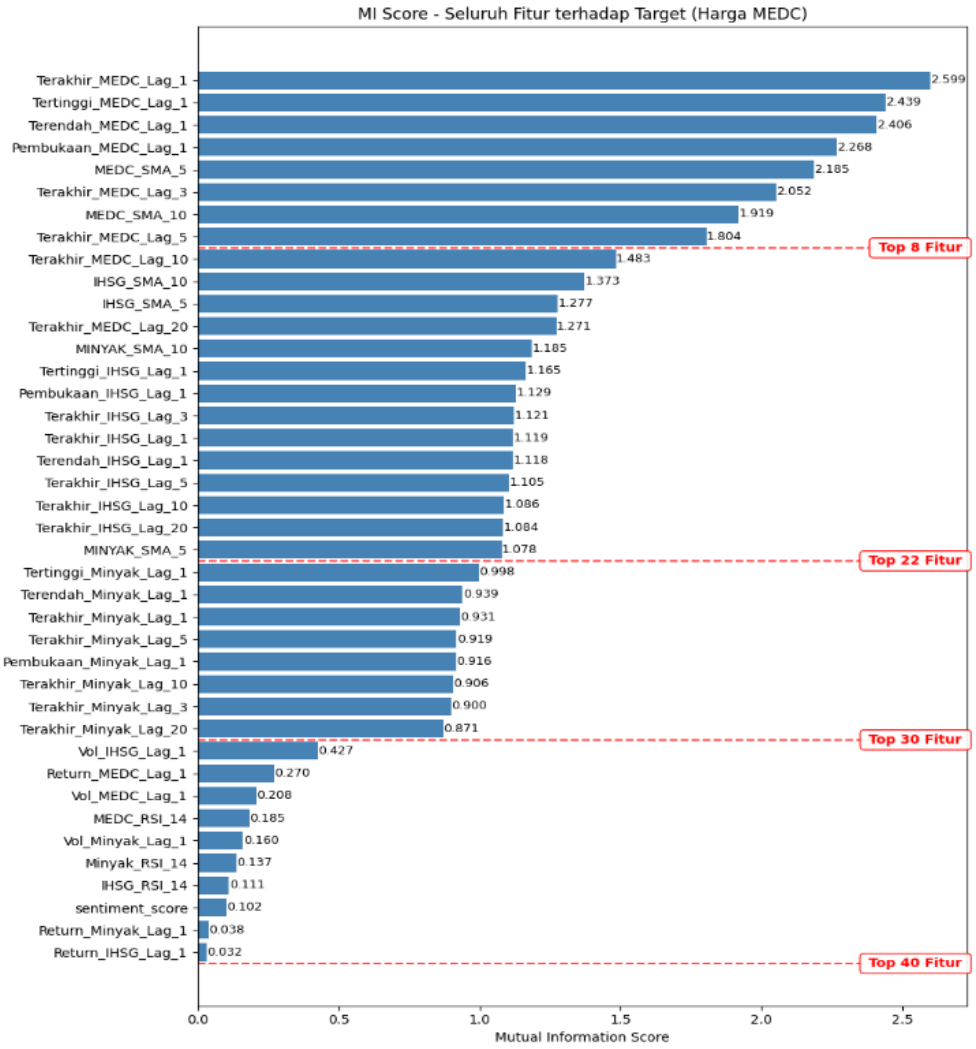
Tabel 5. Hasil Uji Wilcoxon Signed-Rank Test

Metrik	Model 1	Model 2	Statistik	p-value	Keputusan
RMSE	XGBoost Semua	PatchTST Semua	14,0	0,6406	Tidak Signifikan
	XGBoost FS8	PatchTST FS8	16,0	0,8438	Tidak Signifikan
	XGBoost FS22	PatchTST FS22	4,0	0,0547	Tidak Signifikan
	XGBoost FS30	PatchTST FS30	8,0	0,1953	Tidak Signifikan
MAE	XGBoost Semua	PatchTST Semua	14,0	0,6406	Tidak Signifikan
	XGBoost FS8	PatchTST FS8	17,0	0,9453	Tidak Signifikan
	XGBoost FS22	PatchTST FS22	4,0	0,0547	Tidak Signifikan
	XGBoost FS30	PatchTST FS30	9,0	0,2500	Tidak Signifikan
MAPE	XGBoost Semua	PatchTST Semua	16,0	0,8438	Tidak Signifikan
	XGBoost FS8	PatchTST FS8	16,0	0,8438	Tidak Signifikan
	XGBoost FS22	PatchTST FS22	6,0	0,1094	Tidak Signifikan
	XGBoost FS30	PatchTST FS30	11,0	0,3828	Tidak Signifikan
R^2	XGBoost Semua	PatchTST Semua	10,0	0,3125	Tidak Signifikan
	XGBoost FS8	PatchTST FS8	16,0	0,8438	Tidak Signifikan
	XGBoost FS22	PatchTST FS22	4,0	0,0547	Tidak Signifikan
	XGBoost FS30	PatchTST FS30	7,0	0,1484	Tidak Signifikan

Pengujian Wilcoxon Signed-Rank Test dilakukan untuk membandingkan kinerja XGBoost dan PatchTST pada setiap konfigurasi fitur. Hasil pengujian menunjukkan bahwa seluruh p-value setelah koreksi Bonferroni berada di atas tingkat signifikansi 0,0125. Nilai p-value terendah adalah 0,0547, yang diperoleh pada RMSE, MAE, dan R^2 konfigurasi FS22. Dengan demikian, tidak terdapat perbedaan kinerja yang signifikan secara statistik antara XGBoost dan PatchTST pada seluruh konfigurasi dan metrik yang diuji.

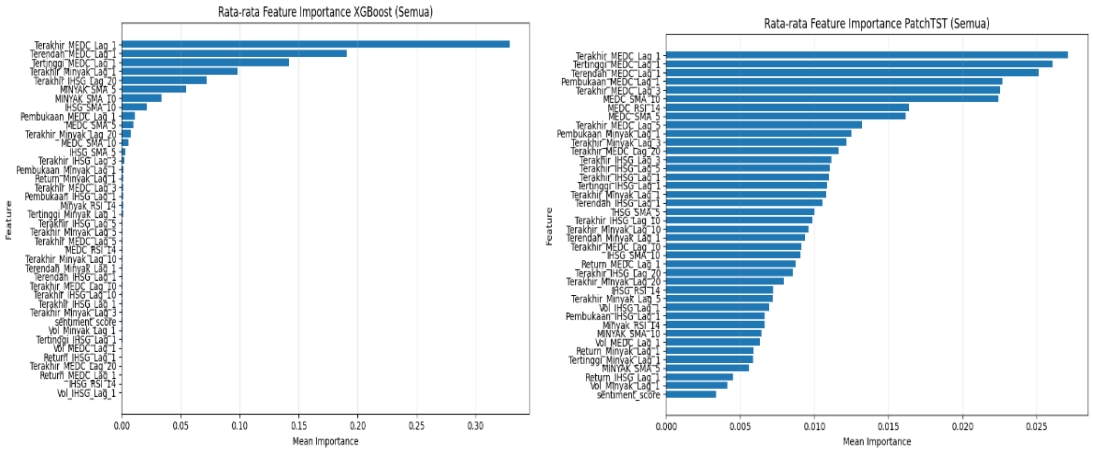
3.4 Analisis Kontribusi Fitur

Analisis kontribusi fitur dilakukan secara bertahap menggunakan *Mutual Information* (MI), *feature importance*, dan SHAP (*SHapley Additive exPlanations*) untuk melihat konsistensi informasi fitur dari tahap *feature selection* hingga proses prediksi model. MI digunakan pada tahap awal untuk mengukur hubungan informasi antara setiap fitur dan target serta menentukan peringkat fitur yang menjadi dasar pembentukan konfigurasi FS8, FS22, dan FS30. Selanjutnya, *feature importance* dan SHAP digunakan untuk menganalisis kontribusi fitur setelah digunakan dalam model. Dengan demikian, ketiga metode tersebut digunakan untuk melihat apakah fitur yang memiliki hubungan informasi tinggi pada tahap *feature selection* juga menunjukkan kontribusi yang besar ketika digunakan oleh model. Nilai MI dapat dilihat pada Gambar 3.



Gambar 3. Hasil Mutual Information (MI)

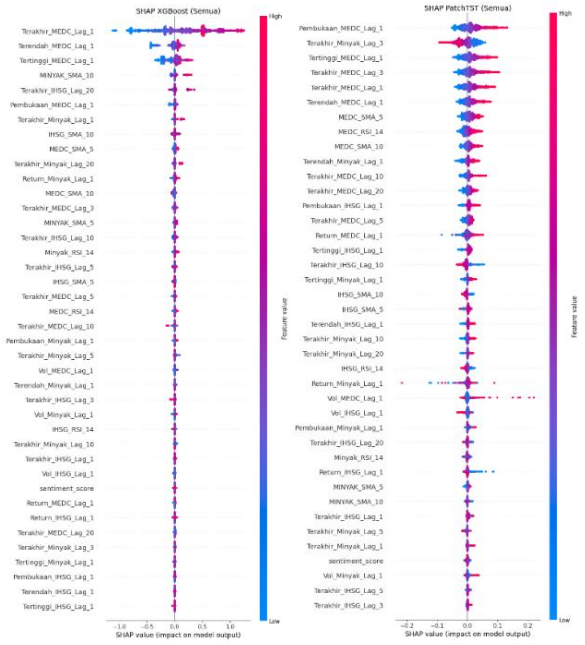
Analisis kontribusi fitur dilakukan dengan Mutual Information (MI), menunjukkan bahwa Terakhir_MEDC_Lag_1 memiliki nilai tertinggi, diikuti oleh beberapa fitur historis MEDC seperti Tertinggi_MEDC_Lag_1, Terendah_MEDC_Lag_1, Pembukaan_MEDC_Lag_1, dan MEDC_SMA_5. Hasil tersebut menunjukkan bahwa fitur historis harga MEDC memiliki hubungan informasi yang kuat dengan target dan selanjutnya menjadi dasar untuk melihat apakah pola dominasi fitur tersebut tetap tercermin dalam *feature importance* dan SHAP pada masing-masing model.



Gambar 3. Hasil Feature Importance

Hasil *feature importance* pada Gambar 3 dapat dilihat jika XGBoost menunjukkan bahwa Terakhir_MEDC_Lag_1 merupakan fitur yang paling dominan. Beberapa fitur historis MEDC lainnya juga memiliki tingkat kepentingan yang relatif tinggi, sedangkan fitur yang berasal dari IHSG dan harga minyak memiliki tingkat kepentingan yang lebih rendah. Hasil ini menunjukkan adanya kesinambungan dengan MI, karena Terakhir_MEDC_Lag_1 tetap menjadi fitur dominan setelah digunakan dalam model XGBoost. XGBoost cenderung lebih banyak memanfaatkan informasi historis harga MEDC dalam menghasilkan prediksi.

Pada PatchTST, Terakhir_MEDC_Lag_1 juga termasuk fitur dengan tingkat kepentingan yang tinggi. Hal ini menunjukkan adanya kesinambungan dengan hasil MI, karena fitur yang memiliki hubungan informasi tertinggi dengan target pada tahap *feature selection* tetap menjadi salah satu fitur penting ketika digunakan dalam model. Namun, berbeda dengan XGBoost, kontribusi fitur pada PatchTST lebih tersebar pada beberapa variabel. Selain fitur historis MEDC, fitur IHSG dan harga minyak juga menunjukkan tingkat kepentingan dalam model. Hal ini menunjukkan bahwa PatchTST tidak hanya memanfaatkan hubungan informasi fitur historis MEDC yang teridentifikasi pada tahap MI, tetapi juga memanfaatkan informasi dari variabel eksternal dalam proses prediksi.



Gambar 4. Analisis SHAP

Hasil SHAP menunjukkan bahwa kontribusi fitur tidak hanya terlihat dari tingkat kepentingannya, tetapi juga dari arah pengaruhnya terhadap output prediksi. Pada XGBoost, Terakhir_MEDC_Lag_1 menjadi fitur yang paling dominan, dengan nilai SHAP positif menunjukkan kecenderungan mendorong prediksi ke nilai yang lebih tinggi, sedangkan nilai SHAP negatif menunjukkan kecenderungan menurunkan prediksi. Pola serupa terlihat pada fitur historis MEDC lainnya. Pada PatchTST, fitur historis MEDC tetap memiliki kontribusi besar, sementara fitur IHSG dan harga minyak juga menunjukkan kontribusi terhadap output prediksi melalui distribusi nilai SHAP positif dan negatif.

Jika dibandingkan berdasarkan urutan fitur, hasil *feature selection* menggunakan MI, *feature importance*, dan SHAP tidak menunjukkan urutan yang sepenuhnya sama. Perbedaan tersebut berkaitan dengan tujuan masing-masing metode. MI digunakan pada tahap *feature selection* untuk mengukur hubungan informasi antara fitur dan target sebelum pemodelan, sedangkan *feature importance* dan SHAP digunakan setelah model dilatih untuk menganalisis penggunaan dan kontribusi fitur terhadap prediksi. Meskipun terdapat perbedaan urutan, Terakhir_MEDC_Lag_1 menunjukkan konsistensi sebagai fitur yang dominan. Sementara itu, sentiment_score memiliki kontribusi yang relatif rendah dibandingkan fitur historis MEDC dan variabel pasar lainnya.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa konfigurasi FS8 memberikan kinerja terbaik pada XGBoost dan PatchTST dalam prediksi harga saham MEDC. Pada konfigurasi tersebut, XGBoost menghasilkan RMSE sebesar 90,89, MAE sebesar 67,69, MAPE sebesar 6,82%, dan R^2 sebesar 0,6184, sedangkan PatchTST menghasilkan RMSE sebesar 53,56, MAE sebesar 41,20, MAPE sebesar 5,19%, dan R^2 sebesar 0,8710. Secara matriks, PatchTST FS8 menghasilkan nilai error yang lebih rendah dan R^2 yang lebih tinggi dibandingkan XGBoost FS8. Selain itu, standar deviasi dan nilai *coefficient of variation* (CV) PatchTST FS8 yang lebih rendah pada sebagian besar metrik menunjukkan variasi kinerja antar-*fold* yang lebih kecil. Namun, hasil uji *Wilcoxon* dengan koreksi *Bonferroni* menunjukkan bahwa perbedaan kinerja antara XGBoost dan PatchTST tidak signifikan secara statistik. Dengan demikian, hasil penelitian menunjukkan adanya perbedaan kinerja secara deskriptif, tetapi belum memberikan bukti statistik yang cukup untuk menyatakan bahwa salah satu model memiliki kinerja yang berbeda secara signifikan dari model lainnya.

Analisis kontribusi fitur menunjukkan adanya kesinambungan antara hasil Mutual Information (MI) dan interpretasi model, terutama pada fitur Terakhir_MEDC_Lag_1 yang menunjukkan dominasi secara konsisten. Pada XGBoost, kontribusi fitur lebih banyak didominasi oleh fitur historis harga MEDC, sedangkan PatchTST menunjukkan distribusi kontribusi yang lebih tersebar dengan turut memanfaatkan informasi dari IHSG dan harga minyak mentah WTI. Sementara itu, sentiment_score menunjukkan kontribusi yang relatif rendah pada kedua model. Hasil tersebut menunjukkan bahwa informasi historis harga MEDC menjadi informasi penting dalam prediksi, sementara variabel eksternal memberikan kontribusi tambahan yang berbeda pada masing-masing model.

5. SARAN

Penelitian selanjutnya dapat menambahkan variabel eksternal, seperti nilai tukar, inflasi, suku bunga, dan indikator makroekonomi lainnya. Pengujian juga dapat diperluas dengan model Machine Learning dan Deep Learning lainnya, termasuk model hybrid, serta optimasi hyperparameter yang lebih mendalam. Selain itu, penggunaan multi-seed dapat dilakukan untuk mengevaluasi kestabilan model terhadap variasi inisialisasi acak. Pengembangan selanjutnya juga dapat diarahkan pada prediksi arah pergerakan harga saham untuk menghasilkan informasi yang lebih mudah diinterpretasikan.

Daftar Pustaka

[1] X. Ji, J. Wang, and Z. Yan, "A stock price prediction method based on deep learning technology," International Journal of Crowd Science, vol. 5, no. 1, pp. 55–72, Apr. 2021, doi: 10.1108/IJCS-05-2020-0012.

[2] K. Kwanda, D. E. Herwindiati, M. D. Lauro, and K. K. C. Id, "Perbandingan LSTM dan Bidirectional LSTM pada Sistem Prediksi Harga Saham Berbasis Website," R2J, vol. 7, no. 1, 2024, doi: 10.38035/rrj.v7i1.

[3] I. Karijadi, Y. Widjaja, D. Retno, S. Dewi, and L. J. Asrini, "Studi Penerapan Deep Learning untuk Prediksi Pergerakan Indeks Harga Saham Gabungan," Jurnal Impresi Indonesia (JII), vol. 4, no. 12, 2025.

[4] K. N. Hermawan, "Pengaruh Harga Minyak Mentah dan Struktur Modal terhadap Profitabilitas Perusahaan Sektor Energi yang Terdaftar di Indeks Saham Syariah Indonesia (ISSI)," Indonesian Journal of Economics, vol. 1, no. 9, 2024.

[5] S. Mazya Permataning Tyas, R. Sarno, and B. Setya Rintyarna, "Analisis Perbandingan Metode Klasifikasi Sentimen Berita Saham: Pendekatan Machine Learning, Deep Learning, Transfer Learning, dan Graf," 2024.

[6] L. Hu et al., "Supervised Machine Learning Techniques: An Overview with Applications to Banking Corporate Model Risk, Wells Fargo," International Statistical Review, vol. 89, pp. 573–604, 2020, doi: 10.1111/insr.12448.

[7] Q. Wen et al., "Transformers in Time Series: A Survey," in Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23), 2023. doi: 10.24963/ijcai.2023/759.

[8] S. Yongchareon, "AI-Driven Intelligent Financial Forecasting: A Comparative Study of Advanced Deep Learning Models for Long-Term Stock Market Prediction," Mach. Learn. Knowl. Extr., vol. 7, no. 3, Sep. 2025, doi: 10.3390/make7030061.

[9] J. M. A. S. Dachi and P. Sitompul, "Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit," JURNAL RISET RUMPUN MATEMATIKA DAN ILMU PENGETAHUAN ALAM, vol. 2, no. 2, pp. 65–71, Jul. 2023, doi: 10.55606/jurrimipa.v2i2.1336.

[10] S. Rohman, I. Wiseto, and P. Agung, "KOMPARASI ALGORITMA INDOBERT, SVM, DAN RANDOM FOREST DALAM ANALISIS SENTIMEN PROGRAM MAKAN BERGIZI GRATIS," 2025.

[11] F. Alvin, N. Anisa, and S. Winarsih, "Perbandingan Kinerja Model IndoBERT, IndoBERTweet, dan Algoritma Klasik pada Analisis Sentimen Isu Indonesia Gelap," Technology and Science (BITS), vol. 7, no. 3, pp. 1601–1613, 2025, doi: 10.47065/bits.v7i3.8636.

[12] S. Rohman, I. Wiseto, and P. Agung, "KOMPARASI ALGORITMA INDOBERT, SVM, DAN RANDOM FOREST DALAM ANALISIS SENTIMEN PROGRAM MAKAN BERGIZI GRATIS," 2025.

[13] A. R. Aisy and G. Karyono, "Analisis Sentimen Terhadap Ulasan Google Play Store Aplikasi Lazada, Shopee, dan Tokopedia Menggunakan Algoritma IndoBERT," Technology and Science (BITS), vol. 7, no. 3, 2025, doi: 10.47065/bits.v7i3.8745.

[14] S. Alam, Z. Suryady, and A. Kusnadi, "Fine-Tuning IndoBERT Untuk Analisis Sentimen Berita Saham Berbahasa Indonesia Dengan Hyperparameter Optimization," JTERA (Jurnal Teknologi Rekayasa), vol. 11, no. 1, Jun. 2026, doi: 10.31544/jtera.v11.i1.2026.51-60.

[15] N. N. Azizah, W. D. Utami, and A. Fanani, "Algoritma XGBoost Untuk Analisis Prediksi Permintaan Kalibrasi di PT PAL Indonesia," JIKO (Jurnal Informatika dan Komputer), vol. 10, no. 1, p. 14, Feb. 2026, doi: 10.26798/jiko.v10i1.2245.

[16] B. Jange, "Prediksi Harga Saham Bank BCA Menggunakan XGBoost," ARBITRASE: Journal of Economics and Accounting, vol. 3, no. 2, pp. 231–237, Nov. 2022, doi: 10.47065/arbitrase.v3i2.495.

[17] L. Huang, X. Zhou, L. Shi, and L. Gong, "Time Series Feature Selection Method Based on Mutual Information," Applied Sciences (Switzerland), vol. 14, no. 5, Mar. 2024, doi: 10.3390/app14051960.

[18] M. Hisyam, Z. Fitri, and H. A. K. Aidilof, "Implementasi Algoritma XGBoost dengan Walk Forward Validation untuk Prediksi Harga Emas Antam," JURIKOM (Jurnal Riset Komputer), vol. 12, no. 4, pp. 403–413, Aug. 2025, doi: 10.30865/jurikom.v12i4.8693.

[19] H. Hewamalage, K. Ackermann, and C. Bergmeir, "Forecast evaluation for data scientists: common pitfalls and best practices," Data Min. Knowl. Discov., vol. 37, no. 2, pp. 788–832, Mar. 2023, doi: 10.1007/s10618-022-00894-5.

[20] E. Fortuna, D. Arman Prasetya, and K. Maulida Hindrayani, "X-Means Clustering for Segmenting Property Customer Payment Behaviors," Journal of Information System and Technology Research journal homepage, vol. 5, no. 1, 2026, doi: 10.55537/jistr.v5i1.1228.

[21] M. A. Rayadin, M. Musaruddin, R. A. Saputra, and I. Isnawaty, "Implementasi Ensemble Learning Metode XGBoost dan Random Forest untuk Prediksi Waktu Penggantian Baterai Aki," BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer, vol. 5, no. 2, pp. 111–119, Aug. 2024, doi: 10.37148/bios.v5i2.128.

[22] M. Syukron, "A Patch-Based Transformer Approach to Nonlinear Dynamics Natural Gas Price Forecasting," (IJCSAM) INTERNATIONAL JOURNAL OF COMPUTING SCIENCE AND APPLIED MATHEMATICS, vol. 11, pp. 76–82, Dec. 2025, doi: 10.12962/j24775401.ijcsam.v11i2.8849.

[23] G. Mo, W. Jia, C. Tan, W. Zhang, and J. Rong, "Risk Prediction of International Stock Markets with Complex Spatio-Temporal Correlations: A Spatio-Temporal Graph Convolutional Regression Model Integrating Uncertainty Quantification," vol. 18, p. 488, 2025, doi: 10.3390/jrfm18090488.

[24] S. F. Stumpfe and S. C. Shongwe, "Comparative Analysis and Optimisation of Machine Learning Models for Regression and Classification on Structured Tabular Datasets," Mathematics, vol. 14, no. 3, Feb. 2026, doi: 10.3390/math14030473.

[25] J. Wainer, "A Bayesian Bradley-Terry model to compare multiple ML algorithms on multiple data sets," Journal of Machine Learning Research, vol. 24, pp. 1–34, 2023, doi: 10.48550/arXiv.2208.04935.

[26] K. Wallapure Manikrao, S. Allagi, W. Y. Leong, and M. Laddi, "An Explainable AI for Stock Market Prediction: A Machine Learning Approach with XAI and Deep Neural Networks," Journal of Computational and Cognitive Engineering, Dec. 2025, doi: 10.47852/bonviewjcce52026428.