



Prediksi Arah Harga Cryptocurrency Menggunakan Hybrid Lstm Encoder dan Xgboost Head: Implementasi dan Evaluasi pada 10 Aset Digital Utama

Ade Kurniawan¹, Mulyono², Hari Suriadi³

¹Program Studi Magister Teknik Informatika, Fakultas Ilmu Komputer, Universitas Putra Indonesia "YPTK" Padang, Indonesia

²Program Studi Teknik Elektro, Fakultas Teknik Universitas Muhammadiyah Sumatera Barat

³Universitas Muhammadiyah Sumatera Barat, Padang, Indonesia

Email: ade.kurniawan.a17@gmail.com, mulyono94@gmail.com, suriadihari6@gmail.com³

ABSTRACT

The primary challenge in cryptocurrency price prediction lies in the market's highly volatile, non-linear, and near-random walk behavior, which makes traditional predictive models unable to achieve consistent accuracy. This study aims to develop and evaluate a hybrid model combining Long Short-Term Memory (LSTM) and XGBoost to predict price direction and returns for ten major cryptocurrencies using daily data from 2023 to 2025. Historical data were processed through feature engineering, normalization, and sliding-window sequence construction, and the models were evaluated using TimeSeriesSplit to prevent data leakage. The results show that the hybrid model consistently outperformed both LSTM and XGBoost, achieving an average directional accuracy of 58.6%, significantly higher than the baselines (51.7% for LSTM and 53.6% for XGBoost). The average RMSE of 0.0289 indicates stable return predictions without systematic bias. Statistical validation through paired t-tests and McNemar tests confirmed the significance of the improvement at $p < 0.001$. A trading simulation using a 1-day holding period produced an annualized return of 41.5% with a Sharpe ratio of 1.12, outperforming the buy-and-hold strategy. These findings highlight that integrating LSTM's temporal representation with XGBoost's non-linear learning capabilities is an effective and computationally efficient approach for cryptocurrency price forecasting, offering practical value for the development of algorithmic trading systems.

Keywords: *Cryptocurrency; Price Prediction; Hybrid Model; LSTM; XGBoost; Algorithmic Trading*

ABSTRAK

Permasalahan utama dalam prediksi harga cryptocurrency terletak pada sifat pasar yang sangat volatil, non-linear, dan sering mendekati *random walk*, sehingga model prediksi tradisional sulit menghasilkan akurasi yang stabil. Penelitian ini bertujuan mengembangkan dan mengevaluasi model hybrid berbasis Long Short-Term Memory (LSTM) dan XGBoost untuk memprediksi arah harga dan return pada sepuluh aset kripto berkapitalisasi besar menggunakan data harian periode 2023–2025. Data historis diolah melalui serangkaian tahapan *feature engineering*, normalisasi, dan pembentukan jendela sekuens, kemudian model dianalisis menggunakan *TimeSeriesSplit* untuk menghindari *data leakage*. Hasil penelitian menunjukkan bahwa model hybrid secara konsisten mengungguli LSTM dan XGBoost dengan akurasi prediksi arah harga rata-rata 58,6%, meningkat signifikan atas baseline (51,7% pada LSTM dan 53,6% pada XGBoost). Selain itu, RMSE return rata-rata 0,0289 menunjukkan bahwa prediksi model stabil dan tidak mengandung *systematic bias*. Uji *paired t-test* dan *McNemar test* membuktikan

peningkatan performa signifikan pada tingkat $p < 0.001$. Simulasi strategi perdagangan *1-day holding period* menghasilkan return tahunan 41,5% dengan Sharpe ratio 1,12, melampaui strategi *buy-and-hold*. Temuan ini menunjukkan bahwa integrasi kemampuan temporal LSTM dan fleksibilitas XGBoost merupakan pendekatan yang efektif dan efisien untuk prediksi harga cryptocurrency, sekaligus memberikan kontribusi praktis bagi pengembangan sistem perdagangan algoritmik.

Kata kunci: Cryptocurrency; Prediksi Harga; Model Hybrid; LSTM, XGBoost

PENDAHULUAN

Pasar cryptocurrency telah berkembang menjadi ekosistem finansial global yang sangat dinamis, ditandai oleh pertumbuhan kapitalisasi pasar yang melampaui 2 triliun dolar serta meningkatnya partisipasi investor institusional dan retail (Corbet et al., 2018). Pergerakan harga cryptocurrency yang berlangsung 24 jam tanpa batasan *trading halt* dan tanpa regulasi pasar yang ketat menjadikannya salah satu jenis aset paling volatil. Karakteristik tersebut menghadirkan tantangan signifikan bagi para analis, peneliti, dan pelaku industri dalam melakukan prediksi harga yang akurat dan stabil.

Sejumlah penelitian awal pada aset kripto mengindikasikan bahwa perubahan harga cryptocurrency cenderung mengikuti pola *random walk*, mirip dengan pergerakan harga saham yang sangat acak (Nadarajah & Chu, 2017). Hal ini menyebabkan model statistik konvensional seperti ARIMA, GARCH, atau VAR seringkali tidak mampu menangkap kompleksitas struktural dan dinamika non-linear yang melekat pada aset digital tersebut. Volatilitas yang tinggi, regime shift, serta noise pasar yang intens membuat prediksi harga kripto menjadi lebih menantang dibandingkan aset keuangan tradisional.

Dalam upaya mengatasi keterbatasan model statistik, pendekatan machine learning dan deep learning mulai banyak digunakan dalam penelitian keuangan, khususnya pada prediksi harga cryptocurrency. Model Long Short-Term Memory (LSTM) menjadi salah satu model favorit karena kemampuannya menangani ketergantungan jangka panjang pada data runtun waktu (Hochreiter & Schmidhuber, 1997). Namun, sejumlah penelitian menunjukkan bahwa LSTM masih memiliki kelemahan dalam menangkap pola non-linear yang tidak sepenuhnya bergantung pada struktur temporal dan sering mengalami *overfitting* pada data dengan volatilitas tinggi.

Di sisi lain, model tree-based seperti Extreme Gradient Boosting (XGBoost) terbukti sangat efektif untuk data tabular dan mampu mengekstraksi interaksi non-linear antar variabel secara lebih stabil (Chen & Guestrin, 2016). Model ini unggul dalam menghasilkan prediksi yang kuat dan efisien, tetapi tidak dirancang untuk belajar dari hubungan temporal antar observasi secara langsung. Karena itu, penggunaan XGBoost secara tunggal dalam prediksi harga kripto juga belum mencapai performa optimal.

Keterbatasan model tunggal tersebut melahirkan kebutuhan akan pendekatan hybrid yang mampu menggabungkan kelebihan LSTM dalam mengekstraksi informasi temporal dengan kekuatan XGBoost dalam menangani struktur non-linear. Pendekatan *stacked generalization* (Wolpert, 1992) menyediakan dasar teoritis bahwa kombinasi beberapa model heterogen dapat menghasilkan prediksi yang lebih akurat dibandingkan model tunggal. Meskipun demikian, implementasi hybrid semacam ini masih jarang diterapkan secara komprehensif pada pasar cryptocurrency dengan jumlah aset yang beragam.

Kajian literatur menunjukkan beberapa kelemahan penelitian sebelumnya. Pertama, sebagian besar studi hybrid LSTM dan XGBoost hanya dilakukan pada satu aset, terutama

Bitcoin, sehingga belum menguji generalisasi model pada aset lain yang memiliki karakteristik volatilitas berbeda (Silvanus et al., 2022). Kedua, banyak penelitian masih menggunakan teknik validasi data yang keliru, seperti *random k-fold*, yang dapat menyebabkan *data leakage* dan memberikan estimasi performa yang tidak akurat (Bergmeir et al., 2018). Ketiga, sebagian studi tidak menyertakan pengujian statistik formal sehingga peningkatan performa model tidak dapat dipastikan signifikan secara ilmiah.

Selain itu, penelitian-penelitian sebelumnya belum banyak menyediakan kerangka implementasi end-to-end yang dapat direplikasi, terutama terkait penyusunan pipeline data, pemilihan fitur teknis, pembentukan *sliding window*, dan integrasi LSTM encoder dengan XGBoost meta-learner. Padahal, kebutuhan akan kerangka implementasi yang jelas sangat penting untuk mendukung penelitian lanjutan dan aplikasi industri, khususnya dalam pengembangan *automated trading systems*.

Di tengah kebutuhan tersebut, penelitian ini mengusulkan arsitektur hybrid dua tahap yang memanfaatkan LSTM sebagai *temporal encoder* untuk mengekstraksi representasi laten dari jendela sekuen harga 60 hari, kemudian dikombinasikan dengan indikator teknis seperti RSI, MACD, dan Bollinger %B sebelum diproses oleh XGBoost sebagai *meta-learner*. Pendekatan ini diharapkan mampu mengatasi keterbatasan masing-masing model dan menghasilkan prediksi arah harga yang lebih akurat dan stabil.

Penelitian ini juga menerapkan validasi *TimeSeriesSplit* dengan skema *forward chaining* untuk memastikan tidak terjadi *look-ahead bias* serta memperkuat validitas prediksi pada data *unseen*. Selain itu, evaluasi dilakukan pada sepuluh cryptocurrency utama berdasarkan kapitalisasi pasar, sehingga hasil penelitian ini tidak hanya menunjukkan performa model secara spesifik pada Bitcoin, tetapi juga konsistensi lintas aset dengan volatilitas berbeda.

Dengan demikian, penelitian ini berkontribusi pada dua aspek utama: (1) kontribusi metodologis berupa model hybrid yang terstruktur, dapat direplikasi, dan diuji secara statistik; serta (2) kontribusi empiris berupa temuan konsisten pada berbagai aset cryptocurrency. Kedua kontribusi ini diharapkan dapat memberikan dasar kuat bagi pengembangan penelitian selanjutnya serta memberikan manfaat praktis bagi pelaku industri yang mengembangkan sistem perdagangan otomatis berbasis machine learning.

TINJAUAN PUSTAKA

Perkembangan penelitian mengenai prediksi harga aset keuangan menggunakan machine learning dan deep learning terus meningkat seiring dengan kompleksitas pasar keuangan modern. Model Long Short-Term Memory (LSTM), yang diperkenalkan oleh Hochreiter dan Schmidhuber (1997), menjadi landasan utama dalam prediksi runtun waktu karena kemampuannya mengatasi *vanishing gradient* dan menangkap ketergantungan jangka panjang. Dalam berbagai studi keuangan, LSTM telah menunjukkan performa yang menjanjikan, khususnya dalam memodelkan pola harga saham dan cryptocurrency yang bersifat non-linear. Namun, sejumlah penelitian menyebutkan bahwa LSTM sangat sensitif terhadap hiperparameter, rawan *overfitting*, serta kurang stabil ketika diterapkan pada data dengan volatilitas tinggi, seperti pasar kripto.

Pada sisi lain, penelitian mengenai model tree-based seperti XGBoost menunjukkan perkembangan signifikan dalam pemrosesan data tabular. Chen dan Guestrin (2016) menegaskan bahwa XGBoost unggul dalam pembelajaran non-linear secara efisien melalui mekanisme regularisasi dan *gradient boosting*. Dalam konteks keuangan, model ini telah sukses digunakan untuk mengidentifikasi pola teknis dan faktor pasar karena kemampuannya

mengolah interaksi variabel secara kompleks. Meskipun demikian, kekurangan utama XGBoost adalah ketidakmampuannya menangkap hubungan temporal secara eksplisit sehingga performanya terbatas pada data yang sangat bergantung pada urutan waktu.

Penelitian mengenai ensemble learning menggambarkan bahwa penggabungan beberapa model yang memiliki karakteristik berbeda seringkali menghasilkan prediksi yang lebih akurat. Wolpert (1992) melalui konsep *stacked generalization* memperkenalkan pendekatan meta-learning yang memungkinkan model tingkat atas (meta-learner) mempelajari pola dari hasil keluaran model dasar. Penelitian lanjutan oleh Kuncheva (2014) serta Zhou (2012) memperkuat bahwa keberagaman model dalam ensemble, baik dari struktur maupun mekanisme pembelajaran, berkontribusi signifikan terhadap peningkatan kinerja prediktif. Hal ini menjadi landasan teoritis yang kuat untuk mengembangkan hybrid model yang menggabungkan LSTM dan XGBoost.

Dalam konteks cryptocurrency, sejumlah penelitian telah mengkaji performa model deep learning dan machine learning secara individual. Wang, Wang, dan Zhang (2020) melaporkan bahwa LSTM mampu menghasilkan RMSE rendah pada prediksi Bitcoin, tetapi akurasi arah harga masih terbatas sekitar 54%. Penelitian Permentier et al. (2024) menggunakan kombinasi beberapa algoritma machine learning namun hanya menghasilkan akurasi 52–56% pada lima aset kripto. Sementara itu, Silvanus et al. (2022) mengusulkan model hybrid LSTM–XGBoost pada Bitcoin dan berhasil menurunkan RMSE sekitar 15% dibanding LSTM tunggal, namun penelitian tersebut tidak menguji signifikansi statistik dan tidak mengevaluasi performa pada aset lain.

Gap penelitian semakin terlihat ketika beberapa studi menunjukkan bahwa sebagian besar penelitian terdahulu hanya berfokus pada satu aset (umumnya Bitcoin), sehingga tidak memberikan gambaran generalisasi model pada aset lain dengan karakteristik volatilitas berbeda. Selain itu, banyak penelitian tidak menggunakan metode validasi time series yang tepat. Bergmeir, Hyndman, dan Koo (2018) menegaskan bahwa penggunaan *random k-fold cross validation* pada time series dapat memicu *data leakage*, membuat hasil evaluasi menjadi bias dan tidak mencerminkan kemampuan model pada data nyata. Dengan demikian, validasi berbasis *forward chaining* menjadi syarat penting yang sering diabaikan dalam penelitian terdahulu.

Penelitian-penelitian terbaru pada area deep learning juga menunjukkan tren penggunaan model lebih kompleks, seperti Transformer dan attention mechanism. Zhang et al. (2023) menemukan bahwa ensemble deep learning mampu mencapai akurasi 60–62% pada prediksi cryptocurrency, tetapi dengan biaya komputasi yang sangat tinggi. Penelitian Dantas et al. (2022) menunjukkan bahwa hybrid neural network ensemble dapat meningkatkan akurasi, tetapi seringkali sulit direplikasi dan tidak dilengkapi dokumentasi metodologis yang memadai. Hal ini mempertegas pentingnya model yang tidak hanya akurat, tetapi juga efisien dan dapat diimplementasikan secara praktis.

Berdasarkan kajian literatur tersebut, dapat disimpulkan bahwa meskipun telah banyak penelitian dilakukan, terdapat beberapa kekurangan utama: (1) model tunggal seperti LSTM atau XGBoost masih terbatas dalam akurasi prediksi; (2) pendekatan hybrid belum diterapkan secara sistematis pada banyak aset; (3) validasi time series yang benar sering diabaikan; (4) sebagian penelitian tidak menyertakan pengujian statistik formal; dan (5) kerangka implementasi reproducible masih jarang tersedia. Oleh karena itu, penelitian ini menempatkan dirinya sebagai upaya untuk mengisi celah tersebut dengan mengembangkan model hybrid

LSTM–XGBoost yang dievaluasi secara komprehensif dan konsisten pada sepuluh aset kripto menggunakan metodologi validasi yang tepat dan analisis statistik yang kuat.

METODOLOGI

Penelitian ini menggunakan pendekatan kuantitatif dengan desain prediktif untuk membangun model machine learning yang mampu memproyeksikan arah harga cryptocurrency berdasarkan data historis. Pemilihan pendekatan prediktif didasarkan pada tujuan penelitian yang fokus pada pengembangan dan evaluasi model hybrid berbasis LSTM dan XGBoost. Penelitian dilakukan secara eksperimental dengan membandingkan performa model baseline dan hybrid untuk menentukan kelayakan model yang diusulkan.

Data yang digunakan dalam penelitian ini berasal dari platform Yahoo Finance, yang menyediakan data harga cryptocurrency secara terbuka. Sepuluh aset utama dipilih berdasarkan kapitalisasi pasar tertinggi, yaitu BTC, ETH, BNB, SOL, XRP, ADA, DOGE, TON, TRX, dan DOT. Periode data yang dianalisis mencakup November 2023 hingga November 2025 dengan interval harian. Setiap aset memiliki sekitar 730 observasi sehingga total dataset mencapai lebih dari 7.000 data poin. Seluruh data diperiksa untuk memastikan tidak adanya *missing value* atau anomali yang dapat memengaruhi kinerja model.

Proses pengolahan data dilakukan dalam beberapa tahapan utama. Tahap pertama adalah menghitung return dan log-return untuk meminimalkan pengaruh skala harga. Indikator teknikal seperti RSI, MACD, dan Bollinger %B dihitung untuk memperkaya atribut model karena indikator-indikator ini terbukti relevan dalam penelitian teknikal sebelumnya. Selanjutnya, seluruh fitur dinormalisasi menggunakan MinMax scaling untuk memastikan stabilitas pelatihan model LSTM. Tahap terakhir adalah membentuk jendela sekuens sepanjang 60 hari untuk menangkap pola temporal sebagai input model.

Model hybrid dibangun melalui dua komponen inti. LSTM digunakan sebagai *encoder* untuk menghasilkan representasi laten dari pola harga pada jendela waktu tertentu. Representasi temporal ini kemudian digabungkan dengan fitur teknikal dan volume, lalu diproses oleh XGBoost sebagai *meta-learner* untuk menentukan arah harga hari berikutnya. Pendekatan ini menggabungkan keunggulan LSTM dalam memahami dependensi waktu dan keunggulan XGBoost dalam mengekstraksi hubungan non-linear antar variabel, sesuai dengan konsep *stacked generalization* yang diperkenalkan Wolpert (1992).

Validasi model dilakukan menggunakan teknik *TimeSeriesSplit* dengan lima bagian untuk memastikan bahwa data uji selalu berada setelah data latih dalam urutan waktu. Teknik ini direkomendasikan dalam analisis time series untuk menghindari *data leakage* dan meningkatkan kemampuan generalisasi model (Bergmeir et al., 2018). Evaluasi dilakukan menggunakan beberapa metrik seperti akurasi, precision, recall, F1-score, serta RMSE untuk melihat kualitas prediksi return. Untuk memastikan peningkatan performa signifikan secara statistik, digunakan uji paired t-test dan McNemar test.

Penelitian ini tidak melibatkan manusia ataupun data pribadi sehingga tidak memerlukan izin etik khusus. Seluruh proses penelitian dilakukan dengan memperhatikan prinsip transparansi ilmiah, replikasi, serta penggunaan sumber data terbuka sehingga hasil penelitian dapat diuji ulang oleh peneliti lain.

RESULTS AND DISCUSSION / HASIL DAN PEMBAHASAN

Untuk memperoleh pemahaman menyeluruh mengenai efektivitas model yang dikembangkan, penelitian ini mengevaluasi performa LSTM, XGBoost, dan model hybrid pada

sepuluh aset cryptocurrency menggunakan data uji dengan pendekatan *forward-chaining*. Evaluasi dilakukan melalui akurasi prediksi arah harga, RMSE prediksi return, serta metrik klasifikasi seperti precision, recall, dan F1-score yang merupakan indikator penting dalam memvalidasi performa model prediktif (Chollet, 2018; Goodfellow et al., 2016). Ringkasan hasil evaluasi tersebut disajikan pada Tabel 1.

Tabel 1 Performa Komprehensif Model Per Cryptocurrency (Test Set)

Cryptocurrency	LSTM Acc (%)	XGB Acc (%)	Hybrid Acc (%)	RMSE Reg	CV Mean (%)	Precision	Recall	F1- Score
BTC-USD	51.2	53.8	58.6	0.0261	57.1	0.61	0.65	0.63
ETH-USD	52.1	54.2	59.3	0.0287	58.9	0.63	0.67	0.65
BNB-USD	50.8	52.6	56.4	0.0342	55.2	0.57	0.61	0.59
SOL-USD	53.0	54.9	61.2	0.0198	61.0	0.65	0.71	0.68
XRP-USD	51.5	53.1	57.8	0.0305	56.9	0.59	0.63	0.61
ADA-USD	50.6	51.9	55.1	0.0368	53.8	0.56	0.59	0.57
DOGE-USD	52.3	54.5	60.1	0.0275	59.5	0.64	0.68	0.66
TON-USD	51.9	53.7	58.9	0.0298	57.4	0.60	0.64	0.62
TRX-USD	52.6	55.1	62.3	0.0225	61.8	0.66	0.72	0.69
DOT-USD	50.9	52.4	56.7	0.0331	55.3	0.58	0.62	0.60
Mean	51.7	53.6	58.6	0.0289	57.7	0.61	0.65	0.63

Hasil pada Tabel 1 menunjukkan bahwa model hybrid menghasilkan performa prediksi yang secara konsisten lebih unggul dibandingkan model baseline. Secara rata-rata, akurasi model hybrid mencapai 58,6%, meningkat signifikan dibandingkan LSTM (51,7%) dan XGBoost (53,6%). Peningkatan sebesar 6,9 poin persentase dibandingkan LSTM dan 5 poin persentase dibandingkan XGBoost ini menegaskan pentingnya kombinasi arsitektur berbasis *temporal deep learning* dan *gradient-boosted decision trees*. Temuan ini sejalan dengan studi sebelumnya yang menyatakan bahwa integrasi model deep learning dan tree-based learners dapat meningkatkan kemampuan generalisasi dan mengurangi kesalahan prediksi pada data berderet waktu yang kompleks (Zhang et al., 2023; Kim & Kang, 2020).

Tingkat akurasi di atas 58% dapat dikategorikan tinggi untuk pasar cryptocurrency, mengingat banyak studi menunjukkan bahwa pergerakan harga aset kripto sering kali mendekati sifat *random walk* (Urquhart, 2016; Corbet et al., 2019). Meskipun demikian, model hybrid mampu menangkap pola non-linear dan dependensi temporal yang tidak dapat dipahami sepenuhnya oleh model tradisional. Dominasi kinerja hybrid semakin jelas pada aset volatil seperti TRX (62,3%) dan SOL (61,2%), di mana pola *volatility clustering* lebih menonjol, sehingga jaringan LSTM dapat mengekstraksi sinyal temporal yang lebih kaya. Sebaliknya, performa menurun pada aset yang memiliki volatilitas relatif lebih rendah seperti ADA dan DOT. Temuan ini menunjukkan bahwa volatilitas menjadi determinan penting dalam efektivitas model prediktif, sebagaimana juga ditemukan dalam penelitian sebelumnya (Atsalakis & Valavanis, 2009).

Dari perspektif prediksi return, nilai RMSE rata-rata 0,0289 memperlihatkan bahwa prediksi model hybrid relatif stabil dan tidak menunjukkan *systematic drift*. Nilai RMSE yang berada di kisaran <0,03 ini menunjukkan bahwa model mampu mengekstraksi pola jangka

pendek dari dinamika harga tanpa mengalami *overfitting*, terutama karena XGBoost menjaga regulasi melalui pengendalian kompleksitas pohon (Chen & Guestrin, 2016). Hasil ini mendukung pandangan bahwa penggabungan model sekuensial dan model pohon penguat dapat menghasilkan estimasi kuantitatif yang lebih baik dalam pasar yang sangat dinamis (Fischer & Krauss, 2018).

Untuk memvalidasi signifikansi peningkatan kinerja, penelitian ini menggunakan *paired t-test* dan *McNemar test*. Uji *paired t-test* menghasilkan nilai $t = 18.2$ dengan $p < 0.001$, yang menunjukkan bahwa peningkatan akurasi model hybrid tidak terjadi secara kebetulan. McNemar test menunjukkan hasil yang konsisten, yaitu Hybrid vs LSTM ($\chi^2 = 127.4$; $p < 0.001$) dan Hybrid vs XGBoost ($\chi^2 = 89.3$; $p < 0.001$). Hasil ini memperkuat bukti bahwa evaluasi empiris tidak hanya menunjukkan perbedaan performa, tetapi juga signifikan secara statistik, sesuai rekomendasi validasi model dalam prediksi keuangan (De Prado, 2018).

Analisis *feature importance* mengungkap bahwa *latent vector* dari LSTM memberikan kontribusi sekitar 45% terhadap total bobot fitur, menunjukkan bahwa pola temporal jangka pendek memainkan peran kunci dalam keberhasilan prediksi. Hal ini relevan dengan teori bahwa pasar keuangan memiliki struktur dependensi waktu, terutama pada jangka pendek, di mana dinamika momentum dan pembalikan tren memiliki efek substansial (Lo, 2004). Indikator teknikal seperti RSI memberikan kontribusi 18%, mendukung literatur yang menunjukkan efektivitas indikator momentum dalam pasar volatil (Bollinger, 2002). Volume perdagangan yang menyumbang ~11,7% juga menunjukkan bahwa likuiditas tetap menjadi variabel penting, sejalan dengan temuan literatur microstructure (Kyle, 1985).

Untuk menguji relevansi praktis dari model, penelitian ini melakukan simulasi strategi perdagangan berbasis *1-day holding period*. Model hybrid menghasilkan return tahunan sebesar 41,5% dengan Sharpe ratio 1,12, jauh melampaui strategi *buy-and-hold* yang hanya menghasilkan return 28,5% dengan Sharpe ratio 0,85. Profit factor sebesar 1,64 menunjukkan keberlanjutan performa model dalam jangka panjang, di mana rasio profit terhadap loss menunjukkan efisiensi yang cukup tinggi. Hasil ini memperlihatkan bahwa model hybrid tidak hanya unggul dari perspektif prediksi, tetapi juga dapat diterapkan secara praktis dalam strategi perdagangan algoritmik.

Perbandingan penelitian ini dengan literatur terkini menunjukkan bahwa performa akurasi 58,6% berada dalam kategori kompetitif. Beberapa penelitian berbasis Transformer melaporkan akurasi sekitar 60–62%, namun dengan biaya komputasi yang jauh lebih besar (Zhang et al., 2023). Model hybrid dalam penelitian ini menawarkan alternatif yang lebih efisien secara komputasi namun tetap handal dalam performa, sehingga memiliki nilai praktis yang tinggi untuk aplikasi dunia nyata. Selain itu, kontribusi penelitian ini menjadi kuat karena mencakup evaluasi pada sepuluh aset, penggunaan *TimeSeriesSplit* untuk menghindari *data leakage*, serta pengujian robust melalui dua pendekatan uji statistik yang valid dalam ranah keuangan.

Secara keseluruhan, hasil penelitian ini memperlihatkan bahwa model hybrid LSTM–XGBoost mampu memberikan prediksi harga cryptocurrency yang lebih akurat, stabil, dan signifikan secara statistik dibandingkan model individual. Selain itu, model hybrid terbukti dapat menghasilkan kinerja investasi yang unggul dalam simulasi perdagangan, sehingga memberikan kontribusi penting terhadap pengembangan sistem prediksi berbasis machine learning dalam konteks pasar keuangan digital yang dinamis. Temuan ini menegaskan bahwa

pendekatan integratif yang memanfaatkan kekuatan deep learning dan model berbasis pohon merupakan strategi yang efektif untuk meningkatkan kualitas prediksi dan performa investasi.

KESIMPULAN

Penelitian ini bertujuan mengembangkan dan mengevaluasi model prediksi harga cryptocurrency berbasis pendekatan hybrid LSTM–XGBoost dengan memanfaatkan data historis sepuluh aset kripto berkapitalisasi besar. Hasil penelitian menunjukkan bahwa model hybrid secara konsisten memberikan performa yang lebih unggul dibandingkan model baseline LSTM dan XGBoost. Model hybrid mencapai akurasi prediksi arah harga rata-rata 58,6%, meningkat signifikan atas LSTM (51,7%) dan XGBoost (53,6%), dengan nilai RMSE yang rendah dan stabil pada seluruh aset. Validasi statistik melalui *paired t-test* dan *McNemar test* memastikan bahwa peningkatan ini signifikan secara ilmiah dan tidak disebabkan oleh variasi acak. Dengan demikian, model hybrid terbukti mampu menangkap pola temporal dan hubungan non-linear yang kompleks pada pasar kripto yang sangat volatil.

Analisis lebih lanjut menunjukkan bahwa representasi laten LSTM berperan besar dalam keberhasilan prediksi melalui kontribusi fitur sebesar 45%, sementara indikator teknikal seperti RSI dan volume juga memberikan peran penting. Model hybrid tidak hanya memberikan keunggulan pada akurasi prediksi, tetapi juga memiliki implikasi praktis yang kuat melalui simulasi strategi trading *1-day holding period*. Hasil simulasi mencatat return tahunan 41,5% dengan Sharpe ratio 1,12, melampaui strategi *buy-and-hold*. Hal ini menegaskan bahwa model hybrid memiliki potensi aplikasi nyata dalam pengembangan sistem perdagangan kuantitatif yang efisien dan menguntungkan.

Secara keseluruhan, penelitian ini memberikan kontribusi penting dalam literatur prediksi harga cryptocurrency melalui evaluasi multi-aset, desain metodologi yang ketat, penggunaan *TimeSeriesSplit* untuk menghindari *data leakage*, serta pengujian statistik yang komprehensif. Temuan penelitian menegaskan bahwa pendekatan integratif antara *deep sequence modeling* dan *gradient-boosted learners* merupakan strategi yang efektif dan efisien dalam meningkatkan performa prediksi di pasar keuangan digital. Penelitian ini membuka ruang untuk pengembangan lebih lanjut, khususnya melalui eksplorasi fitur makroekonomi, model perhatian (*attention-based*), dan strategi trading multi-horizon.

DAFTAR PUSTAKA

- Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. *Expert Systems with Applications*, 36(3), 5932–5941.
- Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120, 70–83.
- Bollinger, J. (2002). *Bollinger on Bollinger Bands*. McGraw-Hill.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 785–794.

- Chollet, F. (2018). *Deep Learning with Python*. Manning.
- Corbet, S., Larkin, C., & Lucey, B. (2019). The contagion effects of the COVID-19 pandemic on the cryptocurrency market. *Finance Research Letters*, 35, 101–110.
- Corbet, S., Meegan, A., Larsen, C., Lucey, B., & Yarovaya, L. (2018). Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Economics Letters*, 165, 28–34.
- Dantas, T. M., Oliveira, F. L. C., & Reiff, M. (2022). Hybrid neural networks ensemble for financial time series forecasting. *Neurocomputing*, 498, 86–99.
- De Prado, M. L. (2018). *Advances in Financial Machine Learning*. Wiley.
- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Kim, J., & Kang, S. (2020). Financial time series forecasting using LSTM and XGBoost. *IEEE Access*, 8, 189215–189228.
- Kuncheva, L. I. (2014). *Combining Pattern Classifiers: Methods and Algorithms* (2nd ed.). John Wiley & Sons.
- Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6), 1315–1335.
- Nadarajah, S., & Chu, J. (2017). On the inefficiency of Bitcoin. *Economics Letters*, 150, 6–9.
- Permentier, E., Mahroji, M., & Setiawan, H. (2024). Optimasi prediksi harga cryptocurrency berbasis model machine learning. *Journal Teknik Informatika ITB*, 12(3), 215–234.
- Silvanus, A., Wijaya, B., & Karjono. (2022). Bitcoin price forecasting using seasonal log-differenced XGBoost and LSTM hybrid model. *Sistemasi: Jurnal Sistem Informasi*, 11(3), 512–528.
- Wang, J., Wang, J., & Zhang, Z. (2020). Stock price prediction using LSTM. *Neural Computing and Applications*, 32(17), 13029–13038.
- Wilder, J. W. (1978). *New Concepts in Technical Trading Systems*. Trend Research.
- Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259.
- Zhang, W., Li, Y., Gao, Q., & Zhou, Y. (2023). Deep learning based ensemble model for cryptocurrency price prediction. *Neural Networks*, 159, 90–101.
- Zhou, Z. H. (2012). *Ensemble Methods: Foundations and Algorithms*. CRC Press.
- Lo, A. W. (2004). The adaptive markets hypothesis. *The Journal of Portfolio Management*, 30(5), 15–29.

- Urquhart, A. (2016). The inefficiency of Bitcoin. *Economics Letters*, 148, 80–82.
- Zhang, X., Li, Y., & Wang, J. (2023). Transformer-based models for cryptocurrency forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 34(4), 1542–1556.