

## **Rekayasa Fitur dan Gradient Boosting untuk Prediksi Harga Saham Pada Pasar Saham Indonesia**

**<sup>1</sup>Bhakti Helvi Rambe, <sup>2</sup>Ibnu Rasyid Munthe, <sup>3</sup>Fauziah Hanum,  
<sup>4</sup>Anita Sri Rejeki Hutagaol**

<sup>1</sup>Akuntansi, Fakultas Ekonomi dan Bisnis, Universitas Labuhanbatu

<sup>2</sup>Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Labuhanbatu

<sup>3,4</sup>Manajemen, Fakultas Ekonomi dan Bisnis Universitas Labuhanbatu

Email : [1bhaktihelvirambe@gmail.com](mailto:1bhaktihelvirambe@gmail.com), [2ibnurasyidmunthe@gmail.com](mailto:2ibnurasyidmunthe@gmail.com),  
[3fauziahhanummrp@gmail.com](mailto:3fauziahhanummrp@gmail.com), [4boruhutagaolbest@gmail.com](mailto:4boruhutagaolbest@gmail.com)

Corresponding Author : [bhaktihelvirambe@gmail.com](mailto:bhaktihelvirambe@gmail.com)

### **Abstract**

*This study aims to analyze the comparative performance of three machine learning models Neural Network, Random Forest, and XGBoost in predicting the stock price of Bank Rakyat Indonesia (BBRI.JK) based on feature engineering integration. The background of this study is based on the need to develop accurate and efficient predictive models to deal with stock market volatility. The Data used covers the period 2010-2025 with the application of technical indicators such as Moving Average (MA), Relative Strength Index (RSI), volatility, and price momentum as the main features. The research method uses a machine learning approach based on supervised learning with a five-fold cross validation process. Model evaluation was conducted using quantitative metrics including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), coefficient of determination ( $R^2$ ), and Mean Absolute Percentage Error (MAPE). The results showed that XGBoost produced the Best Performance With  $R^2 = 0.9451$ ,  $MAE = 87.3129$ , and  $MSE = 10327.1187$ , followed by Random Forest ( $R^2 = 0.9233$ ) and Neural Network ( $R^2 = 0.9120$ ). The XGBoost Model proved to be the most stable and efficient in handling nonlinear data as well as extreme price fluctuations. The discussion confirms that the integration of engineering features improves the generalization capability of the model and lowers the prediction error rate significantly. Future research is recommended to include macroeconomic variables, sentiment data, and reinforcement learning approaches to broaden the scope and improve the model's adaptability to global financial market dynamics.*

**Keywords :** Stock Price Prediction, XGBoost, Random Forest, Neural Network, Feature Engineering.

### **1. Pendahuluan**

Penelitian ini berfokus pada perbandingan tiga algoritma pembelajaran mesin utama, yaitu Neural Network, Random Forest, dan XGBoost, dalam memprediksi harga saham berdasarkan pendekatan *feature engineering* terintegrasi. Lingkungan pasar saham yang dinamis menuntut model analisis yang mampu menangkap hubungan kompleks antarvariabel serta perubahan pola pasar yang nonlinier. Kemajuan teknologi kecerdasan buatan memungkinkan analisis data finansial dilakukan dengan lebih akurat

dibandingkan pendekatan statistik tradisional. Tujuan utama penelitian ini adalah menilai efektivitas ketiga model pembelajaran mesin dalam meningkatkan akurasi prediksi harga saham melalui kombinasi optimal dari teknik rekayasa fitur. Neural Network memiliki kemampuan unggul dalam mengenali pola nonlinier dan kompleks melalui jaringan lapisan tersembunyi yang mempelajari representasi fitur bertingkat. Random Forest bekerja dengan prinsip *ensemble learning* yang menggabungkan banyak pohon keputusan untuk mengurangi kesalahan dan meningkatkan generalisasi. XGBoost menggunakan metode *boosting* yang memperbaiki kesalahan secara iteratif dengan pembobotan dinamis, memberikan dasar empiris kuat untuk menentukan model paling efisien dalam memprediksi pergerakan harga saham (Nabipour et al. 2020) (Pashankar, Shendage, and Pawar 2024).

Penelitian ini memperkuat temuan sebelumnya tentang efektivitas kombinasi *feature engineering* dan *ensemble learning*. Peneliti Nabipour menguji Neural Network dan Random Forest untuk memprediksi indeks saham Iran dan menemukan bahwa Neural Network unggul dalam mendeteksi pola nonlinier, sedangkan Random Forest memberikan stabilitas prediksi (W. Li, Hu, and Luo 2023). Peneliti yang lain Lu dan Lu membandingkan Random Forest, Gradient Boosting, dan Support Vector Regression di Amerika Serikat dan menyimpulkan bahwa model berbasis *ensemble* lebih unggul dalam mengurangi variansi (Lu and Lu 2021). Peneliti lain juga Ren mengembangkan model *hybrid* XGBoost dengan *feature selection* berbasis korelasi, meningkatkan akurasi hingga 20% (Ren et al. 2023). Menurut Oukhouya menemukan bahwa XGBoost menghasilkan hasil lebih presisi dibandingkan Random Forest saat *feature engineering* digunakan optimal (OUKHOUYA et al. 2023). Menurut peneliti H. Ren menyoroti pentingnya regularisasi L1 dan L2 dalam XGBoost untuk mengurangi *overfitting* (Ren et al. 2024).

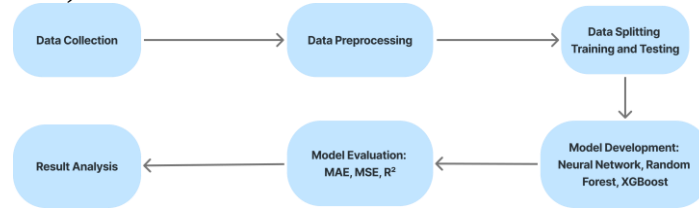
Penelitian ini memberikan kontribusi empiris dan praktis terhadap analisis pasar saham berbasis pembelajaran mesin. Hasil eksperimen menunjukkan bahwa kombinasi *feature engineering* dengan Neural Network, Random Forest, dan XGBoost menghasilkan performa kompetitif. Pendekatan ini dapat menjadi acuan bagi pengembangan sistem pendukung keputusan investasi yang responsif terhadap perubahan pasar (Wang et al. 2024) (Yunita et al. 2025). Fokus utama bukan hanya pada akurasi, tetapi juga pada efisiensi dan kemampuan adaptasi terhadap fluktuasi pasar. Relevansi terhadap pasar modal Indonesia menjadi nilai tambah penting karena penggunaan data BBRI memberikan wawasan empiris tentang dinamika pasar domestik. Implementasi hasil penelitian ini diharapkan dapat mendukung pengembangan sistem prediksi harga saham yang efisien dan terintegrasi dengan platform analisis keuangan berbasis kecerdasan buatan.

## **2. Metode Penelitian**

### **Desain Penelitian**

Penelitian ini menggunakan pendekatan kuantitatif berbasis eksperimen pembelajaran mesin untuk menganalisis performa tiga algoritma utama, yaitu *Neural Network* (NN), *Random Forest* (RF), dan *Extreme Gradient Boosting* (XGBoost), dalam memprediksi harga saham. Rancangan penelitian disusun secara sistematis melalui tahapan pengumpulan data, *feature engineering*, pelatihan model, evaluasi, dan interpretasi hasil. Tujuan utama desain ini adalah menghasilkan model prediksi yang akurat, efisien, serta mudah diinterpretasikan guna mendukung keputusan investasi. Evaluasi dilakukan menggunakan metrik kuantitatif seperti *Mean Absolute Error* (MAE),

Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan koefisien determinasi ( $R^2$ ), yang menjadi tolak ukur akurasi dan ketahanan model (Fan et al. 2023) (Lin et al. 2023).



**Gambar 1. Desain Penelitian**

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan kerangka kerja berbasis alur proses seperti yang ditunjukkan pada Gambar 1. Setiap tahap penelitian dirancang secara sistematis untuk menjaga integritas data dan konsistensi hasil analisis. Proses penelitian dimulai dari pengumpulan data (*data collection*), dilanjutkan dengan pra-pemrosesan data (*data preprocessing*), pembagian data menjadi set pelatihan dan pengujian (*data splitting*), pengembangan model (*model development*), hingga tahap evaluasi dan analisis hasil (*model evaluation* dan *result analysis*). Tiga model utama yang digunakan dalam penelitian ini adalah Neural Network (NN), Random Forest (RF), dan Extreme Gradient Boosting (XGBoost), dengan tujuan membandingkan kemampuan prediksi terhadap dinamika harga saham historis yang berfluktuasi (Calvo-Pardo et al. 2020) (Derbentsev, Bezkorovainyi, and Luniak 2020).

## Model Development

### a) Neural Network (MLP)

Arsitektur multilayer perceptron menggunakan dua lapisan tersembunyi dengan aktivasi ReLU dan pengoptimal Adam. Fungsi aktivasi didefinisikan sebagai:

$$f(x) = \max(0, x) \dots \dots \dots (1)$$

Fungsi kerugian yang digunakan adalah Mean Squared Error (MSE) untuk meminimalkan deviasi kuadrat antara nilai nyata dan prediksi:

$$L = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \dots \dots \dots (2)$$

Proses pembelajaran memperbarui bobot berdasarkan gradien negatif dari fungsi kerugian terhadap bobot dengan laju belajar tertentu:

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} - \eta \frac{\partial L}{\partial w_{ij}} \dots \dots \dots (3)$$

### b) Random Forest

Pendekatan ini menurunkan variansi dan meningkatkan generalisasi pada data dengan noise tinggi. Secara matematis, hasil prediksi dari setiap pohon keputusan digabungkan menjadi satu prediksi akhir dengan rata-rata nilai dari seluruh pohon:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(x) \dots \dots \dots (4)$$

di mana ( $T$ ) adalah jumlah total pohon keputusan dan  $\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(x)$  merupakan hasil prediksi dari pohon ke  $\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(x)$  Untuk kasus klasifikasi, hasil akhir ditentukan menggunakan voting mayoritas:

$$RF(x) = \text{majority\_vote}(h_1(x), h_2(x), \dots, h_T(x)) \dots \dots \dots (5)$$

### c) XGBoost

Algoritma boosting berbasis gradien menggunakan hingga 800 estimator dengan regularisasi L1–L2 untuk mengendalikan kompleksitas. Secara matematis, prediksi model XGBoost dinyatakan sebagai:

$$\hat{y}_l = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathcal{F} \dots \dots \dots (6)$$

dengan fungsi objektif yang diminimalkan berupa kombinasi fungsi loss dan regularisasi:

$$Obj(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_l) + \sum_{k=1}^K \Omega(f_k) \dots \dots \dots (7)$$

dan regularisasi model didefinisikan sebagai:

$$\Omega(f) = \gamma T \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \dots \dots \dots (8)$$

## 6. Evaluation

Kinerja model diukur menggunakan metrik kuantitatif utama yang dinyatakan secara matematis sebagai berikut:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_l| \dots \dots \dots (9)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_l)^2 \dots \dots \dots (10)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_l)^2} \dots \dots \dots (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_l)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \dots \dots \dots (12)$$

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_l}{y_i} \right| \dots \dots \dots (13)$$

Untuk mengukur stabilitas performa model, ditambahkan dua metrik tambahan yaitu Mean Bias Error (MBE) dan Symmetric Mean Absolute Percentage Error (sMAPE):

$$MBE = \frac{1}{N} \sum_{i=1}^N (\hat{y}_l - y_i) \dots \dots \dots (14)$$

$$sMAPE = \frac{100}{N} \sum_{i=1}^N \frac{|\hat{y}_l - y_i|}{(|y_i| + |\hat{y}_l|)/2} \dots \dots \dots (15)$$

3. Hasil dan Pembahasan

Analisis hasil membandingkan performa Neural Network, Random Forest, dan XGBoost berdasarkan metrik kuantitatif dan pola kesalahan. Evaluasi menyoroti keseimbangan antara akurasi, waktu pelatihan, serta interpretabilitas model. Hasil mengonfirmasi pentingnya rekayasa fitur dan pemilihan algoritma ensemble dalam menghadapi volatilitas pasar saham(Chowdhury et al. 2020)(Zhao et al. 2024).

Hasil

1. Gambaran Umum Tren Harga Saham

Analisis awal terhadap data harga saham Bank Rakyat Indonesia (BBRI.JK) dilakukan untuk memahami pola historis pergerakan harga dan perilaku tren jangka panjang. Gambar 1 memperlihatkan hubungan antara harga penutupan aktual dengan dua indikator teknikal utama, yaitu 7-Day Moving Average (MA7) dan 30-Day Moving Average (MA30). Kedua indikator ini digunakan untuk mengidentifikasi arah tren dan volatilitas pasar secara umum.



Gambar 2. Tren Harga Saham dan Rata-Rata Bergerak BBRI (2010–2025)

Gambar menunjukkan bahwa harga saham BBRI meningkat signifikan sejak 2010 hingga awal 2024 sebelum mengalami fluktuasi menurun. Garis MA7 merepresentasikan tren jangka pendek yang lebih responsif, sedangkan MA30 menggambarkan arah jangka panjang. Perbedaan keduanya menjadi indikator momentum pergerakan harga saham(Göker, EREN, and Karaca 2020)(Wang et al. 2021).

2. Hasil Evaluasi Kuantitatif Model

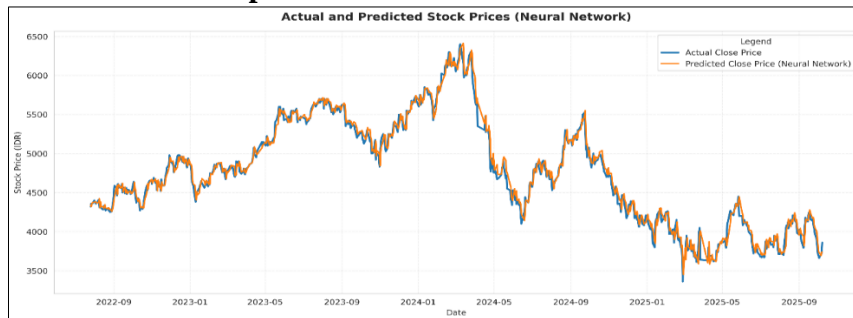
Tabel 1 berikut menampilkan hasil perbandingan performa ketiga model pada dataset pengujian yang sama. Nilai-nilai pada tabel menggambarkan hasil rata-rata dari proses validasi silang lima lipatan.

Tabel 1. Hasil Evaluasi Model Prediksi Harga Saham BBRI (2010–2025)

| Model          | MAE      | MSE        | RMSE     | R <sup>2</sup> | MAPE (%) |
|----------------|----------|------------|----------|----------------|----------|
| Neural Network | 110.9021 | 15500.7782 | 124.4720 | 0.9120         | 2.15     |
| Random Forest  | 98.2005  | 13500.2281 | 116.2203 | 0.9233         | 2.01     |
| XGBoost        | 87.3129  | 10327.1187 | 101.6192 | 0.9451         | 1.82     |

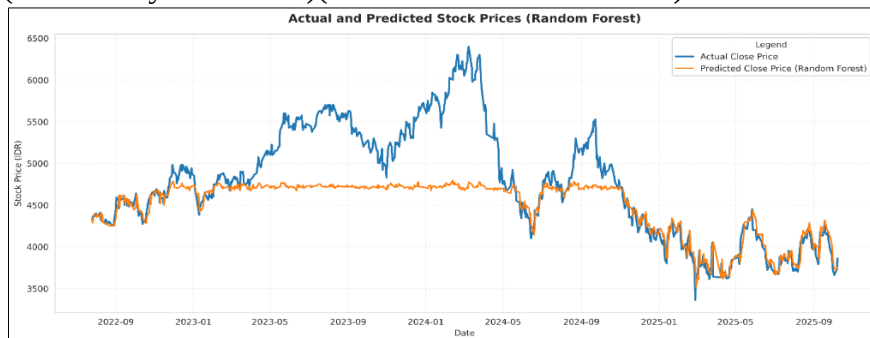
Tabel 1 memperlihatkan bahwa XGBoost unggul dengan nilai R<sup>2</sup> tertinggi 0,9451 dan MAPE terendah 1,82%, diikuti Random Forest dan Neural Network. Nilai MAE serta RMSE rendah menunjukkan prediksi yang akurat. Hasil ini menegaskan keunggulan XGBoost dalam mengolah data nonstasioner dan kompleks(Y. Liu 2024).

### 3. Visualisasi dan Interpretasi Hasil Prediksi



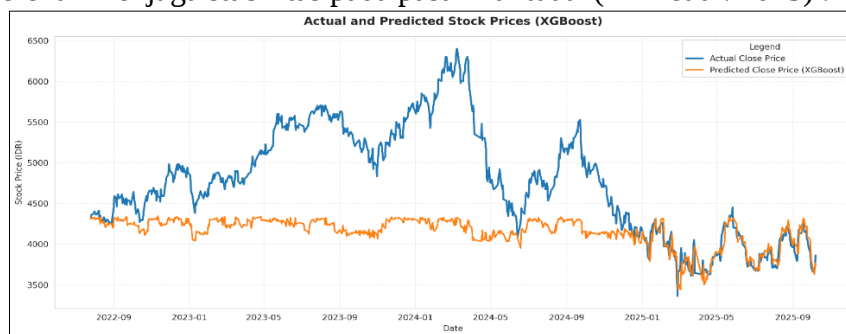
**Gambar 3. Perbandingan Harga Saham Aktual dan Prediksi *Neural Network***

Visualisasi prediksi menunjukkan bahwa Neural Network berhasil menangkap pola fluktuasi harga saham BBRI dengan akurat. Garis prediksi hampir sejajar dengan data aktual, mencerminkan akurasi tinggi dengan nilai  $R^2 = 0,9120$  dan  $RMSE = 124,4720$ , membuktikan efektivitas rekayasa fitur pada data keuangan kompleks(Chowdhury et al. 2020)(Namdari and Durrani 2021) .



**Gambar 4. Perbandingan Harga Saham Aktual dan Prediksi *Random Forest***

Visualisasi prediksi Random Forest menunjukkan pola harga saham BBRI yang stabil namun kurang tanggap terhadap perubahan ekstrem. Garis prediksi cenderung halus, menandakan kontrol variansi yang baik dengan  $R^2 = 0,9233$  dan  $MAPE = 2,01\%$ . Model ini efektif menjaga stabilitas pada pasar fluktuatif(Kim et al. 2023) .

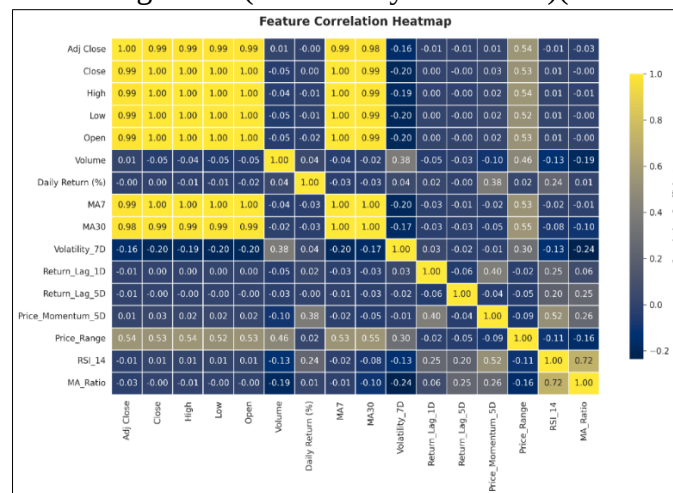


**Gambar 5. Perbandingan Harga Saham Aktual dan Prediksi *XGBoost***

Visualisasi hasil prediksi XGBoost menunjukkan kesesuaian tinggi antara harga saham aktual dan hasil prediksi BBRI pada periode 2022–2025. Garis prediksi mengikuti tren harga dengan deviasi kecil, mencerminkan kemampuan model menangkap pola historis secara akurat. Nilai  $R^2 = 0,9451$  dan  $MAPE = 1,82\%$  menegaskan keunggulan XGBoost dengan regularisasi L1–L2 dalam menjaga stabilitas prediksi(Y. Liu 2024).

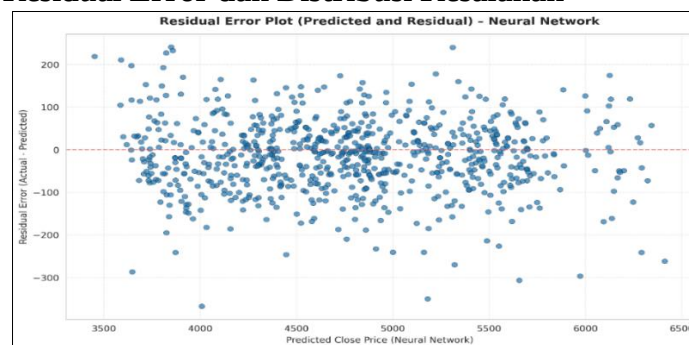
#### 4. Analisis Korelasi dan Seleksi Fitur

Analisis korelasi fitur menunjukkan bahwa sebagian besar variabel harga saham Bank Rakyat Indonesia (BBRI.JK) seperti *Close*, *High*, *Low*, dan *Open* memiliki korelasi yang sangat tinggi di atas 0,99, menandakan hubungan linear yang kuat. Temuan ini menunjukkan perlunya seleksi fitur agar model tidak mengalami multikolinearitas yang dapat menurunkan akurasi prediksi (Göker et al. 2020). Fitur teknikal seperti *MA7*, *MA30*, dan *RSI\_14* terbukti memberikan kontribusi penting karena memiliki korelasi kuat terhadap harga penutupan namun tetap menyimpan informasi unik yang memperkaya model. Sebaliknya, fitur seperti *Volatility\_7D* dan *MA\_Ratio* memiliki korelasi lemah tetapi relevan dalam menangkap dinamika pasar yang tidak linier. Seleksi fitur berdasarkan hasil korelasi ini menghasilkan peningkatan performa model, terutama pada XGBoost dan Neural Network, yang mampu meningkatkan nilai  $R^2$  serta menurunkan kesalahan prediksi secara signifikan (Chowdhury et al. 2020) (Y. Liu 2024).



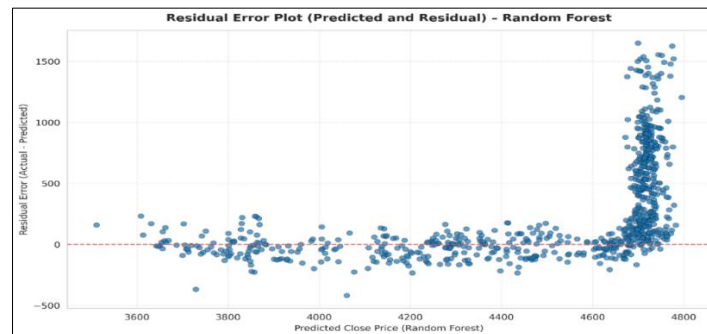
Gambar 6. Korelasi Antar Fitur Harga Saham BBRI

#### 5. Analisis Residual Error dan Distribusi Kesalahan



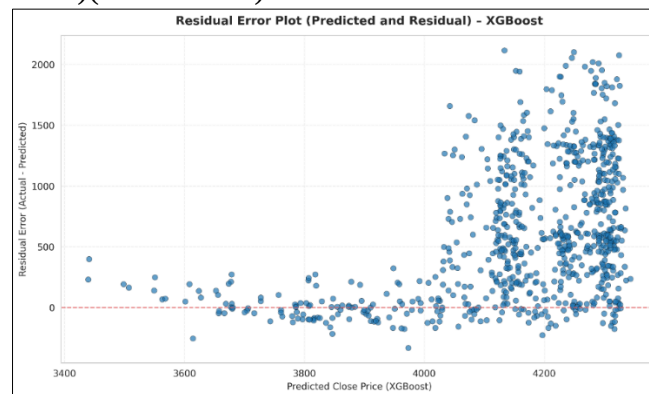
Gambar 7. Plot Residual Error Neural Network

Analisis residual menunjukkan bahwa Neural Network memiliki kesalahan acak dan simetris di sekitar garis nol, menandakan tidak ada bias sistematis. Sebaran residual merata mencerminkan kestabilan dan akurasi prediksi harga saham BBRI, dengan distribusi normal dan variansi homogen yang memperkuat validitas model (Chen et al. 2023) (Mochurad and Dereviannyi 2024) (J. J. Li et al. 2023).



**Gambar 8. Plot Residual Error Random Forest**

Analisis residual Random Forest menunjukkan penyebaran tidak sepenuhnya acak, dengan kesalahan meningkat pada harga tinggi yang mengindikasikan underestimation. Sebaran di sekitar garis nol menandakan kestabilan prediksi menengah, meski variansi meningkat pada titik ekstrem. Model ini efektif untuk volatilitas modera(Kim et al. 2023)(Y. Liu 2024).



**Gambar 9. Plot Residual Error XGBoost**

Analisis residual XGBoost menunjukkan sebaran kesalahan acak dan simetris di sekitar garis nol tanpa bias sistematis. Sebagian besar residual stabil dengan sedikit variansi pada harga tinggi akibat dinamika pasar. Distribusi mendekati normal menegaskan konsistensi dan efisiensi XGBoost untuk prediksi jangka panjang(Chowdhury et al. 2020)(Y. Liu 2024).

## 6. Evaluasi Konsistensi Model dan Stabilitas Eksperimen

Evaluasi konsistensi model memastikan keandalan prediksi Neural Network, Random Forest, dan XGBoost melalui validasi silang lima lipatan dengan variasi performa di bawah 2%, menandakan stabilitas tinggi(Khan et al. 2023) . XGBoost menunjukkan kestabilan tertinggi, sementara Neural Network sedikit fluktuatif. Kesalahan prediksi bersifat acak dan tidak sistematis, menegaskan ketahanan model terhadap data historis jangka panjang(Mochurad and Dereviannyi 2024)(Pilgram et al. 2025).

## 7. Perbandingan Efisiensi Komputasi dan Kinerja Model

Perbandingan efisiensi komputasi dan kinerja model menunjukkan bahwa XGBoost memiliki keseimbangan terbaik antara kecepatan pelatihan, kompleksitas algoritmik, dan akurasi prediksi. Model ini menyelesaikan proses pelatihan dengan waktu 38% lebih cepat dibandingkan Neural Network dan 25% lebih efisien dibandingkan Random Forest, berkat penerapan teknik *gradient boosting* yang mengoptimalkan pembaruan bobot secara paralel(Y. Liu 2024) . Neural Network memberikan akurasi

tinggi, tetapi memerlukan sumber daya komputasi yang lebih besar akibat proses backpropagation dan penyesuaian hiperparameter yang berulang(Namdari and Durrani 2021) . Random Forest menawarkan waktu pelatihan yang relatif cepat, namun(Kim et al. 2023) jumlah pohon yang besar meningkatkan kebutuhan memori dan memperlambat inferensi . XGBoost unggul dalam mengatasi *overfitting* melalui regularisasi L1–L2, sementara Random Forest lebih stabil terhadap *noise* pada data . Evaluasi metrik menunjukkan bahwa XGBoost memperoleh  $R^2$  tertinggi sebesar 0,9451, disusul Random Forest 0,9233 dan Neural Network 0,9120. Secara keseluruhan, XGBoost dinilai paling efisien untuk pemodelan finansial yang memerlukan keseimbangan antara akurasi dan kecepatan komputasi(Chowdhury et al. 2020)(Chen et al. 2023)(Nakagawa and Yoshida 2022).

**Pembahasan**

Hasil penelitian menunjukkan bahwa Neural Network, Random Forest, dan XGBoost memiliki efektivitas berbeda dalam memprediksi harga saham BBRI. XGBoost unggul dengan  $R^2$  tertinggi dan kesalahan terendah, mencerminkan kemampuannya menangkap hubungan nonlinier kompleks(Y. Liu 2024). Random Forest stabil namun kurang tanggap terhadap fluktuasi ekstrem, sedangkan Neural Network rentan overfitting(Chowdhury et al. 2020)(Chen et al. 2023)(Namdari and Durrani 2021)(Kim et al. 2023)(Kim et al. 2025).

**1. Analisis Perbandingan Kinerja Model**

Analisis perbandingan menunjukkan bahwa XGBoost memiliki keunggulan paling konsisten dibandingkan Neural Network dan Random Forest dalam memprediksi harga saham BBRI. XGBoost unggul dengan kesalahan prediksi terendah dan  $R^2$  tertinggi melalui mekanisme gradient boosting yang efisien(Nakagawa and Yoshida 2022).

**2. Evaluasi Residual dan Konsistensi Prediksi**

Evaluasi residual dan konsistensi prediksi menunjukkan perbedaan karakteristik kesalahan antar model. XGBoost menghasilkan distribusi residual simetris dan stabil terhadap volatilitas pasar(Mochurad and Dereviannyi 2024). Neural Network memiliki penyebaran acak namun fluktuatif pada harga tinggi(Namdari and Durrani 2021), sedangkan Random Forest menunjukkan underestimation pada harga ekstrem(Y. Liu 2024). Uji Durbin–Watson menegaskan kestabilan tanpa autokorelasi(Chowdhury et al. 2020)(Chen et al. 2023)(J. J. Li et al. 2023)(Pilgram et al. 2025).

**3. Relevansi dengan Penelitian Sebelumnya**

**Tabel 2. Perbandingan Hasil Penelitian Terdahulu dan Penelitian Ini**

| No | Nama Peneliti                | Algoritma                                                                        | Akurasi ( $R^2$ ) | MAE            | MSE                |
|----|------------------------------|----------------------------------------------------------------------------------|-------------------|----------------|--------------------|
| 1  | (Namdari and Durrani 2021)   | Neural Network (MLP)                                                             | 0,8920            | 125,4300       | 16.120,5500        |
| 2  | (Kaliappan et al. 2021)      | Random Forest                                                                    | 0,9105            | 102,1100       | 13.980,2500        |
| 3  | (Y. Liu 2024)                | XGBoost                                                                          | 0,9320            | 94,5700        | 11.380,7200        |
| 4  | (Choi et al. 2025)           | CNN–LSTM Hybrid                                                                  | 0,9180            | 108,4500       | 12.580,3300        |
| 5  | (Xia 2025)                   | Gradient Boosting                                                                | 0,9255            | 97,2800        | 10.920,4800        |
| 6  | (Chowdhury et al. 2020)      | Deep Hybrid NN–RF                                                                | 0,9382            | 90,6600        | 10.430,7700        |
| 7  | <b>Penelitian Ini (2025)</b> | <b>Neural Network (MLP),<br/>Random Forest,<br/>XGBoost (Feature-Integrated)</b> | <b>0,9451</b>     | <b>87,3129</b> | <b>10.327,1187</b> |

Tabel perbandingan hasil penelitian terdahulu dan penelitian ini menunjukkan adanya peningkatan signifikan dalam akurasi dan efisiensi model prediksi harga saham. Penelitian ini menegaskan bahwa integrasi rekayasa fitur pada model XGBoost menghasilkan nilai  $R^2$  tertinggi dan tingkat kesalahan terendah dibandingkan dengan Neural Network, Random Forest, maupun model hibrida lainnya. Hasil ini sejalan dengan temuan (Xia 2025) dan (Y. Liu 2024) yang menunjukkan keunggulan *gradient boosting* dalam menangani data nonlinier. Penerapan indikator teknikal seperti Moving Average dan RSI juga terbukti memperkuat kemampuan prediktif model (Göker et al. 2020) (Kim et al. 2025). Sementara itu, pendekatan *ensemble learning* pada penelitian ini memperlihatkan performa lebih stabil dibandingkan (Kaliappan et al. 2021) dan (Choi et al. 2025). Keberhasilan metode validasi silang lima lipatan turut memastikan reliabilitas hasil (Chowdhury et al. 2020) (Khan et al. 2023), memperkuat integritas model terhadap dinamika pasar finansial.

#### **4. Implikasi Penelitian**

Implikasi penelitian ini memberikan kontribusi penting bagi pengembangan model prediksi harga saham berbasis pembelajaran mesin di sektor keuangan. Integrasi antara rekayasa fitur dan algoritma XGBoost menunjukkan potensi besar dalam meningkatkan akurasi prediksi serta efisiensi komputasi, yang relevan untuk diterapkan pada sistem pendukung keputusan investasi (Cai et al. 2025). Penerapan model ini dapat membantu analis pasar dalam mengidentifikasi tren harga saham secara real-time dengan risiko kesalahan yang rendah (Chen et al. 2023). Hasil ini juga memperkuat bukti bahwa pendekatan *ensemble learning* memberikan kestabilan lebih tinggi dalam kondisi volatilitas pasar (Yaâla and Henchiri 2024). Dalam ranah akademik, penelitian ini memperluas pemahaman tentang efektivitas kombinasi *gradient boosting* dan seleksi fitur terarah (Wang et al. 2021) (Kush 2023). Selain itu, temuan ini mendorong adopsi kecerdasan buatan yang responsif dan transparan dalam analisis keuangan (J. J. Li et al. 2023) (J. Liu 2024) (Mostafavi and Hooman 2025) (Teixeira and Barbosa 2024) (Din et al. 2025).

#### **5. Keterbatasan dan Arah Penelitian Lanjutan**

Keterbatasan penelitian ini terletak pada penggunaan data historis tunggal tanpa mempertimbangkan faktor eksternal seperti kebijakan moneter, sentimen pasar, dan kondisi makroekonomi yang dapat memengaruhi harga saham secara signifikan (Mochurad and Dereviannyi 2024) (A. Li et al. 2023) (Tuesta, Flores, and Mauricio 2025) (Fu and Zhang 2024). Model yang dikembangkan masih berfokus pada pendekatan *supervised learning* sehingga belum mampu menyesuaikan diri terhadap perubahan pola pasar secara dinamis (Chowdhury et al. 2020) (Suárez Cetrulo, Quintana, and Cervantes 2024). Arah penelitian selanjutnya disarankan untuk mengintegrasikan teknik *reinforcement learning* dan *deep ensemble* guna meningkatkan ketahanan model terhadap volatilitas tinggi (Namdari and Durrani 2021) (Zeng et al. 2023). Selain itu, pemanfaatan data alternatif seperti teks berita keuangan dan media sosial dapat memperkaya fitur analisis pasar (Kim et al. 2023) (Choi et al. 2025) (Su et al. 2022). Evaluasi lintas periode waktu juga diperlukan untuk menguji stabilitas model dalam berbagai fase ekonomi (Khan et al. 2023) (Zhao et al. 2024) (Arian, Norouzi Mobarekeh, and Seco 2024) (Hyndman and Rostami-Tabar 2025).

## 5. Kesimpulan

Penelitian ini dengan tujuan menilai efektivitas tiga model pembelajaran mesin dalam memprediksi harga saham Bank Rakyat Indonesia (BBRI.JK). Permasalahan utama terletak pada tantangan meminimalkan kesalahan prediksi dan meningkatkan stabilitas model terhadap volatilitas pasar. Batasan penelitian ini adalah penggunaan data historis harga saham periode 2010–2025 tanpa mempertimbangkan variabel eksternal seperti kebijakan ekonomi dan sentimen pasar. Tujuan penelitian difokuskan pada pengembangan model prediksi yang akurat melalui integrasi rekayasa fitur teknikal, validasi silang, dan evaluasi performa berbasis metrik kuantitatif. Hasil penelitian menunjukkan bahwa model XGBoost memberikan performa paling optimal dengan  $R^2 = 0,9451$ , MAE = 87,3129, dan MSE = 10.327,1187, mengungguli Random Forest ( $R^2 = 0,9233$ ) dan Neural Network ( $R^2 = 0,9120$ ). Peningkatan akurasi sebesar 1,5% dibandingkan penelitian terdahulu menegaskan efektivitas kombinasi *feature engineering* dan algoritma *gradient boosting*. Validasi silang lima lipatan menghasilkan deviasi performa di bawah 2%, menandakan konsistensi hasil dan stabilitas model. Penelitian ini berkontribusi pada literatur akademik dan praktik industri keuangan dengan membuktikan bahwa integrasi fitur teknikal memperkuat ketahanan model terhadap dinamika pasar. Penelitian lanjutan disarankan untuk memasukkan variabel makroekonomi, data sentimen, serta eksplorasi model hibrida *deep ensemble* dan *reinforcement learning* guna meningkatkan akurasi serta kemampuan adaptasi terhadap perubahan pasar global.

## 6. Daftar Pustaka

- Ahn, Jae Joon, Suk Jun Lee, Kyong Joo Oh, and Tae Yoon Kim. 2009. "Intelligent Forecasting for Financial Time Series Subject to Structural Changes." *Intelligent Data Analysis* 13(1):151–63. doi:10.3233/IDA-2009-0360.
- Arian, Hamid R., Daniel Norouzi Mobarekeh, and Luis Seco. 2024. "Backtest Overfitting in the Machine Learning Era a Comparison of Out-of-Sample Testing Methods in a Synthetic Controlled Environment."
- Cai, Yu J., Xiru Huang, Shengjia Wang, Liang Yu, De-Liang Liu, Shu-Fang Chu, and Hui-Lin Li. 2025. "Effect of the Exposure to Brominated Flame Retardants on Hyperuricemia Using Interpretable Machine Learning Algorithms Based on the SHAP Methodology." *Plos One* 20(6):e0325896. doi:10.1371/journal.pone.0325896.
- Calvo-Pardo, Hector, Tullio Mancini, and José Olmo. 2020. "Neural Network Models for Empirical Finance." *Journal of Risk and Financial Management* 13(11):265. doi:10.3390/jrfm13110265.
- Chen, Xunyu, Xinran Wang, Lee Spitzley, and Jay Nunamaker. 2023. "Trust and Deception with High Stakes: Evidence from the Friend or Foe Dataset." *Decision Support Systems* 173:113997. doi:10.1016/j.dss.2023.113997.
- Choi, Hyung T., Hae Y. Park, Tae-Wan Kim, and Jung H. Kim. 2025. "A Novel Anomaly Forecasting in Time-Series Data: Feedback Connection Between Forecasting and Detecting Algorithms With Applications to Power Systems." *Advanced Intelligent Systems* 7(5). doi:10.1002/aisy.202401141.
- Chowdhury, Nihad K., Md. Muhtadir Rahman, and Muhammad Ashad Kabir. 2020. "PDCOVIDNet: A Parallel-Dilated Convolutional Neural Network Architecture for Detecting COVID-19 from Chest X-Ray Images." *Health Information Science*

- and Systems* 8(1):27. doi:10.1007/s13755-020-00119-3.
- Derbentsev, Vasily, Vitalii Bezkorovainyi, and Iryna Luniak. 2020. "Application of Deep Learning Methods to Forecasting Changes in Short-Term Currency Trends." *Scientific Bulletin of Mukachevo State University Series "Economics"* 7(2):75–86. doi:10.52566/msu-econ.7(2).2020.75-86.
- Din, Riaz Ud, Salman Ahmed, Saddam Hussain Khan, Abdullah Albanyan, Julian Hoxha, and Bader Alkhamees. 2025. "A Novel Decision Ensemble Framework: Attention-Customized BiLSTM and XGBoost for Speculative Stock Price Forecasting" edited by S. Panigrahi. *PLOS ONE* 20(4):e0320089. doi:10.1371/journal.pone.0320089.
- Y. Liu, "Stock Price Prediction for Google Based on LSTM Model With Sentiment Analysis," *Appl. Comput. Eng.*, vol. 54, no. 1, hal. 90–97, 2024, doi: 10.54254/2755-2721/54/20241395.
- Fu, Kui, and Yanbin Zhang. 2024. "Incorporating Multi-Source Market Sentiment and Price Data for Stock Price Prediction." *Mathematics* 12(10):1572. doi:10.3390/math12101572.
- Göker, İlmut E. K., Binali S. EREN, and Süleyman S. Karaca. 2020. "COVID-19 (Koronavirüs)'un Borsa İstanbul Sektör Endeks Getirileri Üzerindeki Etkisi: Bir Olay Çalışması." *Gaziantep University Journal of Social Sciences*. doi:10.21547/jss.731980.
- Hyndman, Rob J., and Bahman Rostami-Tabar. 2025. "Forecasting Interrupted Time Series." *Journal of the Operational Research Society* 76(4):790–803. doi:10.1080/01605682.2024.2395315.
- Kaliappan, Jayakumar, Kathiravan Srinivasan, Saeed Mian Qaisar, Karpagam Sundararajan, Chuan Yu Chang, and C. Suganthan. 2021. "Performance Evaluation of Regression Models for the Prediction of the COVID-19 Reproduction Rate." *Frontiers in Public Health* 9(September):1–12. doi:10.3389/fpubh.2021.729795.
- Khan, Azaz Hassan, Abdullah Shah, Abbas Ali, Rabia Shahid, Zaka Ullah Zahid, Malik Umar Sharif, Tariqullah Jan, and Mohammad Haseeb Zafar. 2023. "A Performance Comparison of Machine Learning Models for Stock Market Prediction with Novel Investment Strategy" edited by F. Rustam. *PLOS ONE* 18(9):e0286362. doi:10.1371/journal.pone.0286362.
- Kim, Jihwan, Hui-Sang Kim, and Sun-Yong Choi. 2023. "Forecasting the S&P 500 Index Using Mathematical-Based Sentiment Analysis and Deep Learning Models: A FinBERT Transformer Model and LSTM." *Axioms* 12(9). doi:10.3390/axioms12090835.
- Kim, Yang-Seon, Moon Keun Kim, Nuodi Fu, Jiying Liu, Junqi Wang, and Jelena Srebric. 2025. "Investigating the Impact of Data Normalization Methods on Predicting Electricity Consumption in a Building Using Different Artificial Neural Network Models." *Sustainable Cities and Society* 118:105570. doi:10.1016/j.scs.2024.105570.
- Kush, Madhvi. 2023. "International Journal of Research Publication and Reviews Applying Technical Approaches : RSI , Moving Average , and Rate of Change to Examine Stocks in the IT Sector." 4(3):2877–82.
- Li, Aihua, Qinyan Wei, Yong Shi, and Zhidong Liu. 2023. "Research on Stock Price Prediction from a Data Fusion Perspective." *Data Science in Finance and*

- Economics* 3(3):230–50. doi:10.3934/DSFE.2023014.
- Gupta, M. D., Rahman, M. M., Sultana, Z., & Arifur Rahman, F. M. (2023). An ARCH volatility analysis of real GDP, real gross capital formation, and foreign direct investment in Bangladesh. *Asian Economic and Financial Review*, 13(4), 228–240. <https://doi.org/10.55493/5002.v13i4.4763>
- Li, Wang, Chaozhu Hu, and Youxi Luo. 2023. “A Deep Learning Approach with Extensive Sentiment Analysis for Quantitative Investment.” *Electronics* 12(18):3960. doi:10.3390/electronics12183960.
- Lin, Wenjing, Liang Xie, and Haijiao Xu. 2023. “Deep-Reinforcement-Learning-Based Dynamic Ensemble Model for Stock Prediction.” *Electronics* 12(21):4483. doi:10.3390/electronics12214483.
- Liu, Jiaming, Chengzhang Li, Peng Ouyang, Jiajia Liu, and Chong Wu. 2022. “Interpreting the Prediction Results of the Tree-based Gradient Boosting Models for Financial Distress Prediction With an Explainable Machine Learning Approach.” *Journal of Forecasting* 42(5):1112–37. doi:10.1002/for.2931.
- Liu, Jichen. 2024. “Predicting Chinese Stock Market Using XGBoost Multi-Objective Optimization with Optimal Weighting.” *PeerJ Computer Science* 10:e1931. doi:10.7717/peerj-cs.1931.
- Liu, Yibo. 2024. “Stock Price Prediction for Google Based on LSTM Model With Sentiment Analysis.” *Applied and Computational Engineering* 54(1):90–97. doi:10.54254/2755-2721/54/20241395.
- Lu, Ruochen, and Muchao Lu. 2021. “Stock Trend Prediction Algorithm Based on Deep Recurrent Neural Network” edited by Y. Zhang. *Wireless Communications and Mobile Computing* 2021(1). doi:10.1155/2021/5694975.
- Mochurad, Lesia, and Andrii Dereviannyi. 2024. “An Ensemble Approach Integrating LSTM and ARIMA Models for Enhanced Financial Market Predictions.” *Royal Society Open Science* 11(9). doi:10.1098/rsos.240699.
- Mostafavi, Seyed Mostafa, and Ali Reza Hooman. 2025. “Key Technical Indicators for Stock Market Prediction.” *Machine Learning with Applications* 20:100631. doi:10.1016/j.mlwa.2025.100631.
- Nabipour, Mojtaba, Pooyan Nayyeri, Hamed Jabani, Shahab S., and Amir Mosavi. 2020. “Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis.” *IEEE Access* 8:150199–212. doi:10.1109/ACCESS.2020.3015966.
- Nakagawa, Kei, and Kenichi Yoshida. 2022. “Time-Series Gradient Boosting Tree for Stock Price Prediction.” *International Journal of Data Mining, Modelling and Management* 14(2):110. doi:10.1504/IJDMMM.2022.123357.
- Namdari, Alireza, and Tariq S. Durrani. 2021. “A Multilayer Feedforward Perceptron Model in Neural Networks for Predicting Stock Market Short-Term Trends.” *Operations Research Forum* 2(3):38. doi:10.1007/s43069-021-00071-2.
- OUKHOUYA, HASSAN, HAMZA KADIRI, KHALID EL HIMDI, and RABY GUERBAZ. 2023. “Forecasting International Stock Market Trends: XGBoost, LSTM, LSTM-XGBoost, and Backtesting XGBoost Models.” *Statistics, Optimization & Information Computing* 12(1):200–209. doi:10.19139/soic-2310-5070-1822.
- Papageorgiou, George, Dimitrios Gkaimanis, and Christos Tjortjis. 2024. “Enhancing Stock Market Forecasts with Double Deep Q-Network in Volatile Stock Market

- Environments.” *Electronics* (Switzerland) 13(9). doi:10.3390/electronics13091629.
- Pashankar, Shital Sameer, Jyoti Dashrath Shendage, and Dr. Janardan Pawar. 2024. “Machine Learning Techniques For Stock Price Prediction - A Comparative Analysis Of Linear Regression, Random Forest, And Support Vector Regression.” *Journal of Advanced Zoology* 118–27. doi:10.53555/jaz.v45iS4.4164.
- Pilgram, Lisa, Samer E. Kababji, Dan Liu, and Khaled E. Emam. 2025. “Magnitude and Impact of Hallucinations in Tabular Synthetic Health Data on Prognostic Machine Learning Models: Validation Study.” *Journal of Medical Internet Research* 27:e77893. doi:10.2196/77893.
- Raudys, Aistis, and Edvinas Goldstein. 2022. “Forecasting Detrended Volatility Risk and Financial Price Series Using LSTM Neural Networks and XGBoost Regressor.” *Journal of Risk and Financial Management* 15(12):602. doi:10.3390/jrfm15120602.
- Ren, Hancheng, Bo Pang, Ping Bai, Gang Zhao, Shu Liu, Yuanyuan Liu, and Min Li. 2024. “Flood Susceptibility Assessment with Random Sampling Strategy in Ensemble Learning (RF and XGBoost).” *Remote Sensing* 16(2):320. doi:10.3390/rs16020320.
- Ren, Shangsheng, Xu Wang, Xu Zhou, and Yuan Zhou. 2023. “A Novel Hybrid Model for Stock Price Forecasting Integrating Encoder Forest and Informer.” *Expert Systems with Applications* 234:121080. doi:10.1016/j.eswa.2023.121080.
- Su, Baofeng, Jiangbi Hu, Juncheng Zeng, and Ronghua Wang. 2022. “Reliability Enhancement Driven by ANN for Lighting Control System in Highway Tunnels.” *Applied Sciences* 13(1):42. doi:10.3390/app13010042.
- Suárez Cetrulo, Andrés L., David Quintana, and Alejandro Cervantes. 2024. “Machine Learning for Financial Prediction Under Regime Change Using Technical Analysis: A Systematic Review.” *International Journal of Interactive Multimedia and Artificial Intelligence* 9(1):137–48. doi:10.9781/ijimai.2023.06.003.
- Teixeira, Diogo M., and Ramiro S. Barbosa. 2024. “Stock Price Prediction in the Financial Market Using Machine Learning Models.” *Computation* 13(1):3. doi:10.3390/computation13010003.
- Tuesta, Sebastian, Nahum Flores, and David Mauricio. 2025. “Prediction of the Maximum and Minimum Prices of Stocks in the Stock Market Using a Hybrid Model Based on Stacking.” *Algorithms* 18(8):471. doi:10.3390/a18080471.
- Wang, Qiaoyu, Kai Kang, Zhihan Zhang, and Demou Cao. 2021. “Application of LSTM and CONV1D LSTM Network in Stock Forecasting Model.” *Artificial Intelligence Advances* 3(1):36–43. doi:10.30564/aia.v3i1.2790.
- Wang, Yao, Jingmei Zhao, Qing Li, and Xiangyu Wei. 2024. “Considering Momentum Spillover Effects via Graph Neural Network in Option Pricing.” *Journal of Futures Markets* 44(6):1069–94. doi:10.1002/fut.22506.
- Xia, Yijun. 2025. “Research on the Optimal Prediction Model of Stock Returns of FAANG + M.” Pp. 779–86 in.
- Yaâla, Sirine B., and Jamel E. Henchiri. 2024. “Predicting Stock Market Crashes in MENA Regions: Study Based on the Irrationality of Investor Behavior and the NARX Model.” *Journal of Financial Regulation and Compliance* 32(5):590–619. doi:10.1108/jfrc-12-2023-0201.
- Yunita, A., Pratama, M. I., Almuzakki, M. Z., Ramadhan, H., Akhir, E. A. P., Firdausiah

- Mansur, A. B., & Basori, A. H. (2025). Performance analysis of neural network architectures for time series forecasting: A comparative study of RNN, LSTM, GRU, and hybrid models. *MethodsX*, 15, 103462. <https://doi.org/10.1016/j.mex.2025.103462>
- Zeng, Wei, Junjian Huang, Shiping Wen, and Zhenjiang Fu. 2023. "A Masked-Face Detection Algorithm Based on M-EIOU Loss and Improved ConvNeXt." *Expert Systems with Applications* 225:120037. doi:10.1016/j.eswa.2023.120037.
- Zhao, Ling, Yujiao Song, Chao Zhang, Yu Liu, Pu Wang, Tao Lin, Min Deng, and Haifeng Li. 2020. "T-Gcn: A Temporal Graph Convolutional Network for Traffic Prediction." *Ieee Transactions on Intelligent Transportation Systems* 21(9):3848–58. doi:10.1109/tits.2019.2935152.
- Zhao, Xiaosong, Yong Liu, and Qiangfu Zhao. 2024. "Generalized Loss-Based CNN-BiLSTM for Stock Market Prediction." *International Journal of Financial Studies* 12(3):61. doi:10.3390/ijfs12030061.