

## PENGEMBANGAN MODEL HYBRID ARIMA-MACHINE LEARNING UNTUK PREDIKSI HARGA SAHAM BCA

Prabowo Dwi Kurniawan<sup>\*1</sup>, Herman Dwi Surjono<sup>2</sup>, Handaru Jati<sup>3</sup>

<sup>1,2,3</sup>Universitas Negeri Yogyakarta, Yogyakarta

Email: <sup>1</sup>prabowodwi.2023@student.uny.ac.id, <sup>2</sup>hermansurjono@uny.ac.id, <sup>3</sup>handaru@uny.ac.id

<sup>\*</sup>Penulis Korespondensi

(Naskah masuk: 18 Maret 2025, diterima untuk diterbitkan: 15 Desember 2025)

### Abstrak

Penelitian ini bertujuan untuk menganalisis kinerja metode *hybrid* antara algoritma *machine learning* dan model ARIMA dalam memprediksi harga saham Bank BCA selama lima tahun terakhir. Data yang digunakan berasal dari saham Bank BCA periode 13 November 2019 hingga 12 November 2024, diperoleh melalui Yahoo Finance. Dataset ini terdiri dari 1210 *record* dengan tujuh variabel: Date, Open, Close, High, Low, Volume, dan Adj Close. Pengujian dilakukan memodelkan data linier menggunakan ARIMA, kemudian memprediksi residual menggunakan algoritma *machine learning* yaitu KNN, Naïve Bayes, Logistic Regression, SVM, Random Forest, dan Gradient Boost. Selanjutnya Prediksi Akhir didapatkan dari penjumlahan Prediksi ARIMA dengan Prediksi Residual oleh Machine Learning. Hasil evaluasi menunjukkan bahwa model hybrid ARIMA–SVM memberikan performa terbaik dengan nilai MSE sebesar 13.341,72, MAE sebesar 89,69, dan MAPE sebesar 0,9078%. Model ini juga memiliki nilai korelasi (R) tertinggi sebesar 0,9785. Sementara itu, model ARIMA–Gradient Boosting juga menunjukkan performa yang kompetitif dengan MSE sebesar 14.126,60 dan MAPE sebesar 0,9434%. Temuan ini menunjukkan bahwa pendekatan hybrid efektif dalam meningkatkan akurasi dan kestabilan prediksi saham, serta dapat dijadikan alternatif yang unggul dalam analisis pasar keuangan berbasis data historis.

**Kata kunci:** *Hybrid, ARIMA, Machine Learning, Saham, BBCA*

## DEVELOPMENT OF A HYBRID ARIMA-MACHINE LEARNING MODEL FOR PREDICTING BCA STOCK PRICES

### Abstract

*This study aims to analyze the performance of a hybrid method combining machine learning algorithms and the ARIMA model in predicting the stock prices of Bank BCA over the past five years. The data used were obtained from Yahoo Finance, covering the period from November 13, 2019, to November 12, 2024. The dataset consists of 1,210 records and includes seven variables: Date, Open, Close, High, Low, Volume, and Adjusted Close. The testing procedure involved modeling the linear component of the data using ARIMA, followed by predicting the residuals with machine learning algorithms, namely K-Nearest Neighbors (KNN), Naïve Bayes, Logistic Regression, Support Vector Machine (SVM), Random Forest, and Gradient Boosting. The final prediction was obtained by summing the ARIMA forecast with the residual predictions from the machine learning models. Evaluation results show that the hybrid ARIMA–SVM model delivered the best performance with an MSE of 13,341.72, MAE of 89.69, and MAPE of 0.9078%, along with the highest correlation (R) value of 0.9785. The ARIMA–Gradient Boosting model also demonstrated competitive performance with an MSE of 14,126.60 and a MAPE of 0.9434%. These findings indicate that the hybrid approach is effective in enhancing the accuracy and stability of stock price predictions and can serve as a promising alternative in historical data-based financial market analysis.*

**Keywords:** *Hybrid, ARIMA, Machine Learning, Stock, BBCA*

### 1. PENDAHULUAN

Saham adalah alat investasi yang memiliki harga yang cenderung berfluktuasi (Soewignjo et al., 2023). Saham memiliki potensi untuk memberikan keuntungan yang signifikan. Namun, karena harga saham cenderung bergerak naik turun dengan cepat,

hal ini membuat investasi saham memiliki tingkat risiko yang cukup tinggi. Keuntungan dari investasi saham diperoleh melalui capital gain dan dividen. Untuk meraih capital gain yang maksimal, penting untuk menentukan harga beli dan jual dengan tepat agar selisihnya optimal (Saputro & Swanjaya, 2023). Untuk mengambil keputusan yang tepat dalam

kebijakan investasi, diperlukan metode yang efektif untuk memprediksi harga saham. Harga penutupan saham yang fluktuatif dan sangat sensitif membuatnya sulit diprediksi secara konvensional. Oleh karena itu, penting bagi investor untuk menggunakan pendekatan yang dapat meminimalkan risiko investasi (Hastomo et al., 2021). Prediksi harga penutupan saham penting untuk menilai kinerja perusahaan. Pasar saham yang efisien juga berhubungan positif dengan pertumbuhan ekonomi suatu negara, baik jangka pendek maupun panjang (Saluza et al., 2021).

Model ARIMA terbukti efektif dalam memprediksi harga saham, terutama dalam jangka waktu pendek (Faisal, 2021). Penggunaan metode ini sangat tepat karena memiliki tingkat akurasi yang tinggi dalam menghasilkan prediksi (Khaira et al., 2021). ARIMA digunakan untuk memodelkan komponen tren dalam data deret waktu, sehingga dapat mengidentifikasi pola jangka panjang dan fluktuasi yang lebih stabil, yang penting dalam memproyeksikan pergerakan harga di masa depan (Pierre et al., 2023). Meskipun demikian Model ARIMA memiliki kelemahan utama karena hanya mampu mengikuti pola urutan waktu linear, yang berarti ia tidak dapat menangkap dinamika kompleks dan non-linear yang sering terjadi dalam pergerakan harga saham (Amrullah et al., 2022).

Fluktuasi harga saham dipengaruhi oleh berbagai faktor, seperti permintaan dan penawaran, kondisi sosial, politik, serta faktor lainnya yang memengaruhi pergerakan harga saham (Puteri, 2023). Salah satu pendekatan yang dapat digunakan untuk memprediksi pergerakan harga saham adalah penerapan metode prediktif, seperti machine learning, yang menawarkan potensi baru dalam menyelesaikan masalah tersebut (Mahendra et al., 2023). Algoritma machine learning, seperti Random Forest, Gradient Boosting Regressor, Support Vector Machine (SVM), dan K-Nearest Neighbors (KNN), semakin populer dalam analisis keuangan berkat kemampuannya dalam menangani data non-linear dan pola kompleks. Meskipun demikian, algoritma machine learning juga memiliki keterbatasan, terutama dalam menangkap pola time series yang eksplisit.

Penggabungan model time series dan machine learning yang dikenal sebagai model hybrid muncul sebagai pendekatan alternatif untuk meningkatkan akurasi prediksi harga saham. Model hybrid memungkinkan model ARIMA untuk menangkap pola tren jangka panjang, sedangkan algoritma machine learning menangkap pola non-linear dan variasi jangka pendek yang mungkin tidak teridentifikasi dengan ARIMA. Pendekatan ini bertujuan untuk mengatasi kelemahan model individu dan memberikan hasil prediksi yang lebih akurat.

penelitian ini bertujuan untuk mengembangkan model hybrid ARIMA-Machine Learning untuk mendapatkan metode terbaik prediksi harga saham

BCA dan melakukan studi komparatif terhadap akurasi prediksi dari masing-masing model. Diharapkan hasil penelitian ini dapat memberikan kontribusi signifikan dalam pengembangan metode prediksi harga saham di Indonesia dan memberikan wawasan baru dalam penggunaan model hybrid untuk prediksi saham di pasar negara berkembang.

## 2. KAJIAN PUSTAKA

### A. Saham BCA

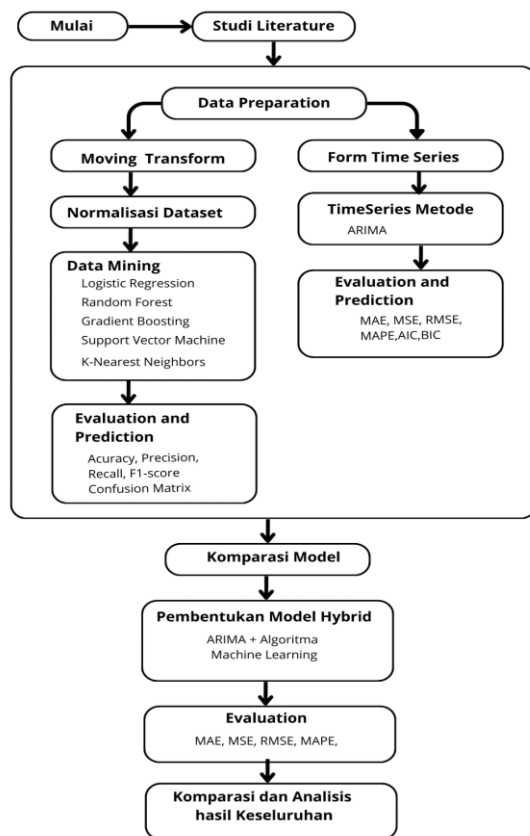
Bank Central Asia (BCA) merupakan salah satu bank swasta terkemuka di Indonesia yang bergerak di sektor industri keuangan, khususnya dalam subsektor perbankan. BCA telah berhasil memperluas jangkauannya tidak hanya di seluruh wilayah Indonesia, tetapi juga di pasar internasional (Aditya et al., 2023). Saham BCA merupakan salah satu saham blue-chip terbesar di Indonesia dengan kapitalisasi pasar yang signifikan, menjadikannya acuan utama bagi banyak investor dalam mengambil keputusan investasi. Harga saham yang fluktuatif membuat investor perlu strategi untuk memprediksi harga dan mengambil keputusan investasi yang tepat.

### B. Model Hybrid

Model hybrid yang menggabungkan berbagai teknik machine learning dan optimasi menawarkan solusi yang efektif untuk menangani masalah kompleks di berbagai bidang. Dengan mengintegrasikan model-model yang berbeda dan menggunakan teknik optimasi seperti tuning hyperparameter dan optimasi multi-objektif, model hybrid dapat meningkatkan akurasi, ketahanan, dan generalisasi prediksi. Selain itu, metode ensemble dan perhatian pada interpretabilitas memastikan model tetap transparan dan dapat dipahami, terutama dalam pengambilan keputusan yang penting. Melalui proses iteratif seperti eksperimen dan adaptasi berkelanjutan, model hybrid dapat terus disempurnakan untuk mencapai hasil yang optimal dan inovatif. (Giroh et al., 2023). Hybrid model ini merupakan penggabungan antara dua metode yang dinilai memiliki tingkat akurasi yang tinggi. Terdapat penelitian yang membandingkan hasil akurasi dari Machine Learning Explicit, Implicit, dan Hybrid model yang diimplementasikan pada diagnosa medis penyakit jantung dengan metode logistic regression dan random forest dimana menunjukkan peningkatan akurasi sebesar 1.32% (Silmi et al., 2022).

## 3. METODE PENELITIAN

Penelitian yang dilakukan untuk mendapatkan hasil analisis saham menggunakan skema alur penelitian sebagai berikut :



Gambar 1. Diagram alur penelitian

#### A. Studi Literature

Dalam mengembangkan penelitian ini dimulai dengan mengumpulkan data pendukung, termasuk informasi tentang teknik pembelajaran mesin untuk prediksi. Sumber data mencakup buku, publikasi penelitian nasional dan internasional, serta database yang diambil dari situs *Yahoo Finance*. (<https://finance.yahoo.com/quote/BBCA.JK/>)

#### B. Data Preparation

Pada tahap ini, dataset direduksi dengan menghapus data kosong dan data duplikat untuk memastikan kualitas data yang digunakan dalam analisis atau model machine learning. Proses penghapusan data kosong dan data duplikat ini biasa disebut dengan proses data cleaning atau data cleansing. Pengecekan awal dilakukan untuk memastikan keberadaan data kosong dan duplikat. Jika tidak ditemukan, seluruh data dalam dataset dapat digunakan.

#### C. Moving Transform

Merupakan sebuah menu untuk mengaplikasikan transformasi data berbasis pergeseran atau perhitungan dalam jendela waktu tertentu, seperti rata-rata bergerak (*moving average*), total kumulatif, atau fungsi statistik lainnya. Salah satu kelebihanannya adalah dapat secara dinamis untuk mengidentifikasi tren, pola, atau perubahan dalam data seiring waktu. Moving Transform sangat berguna dalam analisis

deret waktu (*time series*), memungkinkan analisis yang lebih mendalam terhadap fluktuasi data, baik dalam skala kecil maupun besar.

#### D. Form Time series

Merupakan sebuah menu untuk mengatur data menjadi format deret waktu (*time series*), yang disusun berdasarkan urutan waktu. Fitur ini mempermudah analisis data yang bergantung pada kronologi, seperti tren penjualan, perubahan harga, atau pola cuaca, dengan menyesuaikan variabel waktu dan intervalnya.

#### E. Normalisasi Dataset

Pada tahap ini, dilakukan normalisasi data, yaitu prosedur yang menyiapkan data agar sesuai dengan kebutuhan pengguna dan membantu operasi di masa depan dalam menghasilkan hasil yang lebih baik. Prosedur normalisasi ini mengubah data menjadi skala yang konsisten, memastikan bahwa semua fitur dalam kumpulan data berkontribusi secara setara terhadap analisis atau pelatihan model pembelajaran mesin. Dalam penelitian ini, normalisasi data dilakukan menggunakan metode Min-Max dengan cara sebagai berikut:

$$x' = \left[ \left( \frac{x - \text{oldMin}}{\text{oldMax} - \text{oldMin}} \right) * (\text{newMax} - \text{newMin}) \right] + \text{newMin} \quad (1)$$

Sampel data yang dinormalisasi  $x'$  dapat diperoleh dari sampel data asli  $x$ . Proses ini bergantung pada nilai maksimum dan minimum atribut yang sama di antara instance data. Dalam metode normalisasi, nilai komponen dari sampel data asli akan diubah ke dalam rentang  $[0,1]$ . Dimana  $x'$  adalah sampel data yang dinormalisasi,  $x$  adalah data asli sampel data,  $\text{oldMin}$  adalah data minimum di antara atribut apa pun dari kumpulan data asli,  $\text{oldMax}$  adalah data maksimum di antara atribut apa pun dari kumpulan data asli,  $\text{newMin}$  adalah minimumnya dari kumpulan data yang dinormalisasi, dan  $\text{newMax}$  adalah maksimum kumpulan data yang dinormalisasi. (Muhammad Ali, 2022).

#### F. Data Mining

Penelitian ini menggunakan data harian harga saham Bank BCA selama lima tahun terakhir. Sebelum data tersebut digunakan dalam pembangunan model prediksi, data diproses terlebih dahulu (Supriyanto et al., 2023). Proses ini meliputi pembersihan data, transformasi, dan pemilihan variabel yang relevan untuk memastikan bahwa data yang digunakan berkualitas dan dapat memberikan hasil yang akurat dalam model prediksi.

##### F.1. Metode Random Forest

Random Forest adalah metode pembelajaran ansambel berbasis *bagging* yang menggabungkan beberapa pohon keputusan acak dan menghasilkan prediksi akhir dengan cara rata-rata. Data pelatihan

diambil ulang menggunakan metode *Bootstrap* untuk membentuk kumpulan baru, lalu setiap kumpulan digunakan untuk membangun pohon klasifikasi atau regresi dengan metode CART (*Classification and Regression Trees*). Hasil akhir diperoleh dengan menggabungkan prediksi dari semua pohon keputusan (Wang et al., 2023). Model Random Forest dapat di tunjukan sebagai berikut :

$$\hat{y} = \frac{1}{B} \sum_{i=1}^B T_i(x) \quad (2)$$

Dimana  $\hat{y}$  dalah prediksi akhir, B adalah jumlah total pohon, dan  $T_i(x)$  adalah prediksi dari pohon ke-i.

#### F.2. Metode Support Vector Regression

Support Vector Regression adalah pendekatan terkenal yang pertama kali diperkenalkan oleh Vapnik pada tahun 1999. SVR (Support Vector Regression) memiliki potensi yang sangat baik untuk mengurangi overfitting karena prinsip minimalisasi risiko struktural yang membuat keseimbangan antara keakuratan prediksi dan kompleksitas fungsi regresi. Ketika terdapat data outlier, parameter dalam model regresi konvensional bisa menjadi sangat besar untuk mengakomodasi outlier tersebut, menyebabkan overfitting. Sebaliknya, SVR memiliki kinerja generalisasi yang baik karena fungsi objek optimasinya memastikan fungsi regresi tetap halus dan rata. (Liao et al., 2024).

#### F.3. Metode K-Nearest Neighbor (KNN)

K-Nearest Neighbor adalah algoritma sederhana untuk mengklasifikasikan objek baru dengan membandingkan atributnya dengan data pelatihan yang ada. Algoritma ini tidak membangun model prediktif seperti halnya algoritma lain, melainkan mengandalkan memori dari data pelatihan. Dalam KNN, objek baru diklasifikasikan berdasarkan mayoritas dari K tetangga terdekatnya. Proses ini melibatkan perhitungan jarak antara objek baru dengan sampel pelatihan untuk menentukan K tetangga terdekat (S, 2023).

Tahapan penerapan pendekatan KNN dimulai dengan menentukan nilai jumlah tetangga terdekat (K). Langkah selanjutnya adalah menghitung jarak Euclidean dan mengurutkan data berdasarkan jarak Euclidean terpendek. Setelah itu, dibentuk kelompok K dalam KNN, di mana nilai taksiran paling akurat diprediksi oleh kelompok Tetangga Terdekat. Data kemudian dipisahkan menjadi set pelatihan dan pengujian menggunakan nilai K yang berbeda, dan matriks konfusinya dihitung. Terakhir, nilai presisi, recall, dan F-measure dihitung berdasarkan matriks konfusi tersebut.

Untuk mencari jarak Euclidean digunakan formula sebagai berikut :

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

Dimana

n = Dimensi data

i = Variabel data

x,y = Dua titik dalam ruang Eculidean

d = Jarak

#### F.4. Logistic Regression

Logistic Regression adalah algoritma klasifikasi yang digunakan untuk meramalkan probabilitas hubungan antara variabel dependen dan independen. Algoritma ini mengukur kemungkinan suatu hasil tertentu, sehingga cocok untuk tugas klasifikasi biner maupun multi-kelas (Afiatuddin et al., 2024). Regresi logistik, yang juga dikenal sebagai regresi logit, classifier entropi maksimum (MaxEnt), atau classifier log-linier, adalah teknik yang digunakan untuk memperkirakan probabilitas hasil biner. Teknik ini menerapkan fungsi logistik pada logit, yang menunjukkan logaritma dari probabilitas terjadinya suatu peristiwa. Metode ini memudahkan pemodelan hubungan antara variabel dependen dan independen dalam tugas klasifikasi (Akbar & Rahmaddeni, 2022).

#### F.5. Gradient Boosting

Gradient boosting adalah metode yang kuat dan efektif dalam menangkap ketergantungan fungsi non-linear yang kompleks. Keluarga model ini telah menunjukkan keberhasilan yang signifikan dalam berbagai aplikasi praktis, seperti pengklasifikasian, regresi, dan prediksi (Natekin & Knoll, 2013). Keunggulannya terletak pada kemampuannya untuk meningkatkan kinerja model secara bertahap dengan menggabungkan hasil dari sejumlah model lemah, menghasilkan prediksi yang lebih akurat dan stabil. Dengan kata lain algoritma gradient boosting beroperasi dengan membangun model secara bertahap, di mana setiap model baru dirancang untuk memperbaiki kesalahan yang terjadi pada model sebelumnya (Nasional et al., 2024).

#### F.6. Naïve Bayes

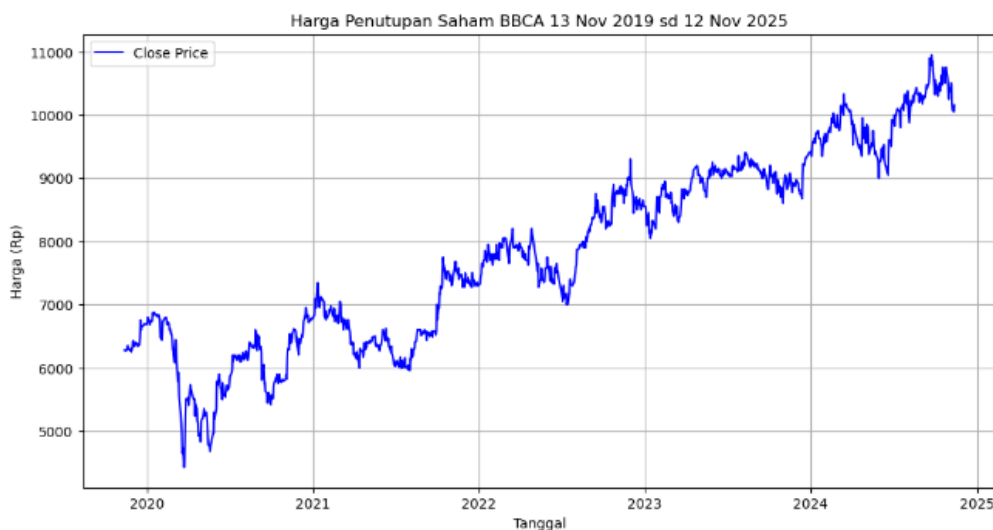
Algoritma klasifikasi Naïve Bayes adalah alat prediktif yang menawarkan keunggulan seperti efektivitas dalam menghasilkan temuan yang valid dan efisiensi dalam pemrosesan input yang cepat (Akbar & Rahmaddeni, 2022). Teorema Bayes dapat diimplementasikan menggunakan persamaan berikut.

$$P(H|X) = \frac{(P(X|H) \times P(H))}{(P(X))} \quad (4)$$

#### G. ARIMA

Model ARIMA (AutoRegressive Integrated Moving Average) adalah metode prediksi statistik yang digunakan untuk menganalisis dan memproyeksikan data deret waktu berdasarkan hubungan korelasi antar data (Saluza et al., 2021). Model ARIMA telah terbukti efektif dalam memberikan prediksi yang akurat untuk periode jangka pendek (Latif et al., 2023).





Gambar 2. Grafik Saham BBKA Periode 5 Tahun

Dalam data time series yang memiliki autokorelasi, terdapat hubungan antara nilai saat ini dengan nilai sebelumnya (lag). Model autoregresi memanfaatkan hubungan ini untuk memprediksi nilai masa depan berdasarkan data historis melalui persamaan regresi. Metode ARIMA sangat tepat digunakan untuk peramalan jangka pendek karena dapat menghasilkan prediksi yang akurat dan mengidentifikasi hubungan statistik yang kuat antara variabel yang diprediksi dengan nilai yang digunakan. Namun, untuk peramalan jangka panjang, tingkat akurasi cenderung menurun (Khaira et al., 2021).

#### H. Hybrid ARIMA – Machine Learning

Metode hybrid ARIMA dan machine learning meningkatkan analisis saham dengan mengintegrasikan keunggulan dari kedua metodologi tersebut. ARIMA sangat efektif untuk peramalan jangka pendek karena kemampuannya untuk mengidentifikasi pola-pola yang ada dalam data time series. Namun, ARIMA kurang mampu menangkap interaksi non-linear dan variabel eksternal yang kompleks. Di sisi lain, algoritma machine learning seperti Random Forest atau SVM dapat mengakomodasi korelasi non-linear dan beradaptasi dengan data yang masuk, namun kesulitan dalam merepresentasikan pola jangka panjang. Pendekatan hibrida ini memanfaatkan keunggulan ARIMA dalam peramalan tren jangka pendek dan kemampuan model machine learning dalam menangani kompleksitas, menghasilkan prediksi saham yang lebih akurat dan dapat diandalkan.

#### I. Evaluation Model

Penilaian model prediksi menggunakan berbagai ukuran untuk mengukur kinerjanya agar dapat memahami efektivitas model secara menyeluruh. Metrik seperti Mean Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Squared Error

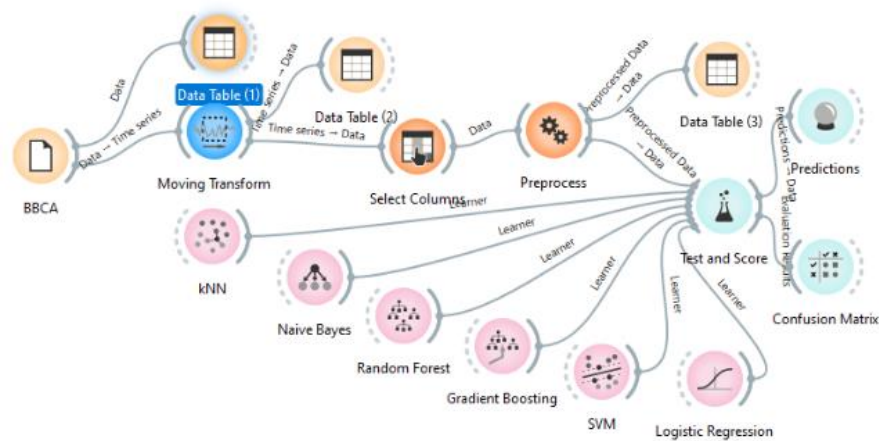
(RMSE) digunakan untuk mengevaluasi keakuratan model regresi dengan mengukur selisih antara nilai prediksi dan nilai aktual. Mean Absolute Percentage Error (MAPE) memberikan penilaian relatif terhadap kesalahan prediksi, sehingga cocok untuk membandingkan performa antar dataset. Selain itu, Akaike Information Criterion (AIC) dan Bayesian Information Criterion (BIC) penting dalam memilih model terbaik, dengan memberi penalti pada model yang terlalu kompleks untuk menghindari overfitting.

### 4. HASIL DAN PEMBAHASAN

Analisis harga saham BBKA menggunakan algoritma machine learning dan ARIMA dilakukan menggunakan kombinasi aplikasi orange dan jupyter notebook. Dataset harga saham BBKA yang didapatkan dari situs yahoo finance pada periode waktu 5 tahun (13 November 2019 – 12 November 2024). Data yang didapatkan berjumlah 1222 record dimana terdiri dari 7 variabel. Dalam proses cleaning data ditemukan nilai record kosong sebanyak 12 sehingga data yang digunakan menjadi 1210. Belum semua data dalam record dataset bernilai numerik sehingga dilakukan konversi. Dalam pengolahan data, semua variabel digunakan yakni Date, Open, High, Low, Close, AdjClose dan Volume. Grafik saham BBKA selama 5 tahun dari 13 November 2019 sampai 12 November 2024 digambarkan pada gambar 2.

#### 4.1. Analisis Prediksi dengan Algoritma Machine Learning

Pada analisis prediksi menggunakan algoritma machine learning, diterapkan beberapa model, yaitu K-Nearest Neighbors (KNN), Naive Bayes, Random Forest, Gradient Boosting, Support Vector Machine (SVM), dan Logistic Regression. Proses analisis ini dilakukan menggunakan aplikasi Orange, yang ditunjukkan pada Gambar 2 berikut.



Gambar 3. Analisis saham BBCA dengan Machine learning

Tabel 1. Hasil Evaluation Model ML pada BBCA

Model	AUC	CA	F1	Prec	Recall	MCC
kNN	0.495	0.507	0.502	0.5	0.507	-0.009
Naïve Bayes	0.482	0.517	0.516	0.516	0.517	0.023
Random Forrest	0.489	0.493	0.489	0.488	0.493	-0.035
Gradient Boosting	0.488	0.502	0.5	0.499	0.502	-0.012
SVM	0.524	0.51	0.511	0.512	0.51	0.014
Logistic Regression	0.544	0.568	0.492	0.57	0.568	0.091

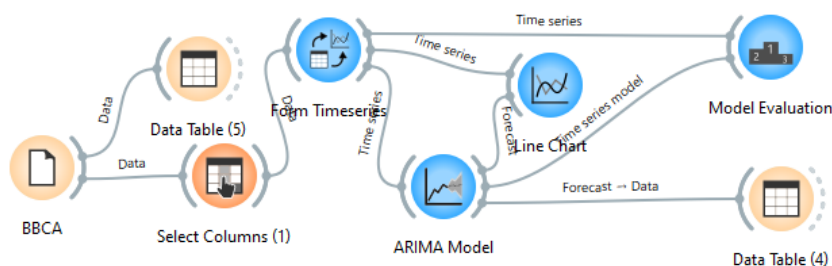
Beberapa tahapan yang dilakukan antara lain moving transform dan normalisasi. Moving transform digunakan untuk melakukan transformasi data berdasarkan jendela (window) yang bergerak yang sangat berguna dalam analisis data time series untuk menghitung nilai statistik tertentu atau menerapkan transformasi pada data dalam interval waktu tertentu. Sedangkan Normalisasi yang dipilih dengan skala -1 hingga 1 adalah teknik preprocessing data yang bertujuan untuk menyamakan rentang nilai dataset agar konsisten. Hal ini penting untuk algoritma pembelajaran mesin, seperti SVM, KNN, yang sensitif terhadap perbedaan skala data. Dengan normalisasi, data yang memiliki variasi besar menjadi lebih seimbang, sehingga mempercepat proses optimasi, meningkatkan akurasi, dan meminimalkan efek outlier. Hasil evaluation model dari algoritma machine learning disajikan pada Tabel 1.

Berdasarkan Tabel 1 diatas, enam algoritma machine learning diuji menggunakan metrik AUC (Area Under the Curve), CA (Classification Accuracy), F1 Score, Precision, Recall, dan MCC (Matthews Correlation Coefficient). Logistic Regression menunjukkan performa terbaik dengan AUC sebesar 0.544, CA sebesar 0.568, dan MCC sebesar 0.091, menjadikannya model paling andal dalam membedakan kelas dan memberikan prediksi yang akurat pada dataset ini. SVM menempati posisi kedua dengan AUC sebesar 0.524, tetapi nilai CA (0.510) dan MCC (0.014) menunjukkan bahwa model ini masih kurang optimal dibanding Logistic Regression. Naive Bayes juga memberikan hasil

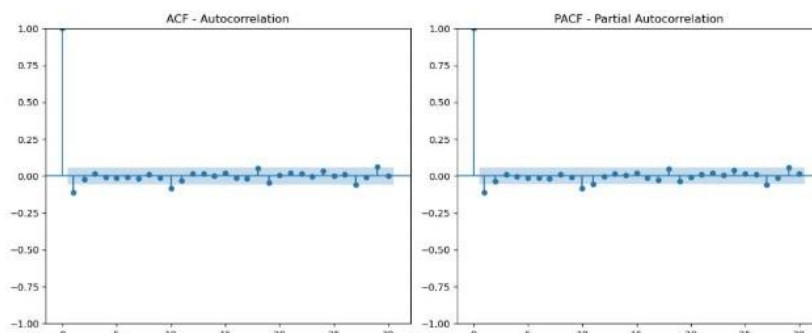
yang cukup kompetitif, dengan MCC positif sebesar 0.023 dan metrik lainnya yang relatif stabil, menjadikannya alternatif yang layak. Model Random Forest dan Gradient Boosting memiliki performa yang hampir serupa dengan AUC masing-masing sebesar 0.489 dan 0.488, namun nilai CA dan MCC yang lebih rendah menunjukkan bahwa kedua model ini kurang mampu menangkap pola data dengan baik. Sementara itu, kNN mencatatkan performa terendah dengan AUC 0.495 dan MCC negatif (-0.009), yang mengindikasikan bahwa model ini tidak cocok untuk dataset ini. Secara keseluruhan, Logistic Regression merupakan algoritma terbaik untuk kasus ini, diikuti oleh SVM dan Naive Bayes, sedangkan model lain memerlukan optimasi lebih lanjut untuk meningkatkan performanya.

Jika dianalisis lebih mendalam, model model yang diuji menunjukkan performa yang secara keseluruhan masih kurang optimal untuk analisis prediksi saham. Nilai AUC terbaik, yaitu 0.544 dari Logistic Regression, mengindikasikan bahwa kemampuan model untuk membedakan antara kelas masih mendekati acak (idealnya  $AUC > 0.7$  untuk performa yang baik). Selain itu, nilai CA (Classification Accuracy) tertinggi sebesar 0.568 juga relatif rendah, menunjukkan bahwa model gagal memberikan akurasi yang signifikan. Hal ini diperkuat oleh nilai MCC yang rendah (maksimal 0.091), yang menandakan korelasi lemah antara prediksi model dengan label sebenarnya.

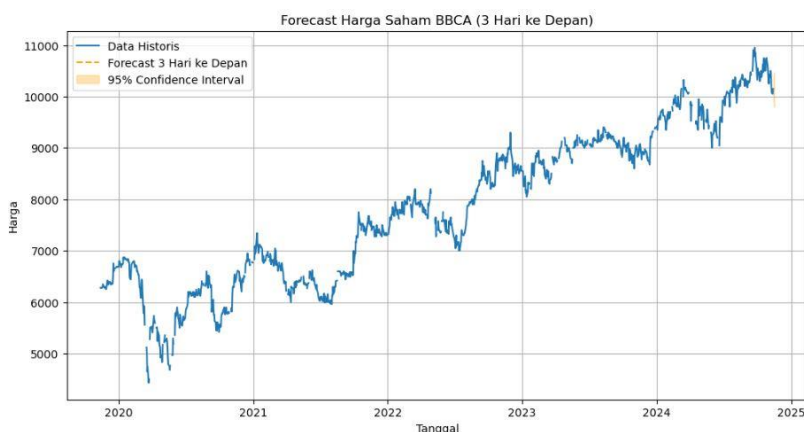




Gambar 4. Analisis Saham dengan ARIMA



Gambar 5. Grafik ACF dan PACF



Gambar 6. Forecast harga saham BBKA 3 hari kedepan

#### 4.2. Analisis Prediksi dengan ARIMA

Dalam analisis prediksi menggunakan ARIMA, data target yang digunakan adalah nilai *Close* untuk memprediksi harga saham secara kontinu. Analisis prediksi dengan Arima menggunakan aplikasi orange ditunjukkan pada Gambar 4.

Analisis awal dilakukan dengan menguji data harga penutupan saham BBKA bersifat stasioner atau tidak. Pengujian dilakukan menggunakan metode Augmented Dickey-Fuller (ADF), yang menunjukkan nilai statistik ADF sebesar -0.92 dengan p-value 0.77. Nilai p-value yang lebih besar dari 0.05 menunjukkan bahwa data belum stasioner pada level aslinya. Langkah selanjutnya melakukan proses differencing sebanyak satu kali. Dari hasil differencing pertama, nilai statistik ADF berubah menjadi -38.83 dengan p-value 0.0000, yang berarti data telah menjadi stasioner dan nilai  $d$  dalam model ARIMA ditetapkan sebesar 1.

Langkah selanjutnya dilakukan analisis grafik Autocorrelation Function (ACF) dan Partial Autocorrelation Function (PACF) terhadap data yang telah didifferencing. Grafik ACF menunjukkan cut-off tajam pada lag ke-1, sedangkan grafik PACF juga menunjukkan penurunan signifikan pada lag ke-1. Berdasarkan karakteristik ini, dapat disimpulkan bahwa nilai orde AR ( $p$ ) adalah 1 dan nilai orde MA ( $q$ ) juga 1. Dengan demikian, model ARIMA yang diusulkan untuk prediksi harga saham BBKA adalah ARIMA(1,1,1). Grafik ACF dan PACF ditunjukkan pada gambar 5.

Model ARIMA (1,1,1) digunakan untuk memprediksi harga saham BBKA dalam jangka pendek, khususnya tiga hari ke depan. Parameter  $p = 1$  merepresentasikan satu lag autoregresif,  $d = 1$  menunjukkan proses differencing satu kali untuk menjadikan data stasioner, dan  $q = 1$  melibatkan satu lag kesalahan sebelumnya.



Tabel 2. Prediksi Saham BBKA dengan ARIMA Model

Day	Close(Forecast)	Close(95% CI Low)	Close (95% CI High)
1	10137.9	9924.25	10351.7
2	10137.18	9854.62	10419.74
3	10137.12	9800.59	10473.65

Tabel 3. Hasil Evaluation Model ARIMA pada BBKA

	RMSE	MAE	MAPE	POCID	R2	AIC	BIC
ARIMA(1,1,1)	113.71	82.22	1.089	39.90	0.994	14838	14853
ARIMA(1,1,0) (in-sample)	113.78	82.14	1.087	39.40	0.994	14837	14847

Grafik hasil prediksi memperlihatkan kecenderungan harga saham BBKA yang tetap mengikuti tren historis, dengan penambahan garis prediksi dan rentang kepercayaan 95% yang menggambarkan estimasi model beserta tingkat ketidakpastiannya secara visual. Rentang prediksi yang sempit mencerminkan tingkat kepercayaan yang tinggi terhadap hasil estimasi model. Grafik hasil prediksi dengan model ARIMA di tunjukan pada gambar 6.

Nilai prediksi 3 hari kedepan menggunakan analisis ARIMA Model ditampilkan pada Tabel 3. Hasil evaluation model ARIMA pada Bank BCA disajikan dalam Tabel 3.

Model ARIMA dengan parameter (1,1,1) dan (1,1,0) menghasilkan performa prediksi yang sangat mirip dalam hal akurasi, ditunjukkan oleh nilai RMSE, MAE, MAPE, dan  $R^2$  yang hampir identik. Nilai MAPE di bawah 1.1% menunjukkan bahwa kedua model mampu memberikan estimasi yang sangat akurat terhadap harga saham BBKA. Namun, ARIMA(1,1,0) sedikit lebih unggul dalam hal kriteria kompleksitas model berdasarkan nilai AIC dan BIC yang lebih rendah, meskipun POCID (directional accuracy)-nya sedikit lebih rendah dibandingkan ARIMA(1,1,1). Hal ini menunjukkan trade-off antara kesederhanaan model dan ketepatan arah tren prediksi.

#### 4.3. Analisa Prediksi model Hybrid ARIMA – Machine Learning

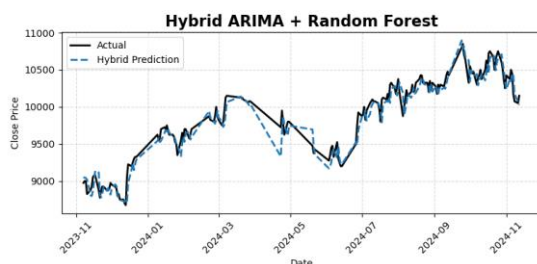
Metode hybrid dalam analisis prediksi saham BCA dilakukan dengan menggabungkan model ARIMA dan algoritma machine learning (ML) untuk meningkatkan akurasi prediksi. Beberapa langkah

yang diterapkan adalah menerapkan model ARIMA pada data historis saham untuk menangkap pola linear dan menghasilkan prediksi awal. Setelah itu, dihitung residual atau selisih antara nilai aktual dan prediksi ARIMA. Residual ini kemudian diprediksi menggunakan algoritma machine learning yang lebih mampu menangkap pola non-linear. Hasil akhir prediksi saham diperoleh dengan menjumlahkan hasil prediksi ARIMA dan prediksi residual dari model machine learning. Pendekatan ini memanfaatkan kelebihan ARIMA dalam menangani komponen time series dan keunggulan machine learning dalam memodelkan ketidakpastian serta pola kompleks, sehingga dapat memberikan hasil prediksi yang lebih akurat dan optimal. Jadi pendekatan konsep yang digunakan adalah

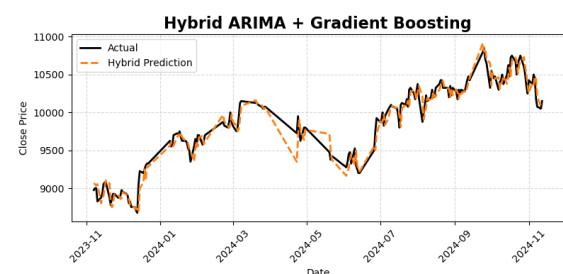
$$\text{Prediksi Akhir} = \text{Prediksi ARIMA} + \text{Prediksi Residual oleh ML}$$

Dalam analisis metode hybrid ARIMA–Machine Learning, tidak semua algoritma machine learning dapat diterapkan secara efektif. Metode seperti Naïve Bayes dan Logistic Regression lebih cocok digunakan untuk permasalahan klasifikasi, bukan regresi. Oleh karena itu, kedua metode tersebut tidak digunakan dalam analisis prediksi harga saham ini karena tidak sesuai dengan karakteristik data dan tujuan prediksi yang bersifat numerik kontinu.

Hasil dari metode Hybrid Arima dengan Beberapa model machine learning antara lain Random Forest, Gradient Boosting, SVM dan KNN ditunjukan pada gambar 7.



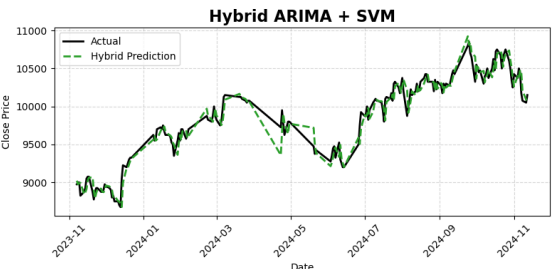
Gambar 7. Grafik Analisis Hybrid ARIMA a-Random Forest



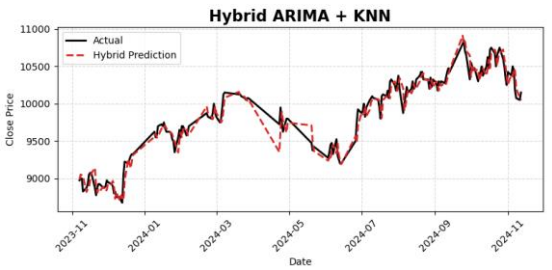
Gambar 8. Grafik Analisis Hybrid ARIMA-Gradient Boosting

Tabel 4. Hasil Evaluation Model Hybrid ARIMA-ML pada BBKA

Model Hybrid	MSE	MAE	MAPE (%)	R	POCID	AIC	BIC
ARIMA - Random Forrest	14987.07	95.96	0.9733	0.977	47.5936	14838.4	14853.93
ARIMA – KNN	14333.20	94.37	0.9592	0.9773	45.9893	14838.41	14853.93
ARIMA – Gradient Boost	14126.60	92.98	0.9434	0.9778	48.1283	14838.41	14853.93
ARIMA – SVM	13341.72	89.69	0.9078	0.9785	47.5936	14838.4	14853.93



Gambar 9. Grafik Analisis Hybrid ARIMA -SVM



Gambar 10. Grafik Analisis Hybrid ARIMA-KNN

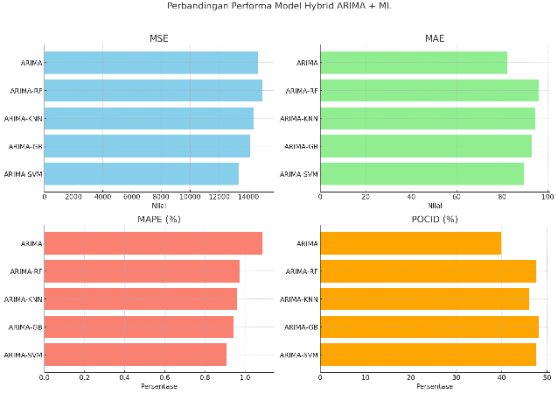
Hasil implementasi metode hybrid ARIMA dengan algoritma machine learning pada prediksi harga saham BCA menunjukkan bahwa pendekatan ini mampu meningkatkan akurasi prediksi dibandingkan penggunaan metode tunggal. Kombinasi ARIMA dengan Random Forest, Gradient Boosting, SVM, dan KNN secara visual menunjukkan kemampuan dalam mengikuti pola tren historis harga saham. Model hybrid ARIMA + Gradient Boosting menghasilkan prediksi yang paling stabil dan mendekati data aktual, sedangkan model lainnya juga mampu menangkap pola pergerakan harga meskipun terdapat variasi dalam sensitivitas terhadap perubahan tren. Pendekatan ini membuktikan bahwa penggabungan model linier ARIMA dengan machine learning yang menangani residual secara non-linier dapat menghasilkan prediksi yang lebih akurat dan adaptif terhadap dinamika pasar saham.

Hasil evaluasi model hybrid ARIMA – Machine Learning disajikan dalam Tabel 4.

Selanjutnya, untuk mengetahui performa model sebelum dan sesudah penerapan model hybrid ARIMA–Machine Learning, disajikan pada Gambar 11

Hasil evaluasi performa model ditampilkan dalam bentuk visualisasi grafik batang horizontal yang membandingkan model ARIMA dengan berbagai kombinasi hybrid ARIMA dan algoritma machine learning, yaitu Random Forest, K-Nearest

Neighbors (KNN), Gradient Boosting, dan Support Vector Machine (SVM).



Gambar 11. Grafik Batang Akurasi Prediksi Analisis Saham BBKA sebelum dan sesudah Hybrid

Berdasarkan metrik Mean Squared Error (MSE) dan Mean Absolute Percentage Error (MAPE), model hybrid ARIMA–SVM menunjukkan performa terbaik dengan nilai MSE sebesar 13.341,72 dan MAPE sebesar 0,9078%, yang menandakan tingkat kesalahan prediksi yang paling rendah dibandingkan model lainnya. Sementara itu, model ARIMA murni memiliki nilai Mean Absolute Error (MAE) terendah, yaitu 82,22, namun tidak diikuti oleh kinerja superior pada metrik lainnya. Dari sisi kemampuan memprediksi arah perubahan harga saham (POCID), model hybrid ARIMA–Gradient Boosting unggul dengan nilai tertinggi sebesar 48,13%, menunjukkan keandalannya dalam mengidentifikasi arah tren pasar secara lebih akurat. Secara umum, hasil ini mengindikasikan bahwa pendekatan hybrid memberikan peningkatan performa prediksi dibandingkan model ARIMA tunggal, terutama dalam konteks akurasi arah tren dan kesalahan persentase.

## 5. KESIMPULAN

Berdasarkan hasil evaluasi performa model hybrid ARIMA dengan berbagai algoritma machine learning, dapat disimpulkan bahwa pendekatan hybrid secara konsisten menunjukkan kemampuan prediksi yang baik terhadap pergerakan harga saham BCA. Model ARIMA–SVM memberikan performa terbaik secara keseluruhan, diikuti oleh ARIMA–Gradient Boosting, ARIMA–KNN, dan ARIMA–Random Forest. Seluruh model hybrid mampu menangkap pola data dengan tingkat kesalahan yang relatif rendah dan korelasi yang tinggi terhadap data aktual. Hal ini menunjukkan bahwa integrasi antara

model time series ARIMA dan algoritma machine learning efektif dalam meningkatkan akurasi dan kestabilan prediksi harga saham, serta layak dipertimbangkan sebagai pendekatan yang unggul dalam analisis pasar keuangan.

## DAFTAR PUSTAKA

- ADITYA, P. T., ANDRIYADI, W., & SIDJAYA, J. A. 2023. Analisis Manajemen Strategik: PT Bank Central Asia Tbk (BCA). *Jurnalku*, 3(1). <https://doi.org/10.54957/jurnalku.v3i1.320>
- AFIATUDDIN, N., WICAKSONO, M. T., AKBAR, V. R., RAHMADDENI, R., & WULANDARI, D. 2024. Komparasi Algoritma Machine Learning dalam Klasifikasi Kanker Payudara. *Jurnal Media Informatika Budidarma*, 8(2), 889. <https://doi.org/10.30865/mib.v8i2.7457>
- AKBAR, F., & RAHMADDENI. 2022. Jurnal Politeknik Caltex Riau Komparasi Algoritma Machine Learning untuk Memprediksi Penyakit Alzheimer. *Jurnal Komputer Terapan*, 8(2), 236–245. <https://jurnal.pcr.ac.id/index.php/jkt/>
- AMRULLAH, A., SOESANTO, O., & MAISARAH, M. 2022. Penerapan Metode Hybrid Arima-Ann Untuk Memprediksi Harga Saham Pt. Bni (Persero) Tbk. *RAGAM: Journal of Statistics & Its Application*, 1(1), 52. <https://doi.org/10.20527/ragam.v1i1.7328>
- FAISAL, A. 2021. PREDIKSI SAHAM TELKOM DENGAN METODE ARIMA. *Jurnal Bisnis, Logistik Dan Supply Chain (BLOGCHAIN)*, 1(2). <https://doi.org/10.55122/blogchain.v1i2.298>
- GIROH, H., KUMAR, V., & SINGH, G. (2023). Improving the Performance of Hybrid Models Using Machine Learning and Optimization Techniques. *International Journal of Membrane Science and Technology*, 10(2), 3396–3409. <https://doi.org/10.15379/ijmst.v10i2.3138>
- HASTOMO, W., KARNO, A. S. B., KALBUANA, N., NISFIANI, E., & ETP, L. 2021. Optimasi Deep Learning untuk Prediksi Saham di ... (*Jurnal Edukasi Dan ...*, 7(2), 133–140. <https://jurnal.untan.ac.id/index.php/jepin/article/view/47411>
- KHAIRA, U., UTOMO, P. E. P., SURATNO, T., & GULO, P. C. S. 2021. Prediksi Indeks Harga Saham Gabungan (IHSG) Menggunakan Algoritma Autoregressive Integrated Moving Average (ARIMA). *JUSS (Jurnal Sains Dan Sistem Informasi)*, 2(2), 11–17. <https://doi.org/10.22437/juss.v2i2.8449>
- LATIF, N., SELVAM, J. D., KAPSE, M., SHARMA, V., & MAHAJAN, V. 2023. Comparative Performance of LSTM and ARIMA for the Short-Term Prediction of Bitcoin Prices. *Australasian Accounting, Business and Finance Journal*, 17(1). <https://doi.org/10.14453/aabfj.v17i1.15>
- LIAO, Z., DAI, S., & KUOSMANEN, T. 2024. Convex support vector regression. *European Journal of Operational Research*, 313(3), 858–870. <https://doi.org/10.1016/j.ejor.2023.05.009>
- MAHENDRA, B. H., CHAERANI, L., & GUMAY, G. 2023. Analisis Perbandingan Prediksi Harga Saham menggunakan Algoritma Artificial Neural Network dan Linear Regression. *Jurnal Ilmiah Komputasi*, 22(2), 303–312. <https://doi.org/10.32409/jikstik.22.2.3357>
- MUHAMMAD ALI, P. J. 2022. Investigating the Impact of Min-Max Data Normalization on the Regression Performance of K-Nearest Neighbor with Different Similarity Measurements. *Aro-the Scientific Journal of Koya University*, 10(1), 85–91. <https://doi.org/10.14500/aro.10955>
- NASIONAL, J., INFORMASI, S., CHRISTINA, N., & LINDA, T. 2024. Komparasi Algoritma Naïve Bayes dan Gradient Boosting untuk Prediksi Pasien Diabetes. 02, 118–125.
- NATEKIN, A., & KNOLL, A. 2013. Gradient boosting machines, a tutorial. *Frontiers in Neuroinformatics*, 7(DEC). <https://doi.org/10.3389/fnbot.2013.00021>
- PIERRE, A. A., AKIM, S. A., SEMENYO, A. K., & BABIGA, B. (2023). Peak Electrical Energy Consumption Prediction by ARIMA, LSTM, GRU, ARIMA-LSTM and ARIMA-GRU Approaches. *Energies*, 16(12). <https://doi.org/10.3390/en16124739>
- PUTERI, D. I. 2023. Implementasi Long Short Term Memory (LSTM) dan Bidirectional Long Short Term Memory (BiLSTM) Dalam Prediksi Harga Saham Syariah. *Euler : Jurnal Ilmiah Matematika, Sains Dan Teknologi*, 11(1). <https://doi.org/10.34312/euler.v11i1.19791>
- S, M. A. 2023. Prediksi Terkena Diabetes menggunakan Metode K-Nearest Neighbor (KNN) pada Dataset UCI Machine Learning Diabetes. *Indonesian Journal of Applied Mathematics*, 3(2), 15. <https://doi.org/10.35472/indojam.v3i2.1577>
- SALUZA, I., SARTIKA, D., ASTUTI, L. W., FARADILLAH, F., DESITAMA, L., & PURNAMASARI, E. D. 2021. Prediksi Data Time Series Harga Penutupan Saham Menggunakan Model Box Jenkins ARIMA. *Jurnal Ilmiah Informatika Global*, 12(2). <https://doi.org/10.36982/jiig.v12i2.1940>
- SAPUTRO, D., & SWANJAYA, D. 2023. Analisa Prediksi Harga Saham Menggunakan Neural Network Dan Net Foreign Flow. *Generation Journal*, 7(2).

- <https://doi.org/10.29407/gj.v7i2.20001>  
SILMI ATH THAHIRAH AL AZHIMA, D. DARMAWAN, N. FAHMI ARIEF HAKIM, I. KUSTIAWAN, M. AL QIBTIYA, N. S. S. 2022. Hybrid Machine Learning Model Untuk Memprediksi Penyakit. *Jurnal Teknologi Terpadu*, 8(1), 40–46.
- SOEWIGNJO, S., SEDIONO, MARDIANTO, M. F. F., & PUSPORANI, E. 2023. Prediksi Harga Saham Bank BCA (BBCA) Pasca Stock Split dengan Artificial Neural Network dengan Algoritma Backpropagation. *G-Tech: Jurnal Teknologi Terapan*, 7(4), 1683–1693. <https://doi.org/10.33379/gtech.v7i4.3363>
- SUPRIYANTO, S., UTAMI, A. P., & ISTIKANAAH, N. 2023. Model Peramalan Harga Saham Menggunakan Metode Arima – Garch (Studi Kasus Saham Pt. Unilever Indonesia). *Jurnal Ilmiah Matematika Dan Pendidikan Matematika*, 15(1), 1. <https://doi.org/10.20884/1.jmp.2023.15.1.8658>
- WANG, G., LYU, Z., & LI, X. 2023. An Optimized Random Forest Regression Model for Li-Ion Battery Prognostics and Health Management. *Batteries*, 9(6). <https://doi.org/10.3390/batteries9060332>