



Comparison Random Forest Regression and Linear Regression For Forecasting BBKA Stock Price

Arif Mudi Priyatno¹, Lailatul Syifa Tanjung², Wahyu Febri Ramadhan³, Putri Cholidhazia⁴, Putri Zulia Jati⁵, Fahmi Iqbal Firmananda⁶

Fakultas Ekonomi dan Bisnis, Universitas Pahlawan Tuanku Tambusai Email: arifmudi@universitaspahlawan.ac.id⁽¹⁾

Fakultas Teknik, Universitas Pahlawan Tuanku Tambusai Email: lailashifa1205@gmail.com⁽²⁾

Fakultas Ekonomi dan Bisnis, Universitas Pahlawan Tuanku Tambusai Email: wahyu.febri.id@universitaspahlawan.ac.id⁽³⁾

Fakultas Sains dan Teknologi, Universitas Al-Azhar Indonesia Email: putricholidhazia@gmail.com⁽⁴⁾

Fakultas Ilmu Hayati, Universitas Pahlawan Tuanku Tambusai Email: putrizuliajati01@gmail.com⁽⁵⁾

Fakultas Ekonomi dan Bisnis, Universitas Pahlawan Tuanku Tambusai Email: fahmiiqbalfirmananda@universitaspahlawan.ac.id⁽⁶⁾

DOI: 10.31004/jutin.v6i3.16933

✉ Corresponding author:

[arifmudi@universitaspahlawan.ac.id]

Article Info

Abstrak

Kata kunci:
Saham
Pasar Saham
Machine learning
Random Forest regression
Linear regression
PT Bank Central Asia Tbk

Perdagangan saham adalah instrumen pasar keuangan yang populer di seluruh dunia. Di Indonesia, pasar saham adalah Bursa Efek Indonesia (BEI), dan salah satu saham aktif yang diperdagangkan adalah PT Bank Central Asia (BBCA). Namun, pergerakan harga saham sangat sulit diprediksi karena dipengaruhi oleh banyak faktor. Investor menggunakan analisis fundamental dan analisis teknikal untuk pengambilan keputusan, tetapi hasilnya cenderung bervariasi. Peramalan harga saham dapat dilakukan dengan menggunakan machine learning, khususnya dengan algoritma *machine learning random forest regression* dan *linear regrssion*. Penelitian ini membandingkan *machine learning random forest regression* dan *linear regrssion* tersebut untuk meramalkan harga saham BBCA dengan harapan memberikan solusi yang lebih akurat dan efektif bagi para investor dalam mengambil keputusan investasi dan perdagangan saham BBCA. Hasil evaluasi rata-rata *cross-validation means squared error* (MSE), *root means squared error* (RMSE), *means absolute error* (MAE), dan *means absolute percentage error* (MAPE) pada *linear regression* yaitu 0.12848, 0.35807, 0.29570, dan 0.0036%; sedangkan *random forest regression* yaitu 27473.76, 158.04, 142.70, 1.7153%. Hal ini menunjukkan bahwa *linear regression* memberikan performa yang lebih baik dalam melakukan peramalan.

Keywords:

*Stock
Stock Market
Machine learning
Random Forest regression
Linear regression
PT Bank Central Asia Tbk*

Abstract

Stock trading is a popular financial instrument worldwide. In Indonesia, the stock market is known as the Indonesia Stock Exchange (BEI), and one actively traded stock is PT Bank Central Asia (BBCA). However, predicting stock price movements is challenging due to various influencing factors. Investors use fundamental and technical analyses for decision-making, but results often vary. Machine learning, particularly random forest regression and linear regression algorithms, can be used for stock price forecasting. In this paper, we compares these two machine learning methods to forecast BBCA stock prices, aiming to provide more accurate and effective solutions for investor's investment and trading decisions. The evaluation results of cross-validation mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) for linear regression were 0.12848, 0.35807, 0.29570, and 0.0036%, respectively, while for random forest regression were 27473.76, 158.04, 142.70, and 1.7153%. These findings indicate that linear regression outperforms in forecasting performance.

1. INTRODUCTION

Perdagangan saham merupakan salah satu instrumen pasar keuangan yang sangat populer di seluruh dunia (Rustam & Kintandani, 2019). Saham adalah surat tanda bukti kepemilikan dalam sebuah perusahaan dan perdagangan saham dilakukan di pasar saham (Azizah et al., 2020; K. S et al., 2021; S. S et al., 2022). Di Indonesia, pasar saham utama adalah Bursa Efek Indonesia (BEI) (Putri & Dhini, 2019). Salah satu saham yang aktif diperdagangkan di BEI adalah PT Bank Central Asia (BBCA), sebuah bank terkemuka di Indonesia. Saham BBCA termasuk dalam Indeks LQ45 yang mencerminkan performa 45 saham terlikuid di pasar (IDX, 2023a).

Sejak perusahaan ini melakukan penawaran saham perdana (initial public offering) pada 31 Mei 2000, saham BBCA telah menarik minat para investor sebagai instrumen perdagangan dan investasi (IDX, 2023b). Karena reputasinya yang kuat dan kinerja yang konsisten, saham BBCA sering menjadi pilihan bagi investor yang mencari peluang dalam pasar saham Indonesia.

Namun, pergerakan harga saham sangat sulit diprediksi (Sakhare & Sagar Imambi, 2019; Zhang et al., 2022). Harga saham dipengaruhi oleh berbagai faktor, termasuk aktivitas jual beli oleh para investor (Bourezk et al., 2019; Gao et al., 2021). Saat ini, investor menggunakan dua pendekatan umum untuk pengambilan keputusan investasi atau perdagangan saham, yaitu analisis fundamental dan analisis teknikal (Ghosh et al., 2019; Harahap et al., 2020; Nti et al., 2020).

Analisis fundamental melibatkan penilaian terhadap kinerja dan prospek fundamental perusahaan, termasuk analisis terhadap laporan keuangan, pertumbuhan pendapatan, dan kondisi industry (Nourbakhsh & Habibi, 2022). Di sisi lain, analisis teknikal melibatkan penggunaan grafik harga dan indikator teknikal untuk mengidentifikasi pola dan tren dalam pergerakan harga saham (Heinz et al., 2021). Namun, hasil analisis fundamental dan analisis teknikal cenderung bervariasi antara investor, karena mereka mendasarkan keputusan mereka pada persepsi dan penilaian yang berbeda. Oleh karena itu, meramal pergerakan harga saham secara akurat merupakan tantangan menarik yang perlu dipelajari lebih lanjut.

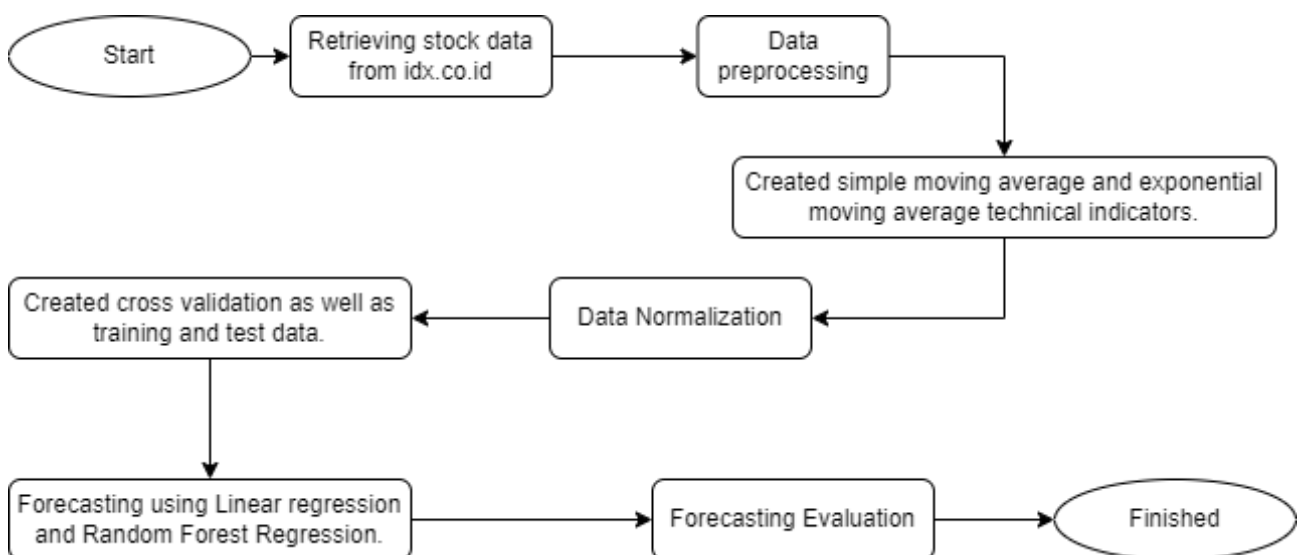
Peramalan harga saham bisa dilakukan menggunakan kecerdasan buatan khususnya machine learning (Saini & Sharma, 2022). Peramalan harga saham menggunakan machine learning dapat memanfaatkan data harian transaksi seperti *close*, *high*, *low*, *open*, dan *volume*, serta data teknikal

indicator seperti *moving average*, *rsi*, dan lain sebagainya. Dalam penelitian yang dilakukan oleh Kumar & Chaudhry (2021) mengenai review dan analisis prediksi harga saham menggunakan data mining, hasilnya menunjukkan bahwa penggunaan regresi linear berhasil mencapai tingkat akurasi di atas 80 persen. Sementara itu, penelitian yang dilakukan oleh Pujitha & Mahesh (2021) membahas tentang prediksi harga saham menggunakan bahasa pemrograman Python, dan menyatakan bahwa penggunaan regresi linear berhasil mencapai tingkat akurasi sebesar 87 persen. Dalam penelitian yang dilakukan oleh Raya et al (2022) menyatakan bahwa random forest merupakan algoritma terbaik berdasarkan pengukuran MAE (*mean absolute error*), algoritma random forest dibandingkan dengan algoritma *support vector regression* (SVR), ARIMA, dan XGBoost. Sementara itu, penelitian yang dilakukan Basak et al (2019) tentang prediksi pergerakan harga saham menunjukkan bahwa random forest berhasil mendapatkan akurasi tertinggi 94,76 persen. Dengan demikian, perlu dilakukan penelitian lebih lanjut untuk membandingkan kinerja antara random forest regression dan regresi linear dalam konteks peramalan harga saham BBCA, dengan tujuan untuk menentukan metode yang paling efektif dalam melakukan peramalan harga saham tersebut.

Pada Penelitian ini, kami memfokuskan pada perbandingan machine learning random forest regression dan linear regression untuk peramalan harga saham BBCA. Data menggunakan data historical dan teknikal indicator simple moving average serta exponential moving average. Metode machine learning digunakna yaitu random forest regression dan linear regression. Dengan membandingkan kedua metode tersebut, kami berharap dapat memberikan solusi yang lebih akurat dan efektif dalam meramalkan pergerakan harga saham BBCA. Dengan adanya solusi yang lebih andal, para investor dapat mengambil keputusan yang lebih baik dalam investasi dan perdagangan saham BBCA.

2. METHODS

Proses penelitian dilakukan yaitu pengambilan data saham, pre-procesing data, pembuatan teknikal indikator *simple moving average* dan *exponential moving average*, penggabungan data harian saham dan data teknikal indikator, pembuatan *cross-validation* berserta dengan data latih dan uji, peramalan harga saham menggunakan *linear regression* dan *random forest regression*, dan evaluasi peramalan. Gambar 1 merupakan block alur proses penelitian ini.



Gambar 1 Block Proses Penelitian

2.1. Pengambilan Data

Data dikumpulkan dari web bursa efek indonesia (idx.co.id). Data saham BBCA diambil dari tanggal 21 April 2015 hingga 24 maret 2023. Field berhasil diambil yaitu *date*, *previous*, *open_price*, *first_trade*, *high*, *low*, *close*, *change*, *volume*, *value*, *frequency*, *index_individual*, *offer*, *offer_volume*, *bid*, *bid_volume*, *listed_shares*, *tradeable_shares*, *weight_for_index*, *foreign_sell*, *foreign_buy*, *delisting_date*, *non_regular_volume*, *non_regular_value*, dan *non_regular_frequency*. Jumlah data didapatkan yaitu 2006 row data.

2.2. Pre-processing data

Pre-processing merupakan langkah yang dilakukan guna mempersiapkan data agar dapat digunakan pada tahap selanjutnya. Dalam konteks saham, terdapat aksi korporasi yang umum dilakukan, yaitu stock split. Oleh sebab itu, penyesuaian terhadap saham harus dilakukan sesuai dengan perbandingan yang berlaku pada stock split tersebut.

Stock split merupakan pembagian saham yang dilakukan dengan memperbanyak jumlah saham yang ada. Misalnya, jika terjadi stock split 2:1, maka setiap saham akan dibagi menjadi dua saham baru dengan harga yang setengah dari harga semula. Proses pre-processing diperlukan untuk menghitung perbandingan stock split dan melakukan penyesuaian harga saham sesuai dengan perbandingan yang telah ditentukan.

2.3. Pembentukan Teknikal Indikator

Teknikal indikator digunakan pada penelitian ini yaitu *simple moving average* dan *exponenetial moving average*. *Simple Moving Average* (SMA) adalah indikator yang digunakan oleh para ahli ekonomi dan trader untuk mengidentifikasi tren pasar dengan menghitung rata-rata harga dalam periode waktu tertentu. SMA membantu menghaluskan fluktuasi harga harian dan memberikan gambaran tentang arah tren pasar dalam jangka waktu yang ditentukan. Persamaan 1 merupakan persamaan untuk membuat *simple moving average*.

$$SMA = (Harga1 + Harga2 + Harga3 + \dots + HargaN) / N \quad (1)$$

Dalam Persamaan 1, "Harga1" hingga "HargaN" mewakili harga-harga penutupan aset keuangan dalam suatu periode waktu tertentu. "N" merupakan jumlah periode waktu yang digunakan untuk menghitung SMA. Untuk menghitung SMA, Anda perlu menjumlahkan harga-harga penutupan dalam periode waktu yang ditentukan dan kemudian membaginya dengan jumlah periode tersebut. Misalnya, jika Anda ingin menghitung SMA selama 5 hari dengan harga penutupan pada hari pertama hingga hari kelima, Anda akan menjumlahkan harga-harga penutupan tersebut dan membaginya dengan 5. Perhatikan bahwa SMA menggunakan metode perhitungan rata-rata sederhana, di mana setiap harga memiliki bobot yang sama dalam perhitungan.

Exponential Moving Average (EMA) adalah indikator yang menggunakan perhitungan rata-rata bergerak dengan memberikan bobot yang lebih tinggi pada harga-harga terbaru. EMA memberikan sinyal perubahan tren lebih cepat daripada SMA dan lebih responsif terhadap pergerakan harga terkini. Dengan memberikan penekanan yang lebih besar pada pergerakan harga terbaru, EMA membantu trader dalam mengidentifikasi perubahan tren dan mengambil keputusan investasi yang tepat. Persamaan 2 merupakan persamaan untuk menghitung *exponential moving average*.

$$\text{EMA}(t) = (\text{Harga}(t) - \text{EMA}(t-1)) * \text{Multiplier} + \text{EMA}(t-1) \quad (2)$$

Dalam persamaan di atas, "EMA(t)" adalah nilai EMA pada periode waktu "t". "Harga(t)" adalah harga terkini pada periode waktu "t". "EMA(t-1)" adalah nilai EMA pada periode sebelumnya. "Multiplier" adalah faktor pengali yang digunakan untuk memberikan bobot pada harga terkini. Nilai EMA pada periode pertama (EMA(1)) biasanya diambil sebagai harga terkini, sehingga EMA(1) = Harga(1). Untuk menghitung EMA pada periode waktu selanjutnya, Anda perlu menggunakan nilai EMA pada periode sebelumnya (EMA(t-1)) dan harga terkini (Harga(t)). Multiplier bergantung pada jumlah periode waktu yang digunakan dan rumus yang dipilih. Persamaan 3 merupakan cara untuk menghitung Multiplier secara umum.

$$\text{Multiplier} = 2 / (N + 1) \quad (3)$$

Dalam rumus di atas, "N" adalah jumlah periode waktu yang digunakan. Multiplier tersebut memberikan bobot yang lebih tinggi pada harga terkini dan mengurangi bobot secara eksponensial saat melihat harga-harga sebelumnya.

2.4. Normalisasi Data

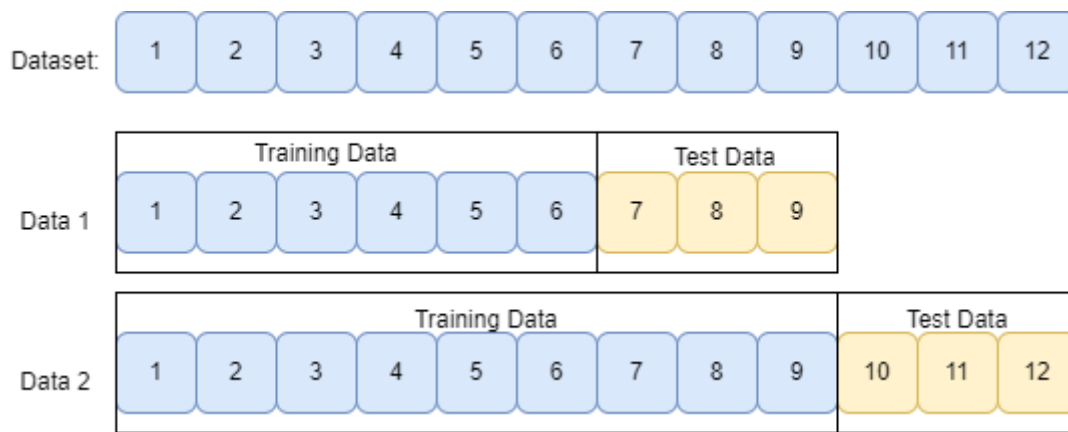
Data histori saham dan teknikal indikator perlu dilakukan normalisasi. Teknik normalisasi data digunakan pada penelitian ini adalah Teknik normalisasi min-max. Data normalisasi Min-Max adalah salah satu teknik normalisasi yang digunakan untuk mengubah nilai-nilai dalam dataset menjadi rentang nilai yang spesifik, yaitu dalam kisaran tertentu seperti [0, 1]. Teknik normalisasi ini mengubah setiap nilai dalam dataset ke dalam rentang tersebut. Persamaan 4 merupakan formula untuk melakukan perhitungannya (Kanwal et al., 2022).

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4)$$

Dimana X_{norm} merupakan nilai yang telah dinormalisasi dari nilai asli (X), X_{min} merupakan nilai terkecil dalam dataset, dan X_{max} merupakan nilai terbesar dalam dataset. Dengan menggunakan normalisasi Min-Max, semua nilai dalam dataset akan diubah ke dalam rentang [0, 1]. Nilai terkecil dalam dataset akan diubah menjadi 0, dan nilai terbesar dalam dataset akan diubah menjadi 1. Nilai-nilai di antara nilai terkecil dan terbesar akan dipersempit menjadi kisaran antara 0 dan 1 sesuai dengan proporsinya dalam dataset.

2.5. Cross-validation

Cross-validation pada data time series adalah metode yang digunakan untuk mengevaluasi kinerja model prediksi pada data time series. Perbedaannya dengan cross-validation pada data non-time series adalah perlunya mempertimbangkan urutan waktu dalam pemisahan dan evaluasi data. Dalam cross-validation data time series, urutan waktu menjadi faktor penting yang harus diperhatikan. Hal ini karena dalam data time series, observasi pada waktu yang berdekatan cenderung saling terkait dan mungkin memiliki pola atau tren tertentu. Oleh karena itu, langkah-langkah dalam cross-validation data time series sedikit berbeda dari cross-validation pada data non-time series. Penelitian ini menggunakan *cross-validation time series* untuk data latih dan data ujinya. Gambar 2 merupakan ilustrasi dari *cross-validation time series*.



Gambar 2 Konsep Cross-validation Data Time Series

2.6. Peramalan Harga saham

Peramalan harga saham pada penelitian ini yaitu membandingkan metode *machine learning random forest regression* dan *linear regression*. *Random Forest Regression* adalah salah satu algoritma machine learning yang digunakan untuk pemodelan regresi. Algoritma ini berdasarkan konsep "Random Forest" yang merupakan ensambel dari banyak decision tree (Pedregosa F. et al., 2011). Tahapan Algoritma Random Forest regression (Breiman, 2001; Liaw & Wiener, 2002).

1. Inputan

- Dataset yang berisi data training dengan fitur-fitur (independent variables) dan nilai target (dependent variable) yang ingin diprediksi.
- Jumlah pohon ($n_{estimators}$) yang akan digunakan dalam ensemble model.
- Jumlah fitur yang dipertimbangkan saat mencari pemisah terbaik untuk setiap pohon ($max_features$).
- Kriteria untuk mengukur kualitas pemisah (criterion), misalnya "mse" (Mean Squared Error) untuk regresi.

2. Pengaturan Parameter

- Tentukan jumlah pohon ($n_{estimators}$) dalam ensemble model. Biasanya, semakin banyak pohon akan meningkatkan akurasi, tetapi juga akan memperpanjang waktu pelatihan.
- Tentukan jumlah fitur yang akan dipertimbangkan saat mencari pemisah terbaik ($max_features$). Jumlah fitur ini dapat berpengaruh pada variasi dan keakuratan model.
- Pilih kriteria untuk mengukur kualitas pemisah (criterion), seperti "mse" untuk regresi. Criterion ini menentukan bagaimana pemilihan fitur terbaik dalam memisahkan data.

3. Pembuatan Model

Untuk setiap pohon dalam ensemble, lakukan langkah-langkah berikut:

- Acak (bootstrap) sampel dari dataset pelatihan dengan penggantian. Hal ini akan membuat setiap pohon memiliki data latih yang unik.
- Pilih fitur secara acak sebanyak yang ditentukan oleh parameter $max_features$ untuk membangun pemisah terbaik dalam setiap node pohon.
- Bangun pohon dengan mengulangi proses pembagian data hingga mencapai titik berhenti (misalnya, jika node mencapai jumlah data minimum atau node tidak dapat dibagi lagi).

4. Peramalan

Untuk setiap data uji, lakukan langkah-langkah berikut:

- Kirimkan data uji melalui setiap pohon dalam ensemble.
- Hitung prediksi dari setiap pohon berdasarkan pemisah yang telah dibangun.
- Agregasikan prediksi dari semua pohon (misalnya, hitung rata-rata prediksi) untuk mendapatkan prediksi akhir model.

5. Output

Output dari algoritma Random Forest Regression adalah prediksi nilai target (dependent variable) berdasarkan fitur-fitur (independent variables) yang diberikan.

Machine learning Linear regression adalah salah satu teknik dalam statistika dan *machine learning* yang digunakan untuk memodelkan hubungan linier antara sebuah variabel independen (variabel prediktor) dengan sebuah variabel dependen (variabel target). Tujuan utama linear regression adalah untuk menemukan garis regresi terbaik yang bisa mewakili hubungan linier antara variabel prediktor dan variabel target. Langkah-langkah algoritma machine learning linear regression (Hastie et al., 2009; James et al., 2021).

1. Inputan

- Dataset yang berisi data training dengan fitur-fitur (independent variables) dan nilai target (dependent variable) yang ingin diprediksi.
- Data testing (opsional) untuk melakukan prediksi pada data yang belum pernah dilihat sebelumnya.

2. Pengaturan Parameter

Tidak banyak parameter yang perlu diatur dalam linear regression. Algoritma ini secara otomatis akan mencari garis (atau hiperbidang dalam kasus regresi multivariat) terbaik yang menggambarkan hubungan linier antara fitur-fitur dan target.

3. Pembuatan Model

Menentukan Variabel:

- Tentukan variabel X sebagai fitur (independent variable) dan variabel y sebagai nilai target (dependent variable).
- Variabel X harus berupa matriks dengan ukuran $(n \times m)$, di mana n adalah jumlah data dan m adalah jumlah fitur.
- Variabel y harus berupa vektor dengan ukuran $(n \times 1)$.

Menghitung Parameter Model:

- Hitung matriks transpose dari X (X^T).
- Hitung matriks perkalian antara X^T dan X ($X^T * X$).
- Hitung matriks invers dari $(X^T * X)^{-1}$.
- Hitung matriks perkalian antara invers $(X^T * X)^{-1}$ dan X^T ($X^T * X)^{-1} * X^T$.
- Hitung vektor bobot (koefisien) w dengan mengalikan $(X^T * X)^{-1} * X^T$ dengan variabel y ($w = (X^T * X)^{-1} * X^T * y$).

4. Peramalan

Setelah mendapatkan vektor bobot (w), kita dapat melakukan prediksi pada data testing (jika ada) atau pada data yang belum pernah dilihat sebelumnya dengan mengalikan matriks dari fitur (X) dengan vektor bobot (w).

5. Output

Output dari algoritma Linear Regression adalah prediksi nilai target (dependent variable) berdasarkan fitur-fitur (independent variables) yang diberikan.

2.7. Evaluasi Peramalan

Evaluasi peramalan harga saham pada penelitian ini menggunakan *mean square error* (MSE), *root mean squared error* (RMSE), *mean absolute error* (MAE), dan *Mean absolute percentage error* (MAPE). MSE mengukur rata-rata dari kuadrat selisih antara setiap nilai prediksi dengan nilai sebenarnya (Riko Anshori Prasetya & Mudi Priyatno, 2023). Persamaan 5 merupakan rumus untuk menghitung MSE (Yun et al., 2023). Semakin kecil nilai MSE, semakin baik kinerja model prediksi. MSE memiliki arti kuadrat rata-rata dari error atau kesalahan prediksi, sehingga nilai-nilai yang jauh dari nilai sebenarnya memiliki kontribusi yang lebih besar pada nilai MSE. RMSE adalah bentuk perluasan dari Mean Square Error (MSE) yang telah diakar kuadratkan (Priyatno et al., 2019). Persamaan 6 merupakan cara untuk menghitung RMSE (Sahu et al., 2023).

MAE mengukur rata-rata dari nilai absolut selisih antara setiap nilai prediksi dengan nilai sebenarnya (Olorunnimbe & Viktor, 2023). Persamaan 7 merupakan cara untuk menghitung MAE (Ali et al., 2023). Pada perhitungan MAE, selisih antara nilai sebenarnya dan nilai prediksi diambil nilai absolutnya, sehingga tidak ada perbedaan arah (positif atau negatif) dalam perhitungannya. MAE memberikan ukuran rata-rata dari deviasi absolut atau error prediksi, yang menjadikannya metrik evaluasi yang lebih toleran terhadap outlier dibandingkan dengan RMSE (Root Mean Squared Error) yang mengkuadratkan nilai selisih. MAPE mengukur rata-rata persentase kesalahan prediksi dibandingkan dengan nilai sebenarnya (Li et al., 2023). Persamaan 8 merupakan cara untuk menghitung MAPE (Chen, 2023). Sebagai panduan umum dalam interpretasi nilai MAPE yaitu Jika $MAPE < 10\%$, maka prediksi dianggap sangat baik. Jika $10\% \leq MAPE < 20\%$ maka prediksi dianggap baik. Jika $MAPE \geq 20\%$ maka prediksi dianggap memiliki tingkat kesalahan yang signifikan dan perlu dipertimbangkan untuk diperbaiki.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \tag{6}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \tag{7}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100 \tag{8}$$

dimana n merupakan jumlah total dari keseluruhan data. Y_i merupakan nilai sebenarnya dari data ke- i , dan \hat{Y}_i merupakan nilai hasil dari prediksi data ke- i .

3. RESULT AND DISCUSSION

Data histori PT Bank Central Asia Tbk (BBCA) diambil langsung dari website bursa efek indonesia (idx.co.id). Jumlah data didapatkan adalah 2006 row data. Data histori diambil dari tanggal 21 April 2015 hingga 24 maret 2023. Gambar 3 merupakan visualisasi harga penutupan saham BBCA yang diambil. Saham BBCA melakukan aksi korporasi *stock split* dengan perbandingan 1:5 pada tanggal 13 oktober 2021.



Gambar 3 Visualisasi harga penutupan saham BBCA (a) harga asli dari idx.co.id (b) harga setelah dilakukan pre-processing

Data hasil preprocessing dilakukan pembentukan teknika indikator berdasarkan harga penutupan. Pembentukan teknikal indikator *simple moving average* dan *exponential moving average*

menggunakan periode 3 hari hingga 75 hari. Pembentukan *simple moving average* menggunakan Persamaan 1, serta pembentukan *exponential moving average* menggunakan Persamaan 2. Total kolom hasil teknikal indikator yaitu 144 kolom. Tabel X merupakan sample hasil dari pembentukan teknikal indikator. Sample pada Tabel 2 ditunjukkan bahwa SMA_3, SMA_5, SMA_5, EMA_3, EMA_5, dan EMA_7 pada tanggal 21 april 2015 yaitu 2975, 2975, 2988, 2979, 2981, dan 2983.

Tabel 1 Data Teknikal Indikator SMA dan EMA

Tanggal	SMA_3	SMA_5	SMA_7	EMA_3	EMA_5	EMA_7
4/21/2015	2975	2975	2988	2979	2981	2983
4/22/2015	2969	2971	2981	2978	2980	2981
4/23/2015	2981	2979	2978	2987	2985	2985
4/24/2015	2994	2979	2978	2992	2990	2989
4/27/2015	2956	2962	2966	2935	2950	2958

Setelah pembentukan teknikal indikator, data histori dan teknikal indikator saham dilakukan proses normalisasi data. Proses normalisasi menggunakan persamaan 4. Tabel 2 merupakan hasil dari data normalisasi. Sample data hasil normalisasi dengan kolom close, open, high, low, volume, SMA_3, SMA_5, SMA_5, EMA_3, EMA_5, dan EMA_7 berdasarkan Tabel 2 pada Tanggal 21 april 2015 yaitu 0.1032, 0.0964, 0.0921, 0.1102, 0.0943, 0.0945, 0.0881, 0.0898, 0.0917, 0.0904, dan 0.0898.

Tabel 2 Data Hasil Normalisasi

Tanggal	Close	Open	High	Low	Volume	SMA_3	SMA_5	SMA_7	EMA_3	EMA_5	EMA_7
4/21/2015	0.103	0.096	0.092	0.110	0.0943	0.094	0.088	0.089	0.091	0.090	0.089
5	2	4	1	2		5	1	8	7	4	8
4/22/2015	0.099	0.103	0.092	0.113	0.0740	0.093	0.087	0.088	0.091	0.090	0.089
5	6	5	1	9		6	5	7	6	2	5
4/23/2015	0.103	0.103	0.092	0.113	0.0415	0.095	0.088	0.088	0.092	0.091	0.090
5	2	5	1	9		4	7	3	9	0	1
4/24/2015	0.103	0.103	0.092	0.113	0.0468	0.097	0.088	0.088	0.093	0.091	0.090
5	2	5	1	9		3	7	3	7	7	7
4/27/2015	0.081	0.1	0.088	0.091	0.2006	0.091	0.086	0.086	0.085	0.085	0.086
5	8		6	9		7	2	5	2	7	1

Setelah proses normalisasi, data dilakukan pembagian data latih dan data uji menggunakan konsep *cross-validation*. *Cross-validation* (CV) dibuat sebanyak 5 CV. Tabel 3 merupakan pembagian data latih dan uji berdasarkan *cross-validation*. Jumlah data uji digunakan adalah 30 row data. Jumlah data latih berdasarkan *cross-validation* dari 1 hingga 5 secara berurutan yaitu 1780, 1810, 1840, 1870, dan 1900 data.

Tabel 3 Cross-validation data latih dan uji

Cross-validation	Jumlah Data Latih	Jumlah Data Uji
1	1780	30
2	1810	30
3	1840	30

4	1870	30
5	1900	30

Data hasil *cross-validation* selanjutnya dilakukan peramalan pergerakan harga saham menggunakan *machine learning linear regression* dan *random forest regression*. Peramalan pergerakan harga saham dilakukan evaluasi menggunakan metrik evaluasi *means squared error* (MSE), *root means squared error* (RMSE), *means absolute error* (MAE), dan *means absolute percentage error* (MAPE). Tabel 4 merupakan metrik hasil evaluasi pengujian menggunakan *machine learning linear regression*. Tabel 5 merupakan metrik hasil evaluasi *machine learning random forest regression*.

Tabel 4 Evaluasi Pengujian Machine Learning Linear regression

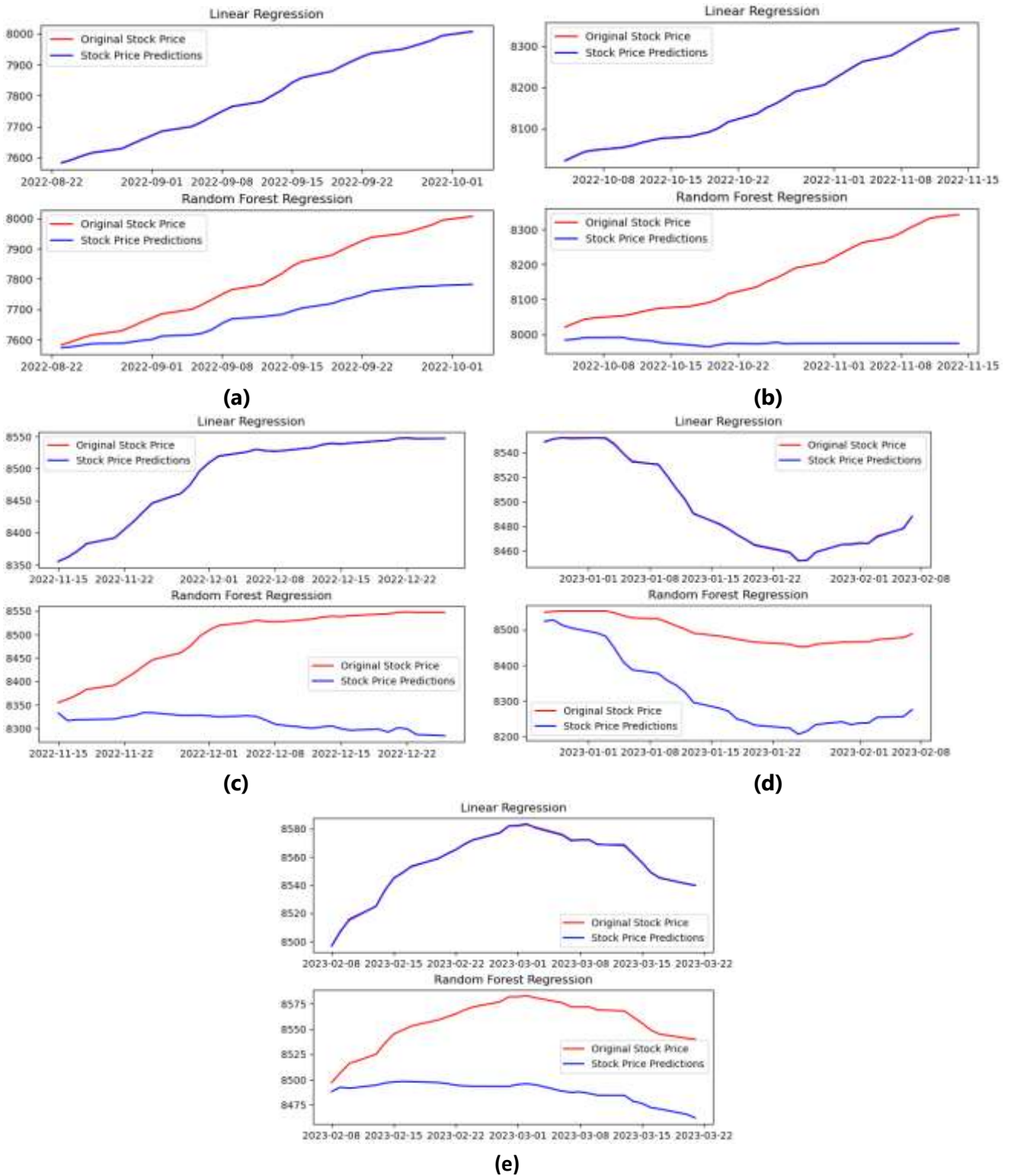
Metrik	CV1	CV2	CV3	CV4	CV5	Means
MSE	0.12514	0.14330	0.12619	0.13820	0.10959	0.12848
RMSE	0.35375	0.37855	0.35524	0.37175	0.33105	0.35807
MAE	0.28856	0.27857	0.31719	0.31319	0.28096	0.29570
MAPE	0.0037%	0.0034%	0.0037%	0.0037%	0.0033%	0.0036%

Tabel 4 menunjukkan bahwa nilai *means squared error* (MSE) terkecil didapatkan pada *cross-validation* ke-5 dengan nilai 0.10959, serta rata-rata *mean squared error* didapatkan 0.12848. Hasil nilai terkecil dari *root means squared error* (RMSE) pada *cross-validation* ke-5 yaitu 0.33105, serta rata-rata *cross-validation root means squared error* yaitu 0.35807. Hasil nilai terkecil *means absolute error* (MAE) pada *cross-validation* ke-5 yaitu 0.28096, serta nilai rata-rata *cross-validation mean absolute error* yaitu 0.29570. Hasil nilai terkecil dari *means absolute percentage error* didapatkan 0.0033% pada *cross-validation* ke-5, serta nilai rata-rata *cross-validation means absolute percentage error* yaitu 0.0036%.

Tabel 5 Evaluasi Pengujian Machine Learning Random Forest Regression

Metrik	CV1	CV2	CV3	CV4	CV5	Means
MSE	18073.10	44338.83	35854.25	33994.63	5108.01	27473.76
RMSE	134.43	210.56	189.35	184.37	71.47	158.04
MAE	118.80	183.24	173.86	169.88	67.71	142.70
MAPE	1.5107%	2.2305%	2.0419%	2.0024%	0.7908%	1.7153%

Table 5 menunjukkan hasil evaluasi peramalan menggunakan *random forest regression*. Hasil terkecil dari *means squared error* yaitu 5108.01 pada *cross-validation* ke-5, serta rata-rata hasil *cross-validation means squared error* yaitu 27473.76. Hasil terkecil *root means squared error* yaitu 71.47 pada *cross-validation* ke-5, serta nilai rata-rata *cross-validation root squared error* yaitu 158.04. Hasil terkecil *means absolute error* yaitu 67.71 pada *cross-validation* ke-5, serta nilai rata-rata *cross-validation means absolute error* yaitu 142.70. Hasil terkecil *means absolute percentage error* yaitu 0.7908% pada *cross-validation* ke-5, serta nilai rata-rata *cross-validation means absolute percentage error* yaitu 1.7153%.



Gambar 4 Perbandingan peramalan harga saham menggunakan *linear regression* dan *random forest regression* (a) *cross-validation 1*, (b) *cross-validation 2*, (c) *cross-validation 3*, (d) *cross-validation 4*, (e) *cross-validation 5*

Gambar 4 merupakan visualisasi hasil peramalan harga saham menggunakan *linear regression* dan *random forest regression* mulai dari *cross-validation* ke-1 hingga ke 5. Gambar menunjukkan bahwa metode *machine learning linear regression* berhasil meramal dengan hasil yang memuaskan dibandingkan *random forest regression*. Metrik evaluasi pada Tabel 4 dan table 5 diperkuat dengan Gambar 4 menunjukkan tingkat keberhasilan dari *linear regression*. Rata-rata metrik *Means absolute percentage error linear regression* yaitu 0.0036%, hal ini menunjukkan bahwa metode *linear regression* berhasil meramal pergerakan harga saham.

Berdasarkan hasil evaluasi pengujian, *machine learning Linear Regression* secara konsisten memberikan performa yang lebih baik dalam mengatasi masalah peramalan dibandingkan dengan *machine learning Random Forest Regression*. Meskipun *machine learning Random Forest Regression* juga menghasilkan prediksi yang baik, *machine learning Linear Regression* memiliki nilai MSE, RMSE, MAE, dan MAPE yang lebih rendah, menunjukkan bahwa *machine learning linear regression* ini memiliki tingkat kesalahan prediksi yang lebih kecil. Oleh karena itu, dalam konteks perbandingan *machine learning linear regression* dan *random forest regression* ini, *machine learning Linear Regression* lebih direkomendasikan karena memberikan performa yang lebih baik dalam melakukan peramalan.

4. CONCLUSION

Penelitian ini membandingkan machine learning linear regression dan random forest regression untuk peramalan harga saham. Data saham digunakan pada penelitian ini yaitu PT Bank Central Asia Tbk (BBCA). Jumlah data digunakan yaitu 2006 row data yang diambil dari tanggal 21 April 2015 hingga 24 maret 2023. Hasil evaluasi rata-rata *cross-validation means squared error (MSE)*, *root means squared error (RMSE)*, *means absolute error (MAE)*, dan *means absolute percentage error (MAPE)* pada *machine learning linear regression* yaitu 0.12848, 0.35807, 0.29570, dan 0.0036%; sedangkan *machine learning random forest regression* yaitu 27473.76, 158.04, 142.70, 1.7153%. Hal ini menunjukkan bahwa *machine learning linear regression* memberikan performa yang lebih baik dalam melakukan peramalan.

5. REFERENCES

- Ali, M., Khan, D. M., Alshanbari, H. M., & El-Bagoury, A. A.-A. H. (2023). Prediction of Complex Stock Market Data Using an Improved Hybrid EMD-LSTM Model. *Applied Sciences*, 13(3), 1429. <https://doi.org/10.3390/app13031429>
- Azizah, M., Irawan, M. I., & Putri, E. R. M. (2020). Comparison of stock price prediction using geometric Brownian motion and multilayer perceptron. *AIP Conference Proceedings*, 2242(May), 030016. <https://doi.org/10.1063/5.0008066>
- Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, 47, 552–567. <https://doi.org/10.1016/j.najef.2018.06.013>
- Bourezk, H., Raji, A., Acha, N., & Barka, H. (2019). An Overview on Sentiment Mining for Stock Market prediction. *2019 1st International Conference on Smart Systems and Data Science (ICSSD)*, 1–8. <https://doi.org/10.1109/ICSSD47982.2019.9002981>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324/METRICS>
- Chen, J. (2023). Analysis of Bitcoin Price Prediction Using Machine Learning. *Journal of Risk and Financial Management*, 16(1), 51. <https://doi.org/10.3390/jrfm16010051>
- Gao, Y., Wang, R., & Zhou, E. (2021). Stock Prediction Based on Optimized LSTM and GRU Models. *Scientific Programming*, 2021, 1–8. <https://doi.org/10.1155/2021/4055281>
- Ghosh, A., Bose, S., Maji, G., Debnath, N., & Sen, S. (2019). Stock Price Prediction Using LSTM on Indian Share Market. *EPiC Series in Computing*, 63, 101–190. <https://doi.org/10.29007/qgcz>
- Harahap, L. A., Lipikorn, R., & Kitamoto, A. (2020). Nikkei Stock Market Price Index Prediction Using Machine Learning. *Journal of Physics: Conference Series*, 1566(1), 012043. <https://doi.org/10.1088/1742-6596/1566/1/012043>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. <https://doi.org/10.1007/978-0-387-84858-7>
- Heinz, A., Jamalooden, M., Saxena, A., & Pollacia, L. (2021). Bullish and Bearish Engulfing Japanese Candlestick patterns: A statistical analysis on the S&P 500 index. *The Quarterly Review of Economics and Finance*, 79, 221–244. <https://doi.org/10.1016/j.qref.2020.06.006>
- IDX. (2023a). *Indeks Saham*. Idx.Co.Id. <https://www.idx.co.id/id/produk/indeks>
- IDX. (2023b). *Profil Perusahaan Tercatat*. Idx.Co.Id. <https://www.idx.co.id/id/perusahaan-tercatat/profil->

perusahaan-tercatat

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning* (Vol. 103). Springer US. <https://doi.org/10.1007/978-1-0716-1418-1>
- Kanwal, A., Lau, M. F., Ng, S. P. H., Sim, K. Y., & Chandrasekaran, S. (2022). BiCuDNNLSTM-1dCNN — A hybrid deep learning-based predictive model for stock price prediction. *Expert Systems with Applications*, 202, 117123. <https://doi.org/10.1016/j.eswa.2022.117123>
- Kumar, A., & Chaudhry, M. (2021). Review and Analysis of Stock Market Data Prediction Using Data Mining Techniques. *2021 5th International Conference on Information Systems and Computer Networks (ISCON)*, 1–10. <https://doi.org/10.1109/ISCON52037.2021.9702498>
- Li, D., Li, Y., Wang, C., Chen, M., & Wu, Q. (2023). Forecasting carbon prices based on real-time decomposition and causal temporal convolutional networks. *Applied Energy*, 331, 120452. <https://doi.org/10.1016/j.apenergy.2022.120452>
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18–22. <http://www.stat.berkeley.edu/>
- Nourbakhsh, Z., & Habibi, N. (2022). Combining LSTM and CNN methods and fundamental analysis for stock price trend prediction. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-022-13963-0>
- Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). Efficient Stock-Market Prediction Using Ensemble Support Vector Machine. *Open Computer Science*, 10(1), 153–163. <https://doi.org/10.1515/comp-2020-0199>
- Olorunnimbe, K., & Viktor, H. (2023). Deep learning in the stock market—a systematic survey of practice, backtesting, and applications. *Artificial Intelligence Review*, 56(3), 2057–2109. <https://doi.org/10.1007/s10462-022-10226-0>
- Pedregosa F., Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, & Édouard Duchesnay. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <http://jmlr.org/papers/v12/pedregosa11a.html>
- Priyatno, A. M., Wiratmo, A., Syuhada, F., & Cholidhazia, P. (2019). Perbandingan Imputasi Dan Parameter Support Vector Regression untuk Peramalan Cuaca. *Jurnal SIMETRIS*, 10(2), 651–660. <https://doi.org/10.24176/simet.v10i2.3402>
- Pujitha, J., & Mahesh, G. (2021). Stock Market Price Predictions Using Python. *2021 International Conference on Forensics, Analytics, Big Data, Security (FABS)*, 1–5. <https://doi.org/10.1109/FABS52071.2021.9702683>
- Putri, H. R., & Dhini, A. (2019). Prediction of Financial Distress: Analyzing the Industry Performance in Stock Exchange Market using Data Mining. *2019 16th International Conference on Service Systems and Service Management (ICSSSM)*, 1–5. <https://doi.org/10.1109/ICSSSM.2019.8887824>
- Raya, M., Srinivasan, D., Vishnu, M., Adedoyin, A., & Sathiyarayanan, M. (2022). Visualizing, Comparing and Forecasting Stock Market Prediction. *2022 IEEE Delhi Section Conference, DELCON 2022*, 1–7. <https://doi.org/10.1109/DELCON54057.2022.9753359>
- Riko Anshori Prasetya, M., & Mudi Priyatno, A. (2023). Penanganan Imputasi Missing Values pada Data Time Series dengan Menggunakan Metode Data Mining. *Jurnal Informasi Dan Teknologi*, 5(2), 56–62. <https://doi.org/10.37034/jidt.v5i1.324>
- Rustam, Z., & Kintandani, P. (2019). Application of Support Vector Regression in Indonesian Stock Price Prediction with Feature Selection Using Particle Swarm Optimisation. *Modelling and Simulation in Engineering*, 2019, 1–5. <https://doi.org/10.1155/2019/8962717>
- S, K., S, K., S, K., & Anuradha, R. (2021). Stock Trend Prediction Using News Headlines. *2021 IEEE India Council International Subsections Conference (INDISCON)*, 1–5. <https://doi.org/10.1109/INDISCON53343.2021.9582219>

- S, S., N, H. P., & PD, R. (2022). Stock market Prediction using Reinforcement Learning Technique. *YMER Digital*, 21(07), 1022–1036. <https://doi.org/10.37896/ymer21.07/83>
- Sahu, S. K., Mokhade, A., & Bokde, N. D. (2023). An Overview of Machine Learning, Deep Learning, and Reinforcement Learning-Based Techniques in Quantitative Finance: Recent Progress and Challenges. *Applied Sciences*, 13(3), 1956. <https://doi.org/10.3390/app13031956>
- Saini, A., & Sharma, A. (2022). Predicting the Unpredictable: An Application of Machine Learning Algorithms in Indian Stock Market. *Annals of Data Science*, 9(4), 791–799. <https://doi.org/10.1007/s40745-019-00230-7>
- Sakhare, N. N., & Sagar Imambi, S. (2019). Performance analysis of regression based machine learning techniques for prediction of stock market movement. *International Journal of Recent Technology and Engineering*, 7(6), 655–662. https://api.elsevier.com/content/abstract/scopus_id/85065176950
- Yun, K. K., Yoon, S. W., & Won, D. (2023). Interpretable stock price forecasting model using genetic algorithm-machine learning regressions and best feature subset selection. *Expert Systems with Applications*, 213, 118803. <https://doi.org/10.1016/j.eswa.2022.118803>
- Zhang, Q., Qin, C., Zhang, Y., Bao, F., Zhang, C., & Liu, P. (2022). Transformer-based attention network for stock movement prediction. *Expert Systems with Applications*, 202, 117239. <https://doi.org/10.1016/j.eswa.2022.117239>