

IMPLEMENTASI METODE *RANDOM FOREST* DALAM MEMPREDIKSI SINYAL PERGERAKAN SAHAM

Moch. Anjas Aprihartha^{1§}, M. Husniyadi², Taufik Nur Alam³

¹Prodi PJJ Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
[Email: anjas.aprihartha@dsn.dinus.ac.id]

² Prodi PJJ Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
[Email: 118202200012@mhs.dinus.ac.id]

³ Prodi PJJ Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
[Email: 118202300068@mhs.dinus.ac.id]

[§]*Corresponding Author*

ABSTRACT

Trading involves purchasing stocks at low prices and then selling them at high prices to generate profits in a short period. Although it offers significant gains, trading is considered a high-risk activity. Careful analysis is required in stock trading to maximize profits and minimize losses. One way to analyze stocks is through technical analysis, a method used to predict price movements by understanding market actions with charts and technical indicators. One statistical method developed to predict stock trends is the random forest method. Random forest is a combination algorithm of several decision trees used to solve prediction or classification problems. The objective of this research is to obtain the best model for predicting stock price movements. Three types of datasets, namely deterministic, nondeterministic, and mixed, are applied for comparison. The data used is daily historical stock price data of Bank Central Asia Tbk (BBCA) for 10 years. The research results reveal that the best model is using the mixed dataset, constructed with $mtry = 6$ and $ntree = 500$. The resulting accuracy is 94,26%, indicating that the model accurately predicts the movement signals of BBCA stock by 94,26%, with the remaining 5,74% misclassification.

Keywords: *deterministic, mixed, nondeterministic, random forest, trading*

1. PENDAHULUAN

Meningkatkan aset melalui investasi saham merupakan salah satu strategi bijaksana dalam mengelola keuangan. Namun, terdapat opsi lain yaitu melalui *trading*. *Trading* merupakan kegiatan yang melibatkan pembelian saham di harga rendah kemudian menjual di harga tinggi untuk memperoleh keuntungan dalam waktu singkat. Meskipun memberikan keuntungan besar namun *trading* dianggap sebagai salah satu kegiatan yang beresiko cukup tinggi (Mahendra *et al*, 2022). Dalam *trading* saham diperlukan analisis yang cermat untuk meningkatkan keuntungan dan mengurangi resiko kerugian. Terdapat dua cara untuk menganalisis saham yaitu melalui analisis fundamental atau analisis teknikal.

Analisis fundamental merupakan metodologi yang umum digunakan investor dalam mengenali suatu saham dengan

memperhatikan kinerja perusahaan itu sendiri, kondisi keuangan perusahaan, dan tinjauan terhadap faktor-faktor ekonomi makro dan mikro yang memengaruhi kinerja perusahaan. Analisis teknikal merupakan metodologi populer yang digunakan dalam industri keuangan untuk memprediksi pergerakan harga di masa depan (Heinz *et al*, 2021). Studi ini mempelajari tentang memahami aksi pasar terutama melalui penggunaan grafik dan indikator teknikal untuk tujuan meramalkan tren harga di masa depan.

Beberapa teknik statistik telah dikembangkan untuk memprediksi tren saham. Pada awalnya menggunakan analisis regresi klasik. Namun data saham dapat dikategorikan sebagai data deret waktu nonstasioner maka teknik pembelajaran mesin dapat digunakan (Thakkar *et al*, 2015). Salah satu teknik pembelajaran mesin adalah metode *random*

forest.

Random Forest adalah algoritma pembelajaran mesin dalam kategori metode ansambel. Artinya algoritma *random forest* bekerja dengan menggabungkan beberapa pohon keputusan untuk membuat model yang menyerupai “hutan”. Kemudian dilakukan prediksi dengan mengambil keputusan paling banyak dari masing-masing pohon individu. Algoritma *random forest* dapat menjalankan tugas dalam mengatasi masalah regresi ataupun klasifikasi. Dalam meningkatkan generalisasi model, algoritma *random forest* akan memilih subset variabel independen secara acak. Algoritma ini terkenal karena akurasi dan kemampuannya yang sangat baik untuk menangani kumpulan data besar dengan dimensi tinggi dan juga membantu menurunkan *overfitting* (Alloubani *et al*, 2024).

Beberapa penelitian yang berkaitan dengan prediksi tren saham menggunakan algoritma *random forest*. Penelitian oleh Thakkar *et al* (2015) membandingkan algoritma *random forest*, *artificial neural network*, *support vector machine*, dan *naive bayes* dalam masalah memprediksi arah pergerakan saham dan indeks harga saham untuk pasar saham India. Studi kasus ini berfokus pada penggunaan parameter teknis sebagai dataset deterministik dan dataset nondeterministik. Variabel independen melibatkan perhitungan sepuluh parameter teknis menggunakan data saham perdagangan (harga pembukaan, tinggi, rendah, dan penutupan). Hasil penelitian menunjukkan bahwa dataset *nondeterministic*, *random forest* mengungguli tiga model prediksi lainnya pada kinerja keseluruhan. Penelitian serupa oleh Pratomo *et al* (2019) yang menerapkan metode *random forest* dalam memprediksi tren saham. Data yang digunakan dalam penelitiannya merupakan data historis harian dari tahun 2011 sampai 2017. Hasil yang diperoleh menunjukkan algoritma *random forest* memberikan akurasi yang lebih baik ketika direpresentasikan dalam dataset nondeterministik dibandingkan dataset deterministik.

Tujuan penelitian ini untuk memperoleh model terbaik dalam memprediksi sinyal pergerakan saham. Sinyal pergerakan saham mengacu pada petunjuk yang dihasilkan dari analisis data historis harga saham untuk menunjukkan arah pergerakan harga saham di masa depan, apakah akan naik atau turun.

Terdapat tiga jenis dataset yang akan digunakan yaitu dataset deterministik,

nondeterministik, dan campuran (*mixed*) sebagai perbandingan dalam menghasilkan model prediksi terbaik. Sepuluh indikator analisis teknikal dapat digunakan dalam memprediksi tren harga saham seperti *Simple Moving Average* (SMA), *Weight Moving Average* (WMA), *Momentum* (M), *Stochastic K%* (SK), *Stochastic D%* (SD), *Relative Strenght Index* (RSI), *Moving Average Convergence Divergence* (MACD), *Larry William's R%* (LWR), *Accumulation/Distribution Oscillator* (ADO), *Commodity Channel Index* (CCI).

Hasil prediksi dengan model *random forest* dapat digunakan dalam pengambilan keputusan bagi *trader*. Apabila indikasi sinyal akan naik maka investor disarankan untuk membeli saham sebaliknya jika sinyal akan turun maka investor disarankan menjual saham.

2. METODE PENELITIAN

2.1 Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang berasal dari data historis harian harga saham Bank Central Asia Tbk (BBCA). Pemilihan saham BBCA didasarkan pada kapitalisasi pasar terbesar pada tahun 2024.

Variabel yang digunakan dalam perhitungan indikator teknis terdiri dari harga pembukaan, harga penutupan, harga tertinggi dan harga terendah. Jumlah amatan yang digunakan sebanyak 2.476 hari terhitung pada periode 2014 sampai 2024. Sumber data diperoleh dari <https://finance.yahoo.com/>.

2.2 Variabel Penelitian

Variabel penelitian terdiri dari variabel independen dan variabel dependen. Variabel independen adalah sepuluh indikator teknikal. Indikator teknikal pada dasarnya merupakan perhitungan dari harga pembukaan, harga penutupan, harga tertinggi dan harga terendah yang umumnya digunakan investor dalam memprediksi harga saham dimasa depan (Murphy, 1999). Adapun sepuluh indikator teknikal sebagai berikut (Thakkar *et al*, 2015).

1. *Simple Moving Average*

$$SMA = \frac{C_t + C_{t-1} + \dots + C_{t-9}}{n} \quad (1)$$

dengan C_t merupakan harga penutupan hari ke- t

2. *Weight Moving Average*

$$WMA = \frac{10.C_t + 9.C_{t-1} + \dots + 1.C_{t-9}}{n + (n-1) + \dots + 1} \quad (2)$$

dengan $n = 10$

3. *Momentum*

$$M = C_t - C_{t-9} \quad (3)$$

4. *Stochastic K%*

$$SK = \frac{C_t - LL_{t-(n-1)}}{HH_{t-(n-1)} - LL_{t-(n-1)}} \times 100 \quad (4)$$

dengan LL_t and HH_t merupakan harga rendah terendah dan harga tinggi tertinggi saat hari ke- t

5. *Stochastic D%*

$$SD = \frac{\sum_{i=0}^{n-1} SK_{t-i}}{10} \% \quad (5)$$

6. *Relative Strenght Index*

$$RSI = 100 - \frac{100}{1 + \frac{\sum_{i=0}^{n-1} UP_{t-i}}{\sum_{i=0}^{n-1} DW_{t-i}}} \quad (6)$$

dengan UP_t dan DW_t merupakan perubahan harga naik dan perubahan harga turun saat hari ke- t

7. *Moving Average Convergence Divergence*

$$MACD = MACD(n)_{t-1} + \frac{2}{n+1} \times (DIFF_t - MACD(n)_{t-1}). \quad (7)$$

dengan $DIFF = EMA(12)_t - EMA(26)_t$, EMA merupakan *exponential moving average*, $EMA(k)_t = EMA(k)_{t-1} + \alpha \times (C_t - EMA(k)_{t-1})$, $\alpha = 2/(k+1)$ is faktor pengalus, k merupakan periode hari ke k pada *exponential moving average*.

8. *Larry William's R%*

$$LWR = \frac{H_n - C_t}{H_n - L_n} \times 100 \quad (8)$$

9. *Accumulation/ Distribution Oscillator*

$$ADO = \frac{H_t - C_{t-1}}{H_t - L_t} \times 100 \quad (9)$$

dengan H_t dan L_t merupakan harga tertinggi dan harga terendah hari ke- t

10. *Commodity Channel Index*

$$CCI = \frac{M_t - SM_t}{0.015 D_t} \quad (10)$$

$$\text{dengan } M_t = \frac{H_t + C_t + L_t}{3}, \quad SM_t = \frac{\sum_{i=1}^n M_{t-i+1}}{n},$$

$$D_t = \frac{\sum_{i=1}^n |M_{t-i+1} - SM_t|}{n}.$$

Sinyal arah pergerakan harga saham (naik/turun) dinyatakan sebagai variabel dependen yang bersifat biner (Thakkar *et al*, 2015).

$$y_t = \begin{cases} 1; C_t > C_{t-1} \\ -1; C_t \leq C_{t-1} \end{cases} \quad (11)$$

dengan $y_t = 1$ menyatakan pergerakan harga

naik pada hari ke t dan $y_t = -1$ menyatakan pergerakan harga turun pada hari ke t .

2.3 Preprocessing Data

Preprocessing data merupakan tahapan awal sebelum melakukan analisis data dengan metode *data mining*. Tujuan dilakukan *preprocessing data* untuk menghasilkan dataset yang berkualitas sehingga dapat menghasilkan model prediksi yang lebih akurat. *Preprocessing data* meliputi penanganan data yang tidak hilang (*missing value*) dan transformasi data.

Transformasi data merupakan teknik dalam mengubah struktur data ke dalam skala yang sama. Setiap variabel yang memiliki jangkauan data yang berbeda akan memengaruhi hasil prediksi dimana variabel dengan jangkauan yang besar lebih berpengaruh kuat dibandingkan dengan variabel dengan jangkauan yang kecil (Prasetyo, 2014). Teknik transformasi yang digunakan sebagai berikut.

1. Trend Deterministic Data Preparation

Transformasi deterministik merupakan teknik mengubah data kontinu suatu variabel independen menjadi data diskret bernilai -1 atau 1. Nilai 1 menunjukkan pergerakan harga naik dari hari sebelumnya sebaliknya nilai -1 menunjukkan pergerakan harga turun dari hari sebelumnya (Thakkar *et.al*, 2015).

2. Sigmoidal Bipolar

Transformasi sigmoidal bipolar merupakan teknik mengubah data kontinu kedalam skala -1 sampai 1. Fungsi sigmoidal bipolar dinyatakan sebagai berikut (Pratomo *et al*, 2019).

$$v'_i = \frac{1 - e^{-x}}{1 + e^{-x}} \quad (12)$$

dengan $x = \frac{k_{ij} - \bar{k}_j}{sd}$, k_{ij} menyatakan data ke- i pada variabel ke- j , $\bar{k}_j$ menyatakan rata-rata variabel ke- j , dan sd menyatakan standar deviasi.

2.4 Random Forest

Random forest merupakan metodologi yang didasarkan pada algoritma klasifikasi *Classification and Regression Tree* (CART) (Brieman, 2001). Pohon yang dihasilkan mewakili aturan keputusan biner yang secara rekursif membagi data pada variabel yang memaksimalkan *information gain* (IG) disetiap pemisah. *Information gain* didefinisikan sebagai

berikut.

$$IG = -\sum_{j=1}^n P(j|t) \log_2 P(j|t) \quad (13)$$

dengan $P(j|t) = \frac{n_j(t)}{n(t)}$ adalah proporsi kelas j pada simpul t untuk $j = 1, 2, \dots, n$, $n_j(t)$ adalah banyaknya pengamatan kelas j pada simpul t , dan $n(t)$ adalah banyaknya pengamatan pada simpul t .

Setelah menghitung *information gain* selanjutnya menghitung *gini index*. Gini index mencerminkan probabilitas kesalahan mengklasifikasikan pengamatan.

$$i(t) = 1 - \sum_{j=1}^n P^2(j|t) \quad (14)$$

Pada prinsipnya metode *random forest* memanfaatkan kumpulan pengklasifikasi lemah (pohon dengan kompleksitas yang direduksi) ke sampel *bootstrap* (sampel acak dari kumpula data yang diambil dengan pengembalian) untuk mencapai hasil yang lebih kuat dan akurat (Nieland *et al.*, 2024). Model *random forest* tidak dibangun oleh semua variabel independen tetapi dengan sejumlah kecil variabel independen.

Misalkan m variabel independen yang dipilih secara acak dari p variabel, dengan $m < p$. Menurut Prasetya (2022), menentukan banyaknya m variabel yang digunakan dalam membuat pohon dapat menggunakan tiga cara yaitu

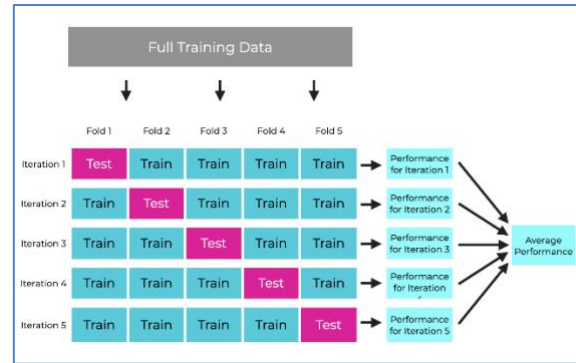
$$m = \frac{1}{2} \times \sqrt{p} \text{ atau } m = \sqrt{p}$$

$$\text{atau } m = 2 \times \sqrt{p}. \quad (14)$$

Dalam mengatasi masalah multikolinearitas, algoritma akan memilih variabel-variabel independen disetiap pemisahan kandidat tanpa pengembalian. Hal ini memungkinkan agar antarvariabel independen tidak saling berkorelasi.

2.5 *k-Fold Cross Validation*

Metode *k-fold cross validation* adalah metode untuk menilai dan menguji efektivitas model pembelajaran mesin (Anandan & Manikandan, 2023). Validasi silang *k-fold* sering digunakan dalam aplikasi berbasis *machine learning* untuk mengatasi masalah *overfitting*. Nguyen *et al.* (2021) mengatakan bahwa validasi silang *k-fold* berhasil meminimalkan masalah *overfitting*.



Gambar 1. *k-fold Cross Validation*

Langkah-langkah dalam menggunakan teknik *k-fold cross validation* sebagai berikut.

1. Menentukan jumlah lipatan (k).
2. Membagi data berdasarkan k kelompok
3. Melakukan pengulangan sebanyak k kali. Pada tiap pengulangan, dipilih satu kelompok sebagai data *testing* dan $k-1$ kelompok sebagai data *training*.
4. Melakukan pelatihan model dengan data *training*.
5. Menguji performa model
6. Ulangi proses 3 sampai 5 untuk setiap kelompok.
7. Hitung rata-rata kinerja setiap model.

2.6 Uji Performa Model

Confusion matrix merupakan tabel yang memberikan hasil dari performa model yang telah diuji dengan data *testing*. Melalui *confusion matrix* maka dapat mempermudah dalam menghitung akurasi. Akurasi merupakan persentase dari total keseluruhan amatan yang terprediksi atau terklasifikasi secara tepat.

Tabel 1. *Confusion Matrix*

		Kelas Prediksi	
		Ya	Tidak
Kelas Observasi	Ya	a	b
	Tidak	c	d

$$\text{Akurasi} = \frac{a+d}{a+b+c+d} \times 100\% \quad (15)$$

Interpretasi kategori nilai akurasi sebagai berikut (Maskoen, T. T., & Masthura, 2017).

- 50% - 59,99% = sangat lemah
- 60% - 69,99% = lemah
- 70% - 79,99% = sedang
- 80% - 89,99% = kuat
- 90% - 100% = sangat kuat

2.7 Teknik Analisis Data

1. Mengumpulkan data historis saham BBKA selama 10 tahun.
2. Menghitung setiap indikator analisis teknikal sebagai variabel independen.
3. Menentukan kelas untuk variabel dependen.
4. Menyusun variabel independen dan variabel dependen menjadi sebuah dataset.
5. Melakukan pembersihan data yang tidak lengkap.
6. Mentransformasi data menggunakan teknik sigmoidal bipolar dan teknik *trend deterministic data preparation*.
7. Membuat model *random forest*.
8. Menguji performa model.
9. Implementasi hasil.

3. HASIL DAN PEMBAHASAN

3.1 Preprocessing Data

Variabel independen dibentuk melalui perhitungan pada 10 indikator teknikal pada persamaan (1) sampai (10) sedangkan variabel dependen ditentukan berdasarkan arah sinyal harga saham dengan persamaan (11). Selanjutnya variabel-variabel tersebut dikumpulkan menjadi dataset.

Pada perhitungan indikator *A/D Oscillator* dan *Relatif Strenght Index* ditemukan *missing value* yang disebabkan pada perhitungan terdapat pembagi yang sama dengan nol. Untuk mengatasi *missing value* maka amatan tersebut dapat diisi dengan nilai rata-rata atau median dari variabel yang terdapat *missing value* (Sari & Azhar, 2023).

Terakhir dilakukan transformasi dengan dua cara yaitu transformasi *trend deterministic data preparation* dan transformasi *sigmoidal bipolar* menggunakan persamaan (12). Dataset yang telah ditransformasi *trend deterministic data preparation* dinamakan dataset deterministik. Dataset yang ditransformasi *sigmoidal bipolar* dinamakan dataset nondeterministik. Sedangkan dataset *mixed* merupakan dataset yang terdiri dari lima variabel independen yang diambil secara acak dari dataset nondeterministik (ADO, LWR, MACD, CCI, dan RSI) dan lima variabel independen sisanya dari dataset deterministik (MA, WMA, M, SK, dan SD).

3.2 Pemodelan *Random Forest*

Pembentukan pengklasifikasi lemah diawali dengan menentukan jumlah variabel yang akan digunakan dalam membangun pohon. Pada 10 variabel independen akan diambil m variabel secara acak. Banyaknya m variabel dihitung menggunakan persamaan (14) sehingga diperoleh $mtry = \{2, 3, 6\}$. Untuk mengetahui performa dari masing-masing m dapat dilakukan pengujian kesalahan model yang disebut estimasi *out of bag error* (OOB).

Proses mendapatkan estimasi OOB dilakukan dengan mengambil amatan secara *bootstrap* sehingga terbentuk sampel *bootstrap*, beberapa amatan yang tidak terambil disebut *out of bag data* (Culter *et al*, 2012). Pada sampel *out of bag data* dibangun pohon-pohon keputusan yang selanjutnya digunakan untuk prediksi sampel data *training*. *Out of bag error random forest* diperoleh dengan menghitung rata-rata tingkat kesalahan prediksi pada setiap pohon dari sampel *out of bag data*.

Tabel 2. Nilai OOB pada $mtry$

$mtry$	Deterministik	Non-deterministik	<i>Mixed</i>
	OOB	OOB	OOB
2	8.49%	14.51%	5.74%
3	8.84%	13.98%	5.09%
6	9.25%	13.17%	4.86%

Pada masing-masing $mtry$ akan diujikan dengan menggunakan 500 pohon untuk membentuk model *random forest*. Berdasarkan Tabel 2, nilai OOB tertinggi adalah 14,51% dengan $mtry = 2$ pada dataset nondeterministik sedangkan nilai OOB terendah adalah 4,86% dengan $mtry = 6$ pada dataset *mixed*. Maka nilai $mtry = 6$ pada dataset *mixed* merupakan yang terbaik.

Tabel 3. Nilai OOB pada $ntre$

$ntre$	Deterministik	Non-deterministik	<i>Mixed</i>
	OOB	OOB	OOB
100	9.02%	13.81%	5.39%
300	8.43%	13.46%	5.09%
500	8.49%	13.17%	4.86%
700	8.49%	13.63%	4.98%
900	8.55%	13.22%	5.09%

Proses selanjutnya menentukan *ntree* terbaik berdasarkan *mtry* terbaik pada masing-masing dataset. *Ntree* merupakan banyaknya pohon yang dibangun untuk menciptakan model *random forest*. Pada Tabel 3, Nilai OOB tertinggi adalah 13,81% dengan *ntree* = 100 pada dataset nondeterministik sedangkan nilai OOB terendah adalah 4,86% dengan *ntree* = 500 pada dataset *mixed*. Maka nilai *ntree* = 500 pada dataset *mixed* merupakan yang terbaik. Pembentukan model *random forest* dilakukan dengan *mtry* dan *ntree* terbaik pada masing-masing dataset. Pada dataset deterministik dengan *mtry* = 2 dan *ntree* = 300, dataset nondeterministik dengan *mtry* = 6 dan *ntree* = 500, dan dataset *mixed* dengan *mtry* = 6 dan *ntree* = 500.

3.3 Evaluasi Performa Model

Langkah awal menetapkan sebanyak *k* lipatan model. Garre *et al.* (2020) mengatakan pendekatan 10-fold *cross validation* dapat meminimalkan *overfitting* untuk beberapa algoritma pembelajaran mesin termasuk *random forest*. Proses pembentukan model *random forest* hingga uji model dengan data *testing* menggunakan *software* R 4.3.2. menggunakan *package* *randomForest*. Ukuran performa model dapat dihitung menggunakan persamaan (15).

Tabel 4. Akurasi 10-fold *Cross Validation*

<i>i-fold</i>	Akurasi		
	Deterministik	Non-deterministik	<i>Mixed</i>
1	91,83%	87,75%	94,69%
2	93,85%	89,75%	94,26%
3	90,57%	86,06%	91,39%
4	93,03%	85,65%	91,39%
5	90,57%	90,16%	95,08%
6	88,11%	86,06%	95,90%
7	90,57%	92,21%	96,72%
8	91,80%	84,01%	94,67%
9	93,44%	83,19%	94,26%
10	95,08%	87,75%	94,26%
Rata-rata	91,88%	87,26%	94,26%

Pada Tabel 4 memperlihatkan akurasi masing-masing lipatan dan rata-rata akurasi setiap dataset. Perbandingan hasil dataset deterministik, nondeterministik, dan *mixed* menghasilkan rata-rata akurasi masing-masing sebesar 91,88%, 87,26%, dan 94,26%. Model *random forest* terbaik menghasilkan rata-rata akurasi tertinggi. Oleh karena itu, model dengan dataset *mixed* merupakan model terbaik. Rata-

rata akurasi sebesar 94,26% menjelaskan bahwa model mampu memprediksi dengan tepat sinyal pergerakan arah saham BBKA sebesar 94,26% sedangkan sisanya 5,74% misklasifikasi.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil pengujian yang dilakukan dengan metode *random forest* diperoleh model terbaik dengan dataset *mixed*. Dataset *mixed* terdiri dari lima variabel independen yang diambil secara acak dari dataset nondeterministik (ADO, LWR, MACD, CCI, dan RSI) dan lima variabel independen sisanya dari dataset deterministik (MA, WMA, M, SK, dan SD). Model ini dibangun dengan *mtry* = 6 dan *ntree* = 500 sehingga dihasilkan model dengan rata-rata akurasi tertinggi sebesar 94,26%. Akurasi sebesar 94,26% menunjukkan bahwa model mampu memprediksi dengan tepat sinyal pergerakan arah saham BBKA sebesar 94,26% sedangkan sisanya 5,74% misklasifikasi.

Dalam penelitian selanjutnya dapat menggunakan variabel independen yang lain untuk meningkatkan akurasi model. Selain itu, pengujian kedepannya dapat membandingkan metode lain seperti *naive bayes*, *k-nearest neighbor*, regresi logistik, dsb.

DAFTAR PUSTAKA

- Alloubani, A., Abuhaija, B., Almatari, M., Jaradat, G., & Ihnaini, B. 2024. Predicting vitamin D deficiency using optimized random forest classifier. *Clinical Nutrition ESPEN*, 60, pp.1-10.
- Anandan, B., & Manikandan, M. 2023. Machine learning approach with various regression models for predicting the ultimate tensile strength of the friction stir welded AA 2050-T8 joints by the K-Fold cross-validation method. *Materials Today Communications*, 34, 105286.
- Breiman, L., 2001. Random Forest. *Machine Learning*. 45, pp.5-32.
- Cutler, A., Cutler, D. R., & Stevens, J. R. 2012. Random forests. *Ensemble machine learning: Methods and applications*, pp.157-175.
- Garre, A., Ruiz, M. C., & Hontoria, E. 2020. Application of Machine Learning to support production planning of a food industry in the

- context of waste generation under uncertainty. *Operations Research Perspectives*, 7, 100147.
- 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, pp.221-244.
- Mahendra, K., Satyahadewi, N., & Perdana, H. 2022. Analisis Teknikal Saham Menggunakan Indikator Moving Average Convergence Divergence (Macd). *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, 11(1).
- Maskoen, T. T., & Masthura, A. 2017. Nilai Area Under Curve dan Akurasi Neutrophil Gelatinase Associated Lipocalin untuk Diagnosis Acute Kidney Injury pada Pasien Politrauma di Instalasi Gawat Darurat RSUP dr. Hasan Sadikin Bandung. *Maj Anest dan Crit Care*, 35 (3), pp.158–64.
- Murphy, J. J. 1999. *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. Penguin.
- Nguyen, X. C., Nguyen, T. T. H., La, D. D., Kumar, G., Rene, E. R., Nguyen, D. D., ... & Nguyen, V. K. 2021. Development of machine learning-based models to forecast solid waste generation in residential areas: A case study from Vietnam. *Resources, Conservation and Recycling*, 167, 105381.
- Nieland, S., Oostendorp, R., Heinrichs, M., & Cyganski, R. 2024. Transferability analysis of user groups in travel behaviour surveys using a random forest classification model. *Transportation Research Procedia*, 76, pp.81-95.
- Pratomo, A., Umbara, R. F., & Rohmawati, A. A. 2019. Prediksi Pergerakan Harga Saham dengan Metode Random Forest Menggunakan Trend Deterministic Data Preparation (Studi Kasus Saham Perusahaan PT Astra International Tbk, PT Garuda Indonesia Tbk, dan PT Indosat Tbk). *eProceedings of Engineering*, 6(1).
- Prasetya, J. 2021. *Perbandingan Analisis Klasifikasi Smote Random Forest Dan Smote K-Nearest Neighbors Pada Data Tidak Seimbang* (Doctoral dissertation, Universitas Gadjah Mada).
- Prasetyo, E. 2014. *Data Mining - Mengolah Data menjadi Informasi Menggunakan Matlab*. Yogyakarta: ANDI
- Sari, W. K., & Azhar, I. S. B. 2023. Perbandingan Kinerja Neural Network dengan Metode Klasifikasi Tradisional dalam Mendiagnosis Penyakit Jantung: Sebuah Studi Komparatif. *JSI: Jurnal Sistem Informasi (E-Journal)*, 15(1).
- Thakkar, P., Patel, J., Shah, S., & Kotecha, K. 2015. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert systems with applications*, 42(1), pp.259-268.